Problem Set #1

MACS 30000, Dr. Evans
Chelsea Ernhofer

Problem 1

Part A-B

Goldrick-Rab, S., Kelchen, R., Harris, D. N., Benson, J. (2016). Reducing Income Inequality in Educational Attainment: Experimental Evidence on the Impact of Financial Aid on College Completion.American Journal of Sociology,121(6), 1762-1817. doi:10.1086/685442

Part C

$$Y_i = \alpha_{0i} + \alpha_{1i}T_i + \alpha_{2i}X_i + \alpha_{3i}(T_i x(X_i) + \alpha_{4i}C_i + \varepsilon_i$$

Part D

In the paper I read, this equation is used to determine two different endogenous variables (both referenced by Y). The first is the binary variable indicating whether or not a student graduated from college. The second is the number of courses completed by the student. Exogenous variables include T, whether or not the student recieved the treatment (added financial aid), X, out-of-pocket costs (as opposed to costs already covered by a previous scholarship, grant, or loan). C represents college fixed-effects and is used to normalize data across multiple public colleges. D is a representation of demographic variables of interest including student race, age, sex, and family income.

Part E

This model is linear because it is used in OLS regression and logistic regression, both of which, in this case, are used to find linear relationships. The model is dynamic because variables are taken from different times in a student's college career, and these variables effect each other over time. The model is stochastic because there is an error term included, suggesting that researchers are not attempting to come to a definite solution.

Part F

One variable that could add meaning to this model is student major. I believe that students enrolled in majors that are historically more challenging (such as engineering) could behave in a systematically different way than students in, for example, the liberal arts. This could, in turn, influence their reactions to added funds and their likelyhood of dropping out or continuing on in their college careers.

Problem 2

Part A-B-C

$$L_i = \alpha_{0i} + \alpha_{1i}O_i + \alpha_{2i}FH_i + \alpha_{4i}F_i + \alpha_{4i}PH_i + \alpha_{4i}A_i + S_i + \varepsilon_i$$

Where L represents predicted lifespan of artist, O represents place of origin or ethnicity, FH represents family medical history, F represents physical fitness, PH represents personal medical history, A represents a continuous pattern of smoking or using other harmful drugs. S represents age at which artist had their first hit.

Part D

Some key factors that I believe would influence lifespan of a musician are place of origin, since persons of different ethnicities are proven to have distinct expected lifespans and family medical history and specifically whether there is a history of heart disease or cancer in the family. Other key factors include personal medical history, such as past heart surgeries or cancers, physical fitness, and any harmful addictions an individual may have.

Part E

I did some research concerning expected lifespans and included the majority of the variables I found through that research. Some indicators, such as genetic dispositions, were extremely complicated so I simplified them down to measures of medical history for the indiviudal and their family. The only factor that is unique to the music industry is the age at which the artist began to experience real success. This is because I believe that, although a musician's lifestyle may be distinct from that of a regular individual, I am not convinced that this difference would have a large influence over lifespan. That is stated with the exception of drugs taken, which is already included in my model. It would be interesting, however, to see if the length of time which an artist spends as a successful musician has an effect on their lifespan. If a significant relationship was found, I would then create an updated model to include more indicators unique to the music industry.

Part F

In order to run a preliminary test and determine whether or not the factors I have chosen are significant, I could simply use this model to 'predict' the lifespans or musicians who have already died, and determine whether there is, indeed, a pattern among those individuals.