

Problem Set #1

MACS 30150, Dr. Evans

Dongcheng Yang

Problem 1 Classify a model from a journal.

Part (a). The model that I choose is from the paper "Sales Taxes and Internet Commerce" published on American Economic Review.

Part (b). Einav, L., Knoepfle, D., Levin, J., Sundaresan, N. (2014). Sales taxes and internet commerce. American Economic Review, 104(1), 1-26.

Part (c). In the following equation, k index the items, i index the viewers of each item. It is assumed that

$$Pr(i \text{ buys } k \mid i \text{ views } k) = \frac{\exp(u_{ik})}{1 + \exp(u_{ik})}$$

where

$$u_{ik} = \alpha_k + \beta \log(1 + \tau_{ik}) + g(d_{ik}) + \gamma \mathbf{1}\{\text{state } i = \text{state } k\}$$

The first term α_k is a fixed effect that captures each item's general desirability. The second term is the effect of the relevant tax rate τ_{ik} and it is equivalent to the combined sales tax in the items ZIP code if i is a same-state buyer, and zero otherwise. The distance between the buyer and seller, denoted d_{ik} , serves as a control to account for the possibility that buyers may prefer nearby items. The last item is a dummy variable indicating whether the buyer is located in the same state as the item.

Part (d). In this model, only the conditional probability variable and u_{ik} are endogenous. The general desirability variable α_k , relevant tax rate variable τ_{ik} , distance variable d_{ik} and dummy variable are all exogenous.

Part (e). This model is a static, nonlinear and deterministic model.

Part (f). To estimate the purchase decision, we should also consider personal preference over different kinds of commodities. Thus, apart from the distance variable indicating consumers' preference over nearby items, some other measurement of individual's preference should be included in the estimation model.

Problem 2 Make your own model.

Part (a) - Part (c) The model that I write is a simple logistic model.

$$Y_i = \begin{cases} 0, & \text{if } P_i < 0.5 \\ 1, & \text{if } P_i \geq 0.5 \end{cases}$$

where

$$P_i = \frac{1}{1 + e^{-z_i}}$$

where

$$z_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{restriction}_i + \beta_3 \text{job}_i$$

The restriction variable is an overall indicator of one's attitude towards marriage. It could be calculated based on personal answers to the questions in a survey. The job variable is a dummy which manifests whether the individual has a stable job. This model is a complete data generating process.

Part (d) The key factors in the model are restriction and job. The restriction variable quantifies personal attitude towards marriage, which is indispensable in the model. The job factor is also important because material conditions are the basis for marriage. The people with a stable job tends to have more admirers and are more likely to be in a long-term relationship.

Part (e) The other potential variables to be included in this model are salary and parents' wealth. The reason that I do not include salary and wealth variables is because the data of money and asset value could easily be inauthentic. Wealthy people tend to hide and an average person is likely to boast.

Part (f) Firstly, we need to collect a dataset containing over 100 individuals' data. Each individual's data includes whether he or she is married, his or her age, overall restriction level to marriage and job indicator. Then, run the logistic regression of the marriage variable on the other three independent variables. We can see from the p-value whether the three variables are significant or not. We can also use scatterplots matrix to see the relationship.