

Problem Set #1

MACS 30100, Dr. Evans

Huanye Liu

Problem 1 Classify a model from a journal.

Part (a) The statistical model introduced in the paper [1] builds a relationship between judicial outcomes and judges' identities in multiethnic societies.

Part (b) The paper *Descriptive Representation and Judicial Outcomes in Multiethnic Societies* [1] is from the American Journal of Political Science. See the References for the detailed citation.

Part (c) The two main regression models are:

$$Pr(y_i = 1|x_{1i}, x_{2i}, X_i, F) = \text{logit}^{-1}(\alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} * x_{2i} + \beta_4 X_i + \gamma_1 F_i) \quad (1)$$

$$p_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} * x_{2i} + \beta_4 X_i + \gamma_1 F_i \quad (2)$$

Part(d) In (1), logit^{-1} is the inverse logistic function used for logistic regression. The endogenous variable is y_i which can represent the leniency or harshness or incarceration depending, and it is a binary variable in all three cases indicating a success or a failure in each case. Leniency is an indicator variable that takes a value of 1 when an appeal against the harshness of a sentence by a defendant has been accepted and a more lenient punishment has been handed down by the district court compared to the magistrate court. Harshness is also an indicator variable, which takes a value of 1 when an appeal by the prosecution against the leniency of the sentence has been accepted and a harsher punishment has been handed down by the district court. Incarceration is a binary variable that equals 0 if the defendant was not sentenced to prison and 1 for all positive prison sentences. All variables on the right side of the equation are exogenous, where x_{1i} is an indicator of the respondent's ethnicity which takes a value of 1 for Arab defendants and 0 for Jewish defendants, and x_{2i} is the mixed-panel indicator that takes a value of 1 if at least one of the panel judges is Arab, and 0 for all-Jewish panels. $x_{1i} * x_{2i}$ is the interaction term and X_i is a vector of all controls including defendant's gender, an indicator of the defendant's past criminal record, the number of female judges, judge age, judicial experience, type of offense, victim's ethnicity, an indicator variables for whether the defendant was convicted in the magistrate court using a plea bargain or trial, and an indicator that takes a value of 1 if the prosecution explicitly requested the court to reject the appeal. In (2), all exogenous variables are the same as in (1), the only difference is the endogenous variable p_i representing the prison term in months.

Part(e) The model represented by (1) is static, nonlinear, stochastic, and the model represented by (2) is static, linear and stochastic.

Part(f) One omitted variable may be the defendant's education level, which may influence decision making from judges of different ethnic group during the defending process in court.

Problem 2 Make our own model

Part(a,b,c) The linear regression model can be represented as the following equation.

$$\begin{aligned} lifespan_i = & \alpha + \beta_1 RACE_{1i} + \beta_2 GENDER_{2i} + \beta_3 BIRTHCOUNTRY_{3i} \\ & + \beta_4 BMI_i + \beta_5 NUMNEGATIVE_i \\ & + \gamma_1 GENRE_i + \gamma_2 INCOME_i \end{aligned} \quad (3)$$

Part(d) From the above equations, the lifespan of a pop musician can be attributed to three aspects. The first aspect relates to the musician's demographic information which I assign three variable to represent: a) the race of musician b) the gender of the musician and c) the country in which the musician was born. The second aspect is about the musician's lifestyle. However, it is hard to directly observe the musician's private life. So I choose two proxies or instrumental variables: a) the BMI index which can be calculated from the musician's weight and height, and b) the number of negative media exposure about the musician, for example, the drug abuse, alcohol abuse, and private life scandals. The third aspect is about the musician's career related information including a) the genre of music which the musician is mainly committed to, and b) the average income level the musician normally earns.

Part(d) As mention in (c), the first aspect is obviously directly related to the musician's lifespan since these are biological traits the musician was born with. The second aspect is about the lifestyle. If the musician tries to live a healthy life, her or his BMI index is likely to be in a normal range and the number of negative exposure in media tends to be low. So these two variables can be instrumental to good estimation of the musician's lifespan. The third aspect is about the musician's occupation. The musician's lifespan may differ with the genre of the music she or he is involved with, and the average income level usually determines the musician's living conditions which may significantly relates to the musician's lifespan.

Part(e) The direct way to test the factors is to collect data of those dead musicians from the web and public media, and then run linear regression on the model (3) to see the significance of all the coefficients(all β s, γ s). Since the data on those chosen variables are relatively easy to collect, this should be a feasible way for a preliminary test.

References

- [1] Guy Grossman, Oren Gazal-Ayal, Samuel D. Pimentel, and Jeremy M. Weinstein. *Descriptive Representation and Judicial Outcomes in Multiethnic Societies*. American Journal of Political Science, Volume 60, Issue 1, pages 44-69, 2016