Problem Set # 1

MACS 30100
1/14/2019
Minkwang Jang

1. Classify a model from a journal

(a) and (b)

Zhang, X. M., & Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review*, 101(4), 1601-15.

This paper discusses a natural field experiment about pro-social behavior using contribution data on Chinese Wikipedia by adding, editing or deleting pages and the effect of group size that is naturally manipulated due to the Chinese government blocking Wikipedia.

(c)

The following model estimates the weekly contributions of each nonblocked contributor i to Chinese Wikipedia articles at week t:

$$\begin{aligned} Contributions_{it} = \beta_0 + \beta_1 AfterBlock_t + \\ \beta_2 PercentageBlocked_i * AfterBlock_t + \\ \beta_3 PercentageBlocked_i + ControlVars_{it} + \epsilon_{it} \end{aligned}$$

$AfterBlock_t$ is a dummy that equals 1 if the time period is after the block, and zero otherwise. $PercentageBlocked_{it}$ measures the percentage of collaborators blocked after the block for contributor i. $ControlVars_{it}$ include Age and Age^2.

(d)

The **exogenous variables** are $AfterBlock_t$, $PercentageBlocked_{it}$, and $ControlVars_{it}$ with consist of Age and Age^2 (age of the contributor).

The **endogenous variables** is $Contributions_{it}$.

(e)

The model is static because $Contributions_{it}$ only depends on the independent variables at time t. It is also linear is because there are no non-linear terms. It is stochastic because of the random error term ϵ_{it}.

(f)

Though it cannot be measured within the setting of this natural field experiment, a psychological measure of 'Need for Recognition' by the contributors, if it were available, could be a valuable addition to the model in illuminating the predictors of pro-social contribution. It could have been operationalized by their preference for having their names published or any other measure in regards to their contributions being visible or not.

2. Make your own model

(a)

$$ln(\frac{P(GetMarried_{it} = 1)}{1 - P(GetMarried_{it} = 1)}) = \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 Educ_i$$
$$+\beta_4 Age_i * Gender_i + \beta_5 Educ_i * Gender_i + \beta_6 CurrentRelationship_{it}+$$
$$\beta_7 MarriageView_i + \beta_8 Religion_t + \beta_9 Location_i + \beta_{10} Occupation_i+$$
$$\beta_{11} Income_{it} + \epsilon_i$$

Whether someone (individual i) decides to get married at a given time t, ($GetMarried_{it}$), could first depend on the person's age, sex, and education level (Age_i, Sex_i, $Educ_i$). The effect of age and education level on the decision to get married would be different for men and women (or any third gender).

(b)

The outcome variable (dependent endogenous variable) $GetMarried_i$ is defined as *1=get married* or *0=not get married*

$$Y_i = \begin{cases} 1 & ln(\frac{P(GetMarried_i=1)}{1-P(GetMarried_i=1)}) > 0 \\ 0 & otherwise \end{cases}$$

(d)

The key factors might be age, gender, education, income and the current status of relationship ($CurrentRelationship_{it}$).

(e)

Age is correlated with many other things going on in life and generally people base their decision to get married in where they are in life, thus age would be a strong key predictor of the decision to get married.

The other variables such as religion, location, and occupation could potentially contribute to the prediction but may be more heterogeneous, thereby weaker predictors for making a prediction for a wider population.

(f)

While some of the variables in my model would be more readily available than others based on archival data or census data (e.g., demographic data such as gender and age), other variables would be harder to yield. A somewhat costly but plausible way is to recruit a nationally representative sample and run a survey, surveying people about the factors in the model as well as their marriage status as a retrospective indicator of the decision to get married.