Jingyang Liu

#12161921

PS 1

Problem 1.

(a) The article is "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets" from American Economic Review.

(b) Detailed citation of the article:

Tahir Andrabi, Jishnu Das, Asim Ijaz Khwaja. 2017. "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets." American Economic Review, 107(6): 1535-1563. https://doi.org/10.1257/aer.20140774

(c) The equation is:

$$Y_{mi2} = \alpha_d + \beta_0 RC_m + \beta_1 GOV_{mi} + \beta_2 HIGH_{mi1} + \beta_3 RC_m \cdot GOV_{mi}$$

$$+ \beta_4 RC_m \cdot HIGH_{mi1} + \beta_4 RC_m \cdot GOV_{mi} \cdot HIGH_{mi1} + \gamma \cdot Y_{mi1}$$

$$+ \delta \cdot X_{m1} + \varepsilon_{mi},$$

(d) In this model, the endogenous variable is Y_{mi2}, which represents the outcome of interest (such as fees or enrollment) for school i in village m in time period 2 (post-intervention year).

The exogenous variables in the model are as follows:

α_d: district fixed-effects

RC_m: treatment dummy assigned to village m. If the village received the treatment, RC_m=1.

GOV_{mi}: a dummy variable 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

Y_{mi1}: the baseline of the outcome variable

X_{m1}: the vector of baseline village-level controls (such as literacy, size and wealth)

(e) This model is a static model because it doesn't contain an element of time and the interactions between variables over time.

This model is nonlinear because it contains multiplication.

This model is stochastic because it also contains random variable.

(f) The valuable variable that the model might miss can be the amount of funds these schools received during the treatment periods. The endogenous various represents the outcome of interest (such as fees), which can be influenced by these funds the schools received. For those schools which got larger amount of funds, they may have lower level of tuition fees or higher enrollment.

Problem 2.

(a)(b)

Since the outcome variable Y_i is a binary dependent endogenous variable, I use the logistic regression, and P is the probability of deciding to get married.

$$g(P(Y_i = 1)) = \beta_0 + \beta_1 Age_i + \beta_2 Gen_i + \beta_3 Inc_i + \beta_4 Edu_i + \beta_5 Race_i + \beta_6 Rel_i + \beta_7 Sex_i + \varepsilon_i$$

And Y_i=1 if $g(P(Y_i = 1)) > 0$. Otherwise, $Y_i = 0$

The endogenous variable in the model is Y_i, which is a dummy variable for whether the individual I decides to get married. If $Y_i = 1$, the individual I decides to marry.

The exogenous variables are as follows:

Age_i: the age of individual i; Gen_i: the gender of individual i, which is a dummy variable. If a

male, $Gen_i = 1$; Inc_i: the income level of individual i, a categorical variable; Edu_i is the education level of i, categorical variable; $Race_i$ is the race of i, categorical variable; Rel_i is the religion of i, categorical variable; Sex_i is the sexual orientation of i, which is a dummy variable, if heterosexual, then $Sex_i = 1$. ε_i denotes error term.

(c) In the model, we can use the sample data to estimate the values of betas. When they are known, using them in conjunction with the rest of this model, to generate values of Yi.

(d) The key factors are age, income, education and religious status.

(e) For age, there are higher social pressure and peer pressure for older people if they have not got married, so they are more willing to decide to marry. For gender, female may be more likely to have plans about marriage. For income, people may not decide to marry when they believe their current income could not support the family or raise the children. People with different educational background have various attitudes towards marriages. People of different racial background have various attitudes towards marriages. Religion beliefs can also encourage or discourage people to marriage. In some areas, same-sex marriage is still illegal, which can influence some same-sex people's decision of marriage.

(f) To conduct preliminary test, we could use data from census bureau or labor force department. Then, we use these open-source data to conduct the preliminary test. However, the available data my not provide all the information we need for the test, so that we can carry out the anonymous survey, online and offline, to create the dataset used for our test. With these data, we can conduct the statistical analysis to determine which factors are significant. In the end, we can also check the robustness of the model.