Problem Set #1

MACS 30000, Dr. Rick Evans

Kevin Sun

January 8, 2018

Problem 1: Classify a model from a journal

Part (b). Find a theoretical or statistical model from a recently published article and give a detailed citation of the article.

Elizabeth Hirsh and Youngjoo Cha, "For Law and Markets: Employment Discrimination Lawsuits, Market Performance, and Managerial Diversity," American Journal of Sociology 123, no. 4 (January 2018): 1117-1160.

In this article, the authors estimate the impact of lawsuit resolutions (regarding discrimination) on subsequent changes in managerial sex and race composition of top publicly-traded companies.

Part (c). Write down the mathematical or statistical model.

$$y_{i,t} = x_{i,t}\beta + \alpha_i + \epsilon_{i,t}$$

This article uses a fixed-effects regression model where $y_{i,t}$ refers to the representation of each group of interest in the companies within the study. There are three groups of interest: white women, black women, and black men; the model estimates each groups level of representation at establishment i in year t. $x_{i,t}$ is a vector of time-varying covariates for establishment i in year t. These covariates are listed as the exogenous variables in (part d). β is a vector of the regression coefficients that are estimated. α represents unobserved stable characteristics of establishment i. Finally, ϵ represents the random error term which is assumed to be normally distributed with mean zero and constant variance.

Part (d). List the exogenous and endogenous variables.

Exogenous variables: lawsuit resolution (dummy), monetary award, negative market return, EEOC plaintiff, policy change mandates, media coverage, multiple, plaintiffs, multiple lawsuits, liberal court, federal contractor, and presidential administration.

Endogenous variables: managerial representation of white women, black women, and black men; this is estimated for both one-year and three-year lags.

Part (e). Classify the model.

This is a dynamic, non-linear (log odds), and stochastic model.

Part (f). List variable or feature model is missing that might be valuable.

Social media coverage may be another valuable variable to consider (i.e. shares on Facebook, views on YouTube, retweets on Twitter, etc.)

Problem 2: Make your own model

$$Y^* = Pr(Y + 1|X) = \beta_0 + \beta_1 PSAMGEN + \beta_2 EDUC + \beta_3 CHLDRN + \beta_4 AGE +$$
$$\beta_5 AGEDIFF + \beta_6 RURAL + \beta_7 COHAB + \beta_8 EMPLYD +$$
$$\beta_9 REL + beta_{10} RACE$$

Where PSAMGEN (1 = partner's gender same as own), CHLDRN (1 = couple desires children), RURAL (1 = couple lives in rural area), EMPLYD (1 = someone in the couple is employed), COHAB (1 = couple lives together) are dummy variables; and where, EDUC (education obtained in years), AGE, AGEDIFF (difference in years between partners), REL (one's religion), and RACE (one's race) are discrete variables.

Part (c and f). Make sure your model is a complete data generating process. How could you simulate data from your model given all the parameters and relationships? How could you do a preliminary test of whether your factors are significant in real life?

For this proposed model, data would be collected from two main sources. The first source would be utilizing census data; this would allow us to find the proportions of people in the entire population that are of a certain gender, have certain years of education, of a certain age, live in rural/urban areas, and are employed. From this, one would be able to obtain specific sub-populations (i.e. college-educated, Christian, Asian females who are not employed) marriage rates from census data.

To generate data for demographic information not included a census, a survey could be designed to capture information on variables of interest such as whether one's partners are the same gender as them or whether one lived with their current partner prior to marriage. These types of questions could asked alongside questions that asked respondents to report their marital status, gender, age, religion, and other demographic information that is of interest for the model and also reported in the census.

As a preliminary test, a survey would be administered to ask people to anonymously report their gender, gender of their partner, education, num. of children desired, age, age of their partner, living location, whether they live with their partner, employment status, race, and religion. Additionally, respondents would be asked to report on extent to which each of these factors acts as pressure in their decision to get married (i.e. low, moderate, significant in their decision to get married). Finally, respondents would then report if they intend on getting married as a yes or no response.

The results from the survey could then be run against census data which provides us with information and nearly all cross-sections of the population and their marriage rates.

Part (d and e). What do you think are key factors that influence this outcome? Why did you decide on those factors and not others?

PSAMGEN: Aside from government policy dictating relationships paradigms, social pressures and tolerance of same-sex relationships among one's immediate family, friend circle, and community circle would influence whether one in a same-sex relationship might consider getting married.

EDUC: One's education level and pursuit of more years of education may necessarily delay when they choose to get married. This was chosen in lieu of income given that education levels are highly correlated with income levels.

CHLDRN: A person's decision on whether or not to have children would strongly impact their decision to get married given that marriage happens to be the government-sponsored family structure that incentivizes having children (i.e. tax-breaks, tax-credits).

AGE: As people age the societal pressure to marry tends to increase.

AGEDIFF: The larger the age difference between couples, the greater the pressure may be against marriage.

RURAL: In less populated areas, the pool of potential partners is smaller which incentivizes people to marry earlier versus waiting for or searching for another potential partner.

COHAB: Living together with one's partner may be seen as a logical stepping stone into marriage.

EMPLYD: Employment and financial stability influences whether a couple decides to marry when considering costs of weddings, home ownership - often associated with marriage.

REL: People who practice certain religions are more inclined to marriage than others.

RACE: People of different races face varying pressures to marry depending on cultural, racial norms.