Problem Set #1

MACS 30100, Dr. Evans
Julian McClellan

Problem 1

Part (a-b). Juhn, Chinhui and Kristin McCue. 2016. "Evolution of the Marriage Earnings Gap for Women." American Economic Review, 106(5): 252-56.

Part (c). $lnY_{it} = \beta X_{it} + \gamma M_{it} + \pi K_{it} + \delta_c + \varepsilon_{it}$ $\varepsilon_{it} = \alpha_i + v_{it}$

i indexes the individual, X are observable characteristics such as age, education, race, ethnicity, M is the married dummy, K are indicators for the presence and age of children, δ_c refers to birth cohort effects, and lnY_{it} refers to the natural log of earnings.

Part (d). The only endogenous variable in the model is lnY_{it}, the natural log of the earnings. Because this variable can easily be transformed to simply be the earnings, we can also say that the earnings themselves are the endogenous variable in the model. (It is what the model is estimating, the response variable.)

The exogenous variables are simply the explanatory variables in the ordinary least squares regression. These are X, the observable characteristics such as age, education, race, ethnicity, M, the married dummy, K, the indicators for the presence and age of children, and δ_c, the birth cohort effects.

Part (e). The model is static, even though the model is indexed over time, each time period is kept distinct from the other periods.

The model is linear because it is an ordinary least squares regression model.

The model is stochastic because it includes an error term to introduce an element of randomness to the model.

Part (f). A useful variable to include in the model would be the number of hours worked during each time period since with it one could separately examine the relative contributions of wages and hours.

Problem 2

Part (a-c). $predicted\,lifespan_{musician} = \beta_0 + \beta_1 expectancy_{demographic\,info} + \beta_2 genre_1 + ... + \beta_{2+I} genre_I + \beta_3 num\,rehab + \epsilon_{musician}$

Part (d). The key factors are essentially all the ones that I included in my model. However I believe some are more important than others.

I think that the most important factor that influences the outcome of how long popular musicians live is simply the expected lifespans of people matching the musician's demographic information. Taking the birth year of the musician, and things like their ethnicity and state of birth can be used to determine the expected lifespan of people matching those demographic characteristics.

Of course, drug overdoses can also be a key factor in determining (terminating) a musician's lifespan, and an element of drug use in general, the number of rehab visits is thrown into the model as a weak proxy to drug usage intensity.

Lastly, the differing genres of music musicians reside in often are indicative of different lifestyles for those musicians and indicator variables for the genre of music are included in my model.

Part (e). I decided on these factors and not others because, firstly, they seemed the most feasible to gather. Demographic information can easily be gathered from a popular musician's Wikipedia page and be used to determine the expected lifespan for people born in that year and with other similar characteristics.

The music genre(s) can easily be gathered from the Wikipedia page as well. The only factor that I chose that might be slightly more difficult to pull from the musician's Wikipedia page is the number of times he or she went to rehab. Assuming the musician was popular enough and that this information was public, this ought to be contained on the page in some natural language form, and thus getting an accurate count would involve some sort of natural language parsing. However, it is possible that the musician was able to keep rehab visits private and/or they are not popular enough to have an exhaustive page to include this information.

Part (f-g). A preliminary test of these factors would probably involve scraping Wikipedia pages of dead popular musicians in order to acquire the genre of music they were active in and their demographic information. For each unique country of origin, I would also have to look up some sort of life expectancy tables for that country and be able to match the year born with things like ethnicity to determine the life expectancy of the musician (the explanatory variable, not predicted lifespan, the response variable).

A preliminary test of the number of rehab visits factor would be much more difficult. It would be much easier to simply eschew this variable in my "preliminary" test, but I suppose I could attempt to utilize some of the things I will learn in James Evans's Content Analysis course to try and build a parser to detect how many times a musician has visited rehab.