

Problem Set #1

MACS 30100, Dr. Evans

Xinyu Cao

Problem 1 Classify a model from a journal

Part (b) This answer will introduce a study from American Economic Review. There is a phenomena in housing market that systematic above-trend increases in prices and transactions during the spring and summer("hot season") and below-trend falls during the autumn and winter("cold season"). The study tried to answer why informed buyers do not try to buy in the lower-price season. Therefore, they study proposed a model of search friction and "thick-market effect" and tested those hypothesis by empirical data.(Ngai and Tenreyro 2014).

Part (c) One of the main regression model is to test the hypothesis that the average quality of match formed in the summer is higher than in the winter. Since the quality of the match could not be observed, the author considered the following proxy as one of the test. The main idea is that the better the fit between buyer and the house, the longer the buyer will stay. The model is formulated as a Poisson regression:

$$\log Y_i = \beta * MovedInSummer_i + \gamma X_i + \alpha_i + \epsilon_i$$

Part (d) exogenous variable: main independent variable of concern: MovedIn-Summer, a dummy variable if the buyer moved into unit in the summer, X, is a vector of other control variables that influence the length of the stay 1) (log) year unit was built, 2) (log) family income, 3) number of persons in the household 4) number of adults 18+ in household 5) children in school dummy. α , fixed effects control: region, central city/suburban status, heating/cooling degree days, CMSA (Consolidated Metropolitan Statistical Area) **exogenous variable:** Y, the length of the stay.

Part (e) This is a static (because we do not actually have a time depended variable in the model), nonlinear(Although we could estimate Poisson by log transformation and maximum likelihood method, Poisson regression is a typical nonlinear regression), and stochastic model(the variable Y is also influenced by some random error).

Part (f) The model is well defined. One of the variable I could think of is the ratio of housing prices and income of the family, which is a good proxy for an indication of the sunk cost, which could help to explain the the length of stay, because the higher the sunk cost, the longer the may stay due to behaviour reason. One control that I think should be useful is the religion of the family. There might be religion difference in the housing market. For example, some religion might value more of the family value, therefore less likely to move.

Problem 2 Build your own model

Part a-c The original question is not clear, we do not know when whether someone decides to get married is that in 1 year, or 5 years or in his life, and we do not know who is going to decide, for example do we need to consider someone who is actually in marriage and decide to broke up and get married with someone else? To simplify our question, I shall restrict my question to whether a single man/woman deiced to get married in one year time.

$$Y_i = \text{logit}^{-1}(\beta_0 + \beta_1 \text{status}_i + \beta_2 \text{income}_i + \beta_3 \text{age}_i + \beta_4 \text{education}_i \\ + \beta_5 \text{gender}_i + \beta_6 \text{race}_i + \beta_7 \text{culture}_i + \beta_8 \text{religion}_i + \epsilon_i)$$

Y_i is a dependent variable that be 1=get married or 0=not get married. logit^{-1} is the inverse logit function, status-relation status, a variable of the length in month of existing partner relationship(if none zero), ϵ_i is error term and distributed as independent identically normal distribution, other variables are quite intuitive, income-log income in terms of dollar of person, age-log age of the person, education-dummy of education of the person, gender-gender dummy of the person, race-race dummy of the person, culture-culture dummy of the person, religion- religion and belief the person.

Part d The main factors that influence this outcome should be relation status, age, and income. The first factor should be relation status, because current status should be a good predict for future marriage, it is reasonable to assume people have higher probability to get married if they are in a stable relation. Income also matters, because financial stability could make people feel secured to get married, form a family and raise children. Age is also an important indicator, because people usually get married at certain age, such as 30s or 40s. Finally, gender matters, because women are usually married earlier than men.

Part e First of all, I include those the main factors , as said in part d, that matters the most for the marriage decision. Except those main factors, I also include some control variable that might also alter the behavior of marriage decision, such as education, race, culture, and religion. Although we do not know yet if those variable will influence the marriage decision for sure, but it is very interesting to know about how they might influence the results. Therefore, it is reasonable to include them as control variable.

Part f We could collect digital trace panel data of age, income, education, gender, race, religion, culture, and marriage of marriage-matching website. For example, we could have an user that is 30 years old, male, annual income of 100,000 dollar, college education, in relationship for 5 month, white and Christian at 2014. Then we could see if he is still on website at 2015 as 1(proxy of married) and not on the website as 0(proxy of not married), thus we are able to make logistic regression based on the panel. Of course, this method could potentially subject to measurement errors, but is good enough to just do a preliminary test.

References

- ¹ L Rachel Ngai and Silvana Tenreyro. Hot and cold seasons in the housing market. *The American Economic Review*, 104(12):3991–4026, 2014.