Problem Set #1

MACS 30150, Dr. Evans
Yiqing Zheng

Problem 1. Classify a model from a journal

Part (a). The model I choose is from paper "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector" published on American Economic Review in 2016.

The paper uses empirical models to investigate whether hospitals with better performance are rewarded with larger market allocation. It finds out that higher quality hospitals have higher market shares and grow more over time.

Part (b). Detailed citation of the article:

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector." American Economic Review, 106 (8): 2110-44.

Part (c). The paper includes both static and dynamic allocation analysis and applies two models as below.

$$\ln N_h = \beta_0^s + \beta_1^s q_h + \gamma_M^s + \varepsilon_h^s \tag{1}$$

In equation 1, N_h is the number of Medicare patients with certain condition treated in hospital h in year 2008. The variable measures the market size of hospital h. γ_M^s are market fixed effects and q_h is the quality of hospital h. The article uses four quality metrics to conduct robustness check across different performance measures. The value of β_1^s reflects the effect of hospital performance on hospital market share in a fixed time point.

$$\Delta_h = \beta_0^d + \beta_1^d q_h + \gamma_M^d + \varepsilon_h^d \tag{2}$$

In equation 2, Δ_h measures the growth rate in admissions of hospital h. Other variables are defined in the same way as equation 1. The value of β_1^d measures how much the growth rate changes with different quality levels. Δ_h is calculated as

$$\Delta_h = \frac{N_{h,2010} - N_{h,2008}}{\frac{1}{2}(N_{h,2010} + N_{h,2008})}, \tag{3}$$

where $N_{h,t}$ is the number of Medicare patients with certain illness condition treated by hospital h in year t.

Part (d). The models and data in the articles are used to estimate parameters instead of forecasting market shares. As the result, in the static model, both hospital quality q_h and hospital market size N_h are exogenous variables and obtained directly from database. Coefficients β_0^s and β_1^s, and ε_h^s are endogenous variables and estimated by the model. Fixed effects measured by γ_M^s are also endogenous variables, for the magnitude of fixed effects is not determined outside the model. Actually, the values of γ_M^s can be considered as coefficients of dummies indicating market categories.

In the dynamic model, Δ_h and q_h are exogenous variables, while β_0^d, β_1^d, γ_M^d, and ε_h^d are endogenous variables.

Part (e). In the static model built by equation 1, there is no time-variate variable, so it is a static model and also linear. Provided that the error term in equation 1 is measured by a probability distribution (e.g. normal distribution) rather than a fixed value, model 1 is a stochastic model.

For the other model in equation 2, it includes a time-dependent variable Δ_h, so it is a dynamic model. It is also linear and stochastic.

Part (f). One of the variables which I consider to be important but is not included in the model is whether the hospital specializes in a certain illness. For example, if hospital h is a cancer center, it is highly possible that its market share in cancer is much larger than other hospitals in the same market area. In this situation, higher market share is attributed to the function of the hospital rather than its treatment quality. If hospital specialty has weak correlation with its quality but attracts a considerable amount of consumers with certain condition, then missing this variable may lead to the underestimation of β_1.

Problem 2. Make your own model

Part (a). I build a logistic model of whether someone decides to get married as below,

$$P(Y_i = 1|X_i) = \frac{e^{\beta_0+\beta_1 age_i+\beta_2 gender_i+\beta_3 income_i+\beta_4 health_i+\beta_5 religion_i}}{1+e^{\beta_0+\beta_1 age_i+\beta_2 gender_i+\beta_3 income_i+\beta_4 health_i+\beta_5 religion_i}}$$

$$Y_i = \begin{cases} 1 \text{ (get married)}, & if\ P(Y_i = 1|X_i) > 0.5 \\ 0 \text{ (not get married)}, & if\ P(Y_i = 1|X_i) \leqslant 0.5, \end{cases}$$

where Y_i is an index of whether the person decides to get married with individual i. The value of it is zero when the decision is not getting married and turns into one under the opposite decision. The probability of getting married is determined by a nonlinear model where age_i is a constant variable meaning the age of individual i. The variable $gender_i$ is a dummy indicating whether i's gender corresponds to the person's sexual orientation. $income_i$ and $health_i$ are both categorical variables which measure the income level and health state of i, respectively. The variable $religion_i$ contains information about the differences in the two person's religions.

Part (b). The dependent variable generated by the model indicates whether the studied person decides to get married with individual i.

Part (c). My model is a complete data generating process. Once given individual's age, gender, income, health state, religion, and corresponding coefficients, the model can simulate the result of whether getting married with i.

Part (d). From my perspective, the key factors in the model are income and health. It is human nature that people want to get married with someone with hither economic status. In addition, a lot of people have fertility desire, so they tend to be careful with choosing spouses in case that their children would have hereditary diseases.

Part (e). Age is included in the model for the reason that people usually prefer to get married with someone with similar ages. $gender_i$ is also an untrivial variable, for it measures that whether i's gender matches the person's sexual orientation. It is the premise of feasible marriage, especially under the context where homosexuality is allowed. Religion is another important factor because for some religions it is forbidden to get married with anyone of other religions. Furthermore, consistency of attitudes towards the world can be reflected by the variable.

Some variables such as education and occupation can also influence marriage decision but are not included in the model. As far as I am concerned, the economic aspect of education and occupation can be captured by income level while gaps in opinions and values generated by education or job differences can be covered by religions.

Part (f). I can firstly run the logistic model with some survey database in which marriage decision, age, gender, income, health, and religion are known. Then I can check the p-value of each parameter to decide whether it is a significant factor in the model. The following prediction can be based on significant factors and their coefficients.