Problem Set # 1

MACS 30100, Dr. Evans
Xi Chen

1. Classify a model from a journal

Part (a-b). Citation

Acconcia, A., Corsetti, G., & Simonelli, S. (2014). Mafia and public spending: Evidence on the fiscal multiplier from a quasi-experiment. *The American Economic Review, 104* (7), 2185-2209.

Part (c). Model

The following model estimates the spending multiplier relating the growth of per capita value added in a province to the year-on-year change in per capita spending on infrastructure in the same province.

$$Y_{i,t} = \beta G_{i,t} + \alpha_i + \lambda_t + \gamma \mathbf{X}_{i,t} + v_{i,t}$$

where i denotes each province in Italy, and t denotes the year; $Y_{i,t}$ denotes the growth of per capita value added in a province; $G_{i,t}$ denotes the year-on-year change in per capita spending on infrastructure in the same province; the parameter β measures the contemporaneous one-year government spending multiplier; α_i denotes the province fixed effect, which controls for national components of public investment and GDP common to all provinces; λ_t denotes the year fixed effect, which control for monetary and fiscal policy at the national level; $\mathbf{X}$ denotes a vector of further control variables; $v_{i,t}$ denotes the error term.

Part (d). Variables

The **exogenous variables** are $G_{i,t}$, which is the year-on-year change in per capita spending on infrastructure in a province; α_i, which is the province fixed effect; λ_t, which is the year fixed effect; $\mathbf{X}$, which represents the control variables including (i) Mafia-type association, (ii) extortion, (iii) Mafia-related murders, (iv) corruption, and (v) the number of corruption crimes reported to the judicial authority). The **endogenous variables** is $Y_{i,t}$, the growth of per capita value added in a province, which is also the output of the model.

Part (e).

The model is static, linear, and stochastic. "Static" is because $Y_{i,t}$ only depends on the independent variables on time t. "Linear" is because there are no non-linear terms. "Stochastic" is because of the random error term.

Part (f).

One more control variable might be added into the model is the provincial government spending on public security.

2. Make your own model

Part (a-c). Model

Since the outcome variable Y_i is binary (*1* or *0*), I use logistic regression, and define P as the probability of deciding to get married. Therefore, by definition, the inverse of the logistic function, g, is

$$g(P(Y_i = 1)) = ln(\frac{P(Y_i = 1)}{1 - P(Y_i = 1)}) = \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 Edu_i + \beta_4 Ses_i + \beta_5 Bmi_i + \beta_6 Race_i + \beta_7 Status_i + \beta_8 Happy_i + \epsilon_i$$

And

$$Y_i = \begin{cases} 1 & g(P(Y_i = 1)) > 0 \\ 0 & otherwise \end{cases} \quad (1)$$

where Age_i, a categorical variable, denotes individual i's age group; Sex_i, a categorical variable, denotes gender; Edu_i, a continuous variable, denotes the number of years of higher education; Ses_i, a categorical variable, denotes the socioeconomic status (High, Middle, Low); Bmi_i, a continuous variable, denotes the Body Mass Index; $Race_i$, a categorical variable, denotes the individual's race group; $Status_i$, a continuous variable, denotes the number of years in a current relationship (if he or she is single, it is zero); $Happy_i$, a continuous variable, denotes the self-perceived happiness or sense of satisfaction from a psychological scale; ϵ_i denotes the random error term.

Part (d). Key Factors

Among all the factors, I think the key ones would be age, gender, socioeconomic status, and the relationship status.

Part (e). Explanation

Firstly, I am going to explain the key factors. For *Age*, there are higher social pressure and/or peer pressure for older people if they haven't got married, so the older people would be more likely to decide to get married. For *Sex*, compared to male, female may be more likely to have plans or have earlier plans for marriage because they have to consider the plan of having a baby. For *Ses*, getting married is a big decision in one's life, so his/her socioeconomic status would play an important role in such big decision. For *Status*, it is obvious that when an individual has a healthy relationship which lasts for a long time, the more likely he or she will decide to get

married. Secondly, except for these key factors, there are still some other factors may make a difference. For example, past studies have showed a significant relationship between education level and the number of children, but it is hard to tell what would the relationship be between education and marriage. BMI score could also affect the possibility of deciding to get married, and my predication is the healthier the BMI score one has, the high probability he or she will decide to get married. Race may affect one's decision due to the cultural or family reasons. Psychological status may also influence one's decision of getting married. It could be the case that the happier single people may feel getting married less attractive because they find their status quo is good, and don't want to change.

Part (f). Preliminary Test

In order to see if the model make sense in real life, I could download some open-source census data, such as from the website of the labor force department. For the purpose of preliminary test, the size of the data would be small. Then I could use these small samples to conduct the preliminary test with the real-world data. However, these census data may not provide all the information I need. Another solution could be collecting data by carrying out a survey study with several participants. The survey would collect all the required information. With these data, I could conduct statistical analysis, such as OLS, to determine which factors are significant.