Problem Set #1

MACS 30100, Dr. Evans
Cheng Yee Lim
12145076

Problem 1: Model Classification

(a)(b)

Stevenson, Betsey, and Justin Wolfers. "Subjective well-being and income: Is there any evidence of satiation?." *The American Economic Review 103, no. 3* (2013): 598-604.

(c)

In Stevenson, Betsey and Wolfer's paper, they investigate whether a satiation point in subjective well-being exists, more specifically if the relationship between income and well-being is different beyond a critical income level. They do so with two models:

$$Well - being_c = \alpha + \beta log(GDP_c) + \varepsilon_c \tag{1}$$

The exogenous variable is the log of the GDP of the country and the endogenous variable is the measured subjective well-being of the country.

$$\begin{aligned} Well - being_c = \alpha + \beta_{poor} I(GDP_c < k) \times (log(GDP)_c) - log(k)) + \\ + \beta_{rich} I(GDP_c >= k) \times (log(GDP)_c) - log(k)) - \varepsilon_c, \end{aligned} \tag{2}$$

where the subscript c denotes country, and the independent variables are the interactions of a dummy variable, categorizing the country's GDP per capita as above or below the a critical point ($k), and log real GDP per capita. The coefficient β_{poor} refers to the gradient of well-being and GDP of "poor" countries (those with GDP $< \$k$), and β_{rich} is the gradient of "rich" countries (those with GDP $\geqslant \$k$).

In model 2, the endogenous variable is also the subjective well-being of the country and the exogenous variables include the log GDP of the country, the dummy variable categorizing whether the country is below or above the critical level of GDP, and the interaction of the former two variables.

(d)

Both model 1 and 2 are static, non-linear, and deterministic models. In model 1, the logarithmic of the GDP of the country allows for a non-linear relationship between well-being and GDP. Similarly, the functional form in model 2 allows for a kink in the model at $k, as the gradients for the "poor" and "rich" countries could be different. Both model 1 and 2 are static models, as they do not have time dimensions. Both models are also deterministic, as there is no randomness involved.

(e)
I would include a variable that denotes government expenditure per capita. The larger the government expenditure per capita, the closer a country is to a welfare state. Welfare states, such as Denmark and Norway, are known for their high happiness levels in their countries. Thus, government expenditure per capita is a valuable variable that has been left out in the model explaining happiness of a country.

Problem 2: Modeling the Lifespan of Popular Musicians

(a)(b)(c)

$$predicted\_lifespan_i = \beta_0 + \beta_1 income_i + \beta_2 smoking_i + \beta_3 drugs_i + \phi X_i + \varepsilon_i \quad (3)$$

Where the dependent variable, $predicted\_lifespan_i$, is a measure of the length of the lifespan of the musician in years. $Smoking_i$ and $drugs_i$ are categorical variables reflecting the extent of smoking of the musician per week and history of substance intake of the musician respectively. X_i is a vector of individual characteristics that include log of real income, gender, marital status, lifestyle habits, such as dietary preferences and frequency of exercising, presence of genetic diseases in family medical history.

(d)(e)
In our model, I hypothesize that $income_i$ will play a more significant role in influencing this outcome compared to an average individual. This is because income reflects not only the individuals ability to consume a healthy diet and have better access to healthcare, but also the musicians popularity. A more popular musician will often have a higher income from their high album and concert sales, but would also be under greater stress to maintain his or her popularity. Thus, the income of a musician can also be a proxy for the stress levels of a musician and I would expect the coefficient of income of the musician to be less positive than the same coefficient for the average individual in a population.

Other key factors will include the categorical variables, $smoking_i$, which reflects the range of the number of cigarettes smoked per week, and $drugs_i$ reflecting the history substance abuse of the musician in the model. According to newspaper reportings and past literature, musicians are more likely to abuse drugs or smoke than an average individual. In fact, the top cause of death amongst performers is lung cancer. Researchers have attributed this to the high stress they are under and the dependence on popularity in their job.

Other factors, such as gender, presence of genetic diseases in family medical history, exercise frequency, and dietary preferences are also factors that affect the lifespan of individuals. These variables are often highly correlated with health, but are not specific in affecting the lifespan of popular musicians. Thus, they will be included as control variables in the regression model instead.

(f)

I would conduct the preliminary test whether the factors are significant in real life with a regression analysis on the dataset, Trends in Life Expectancy by Social Class 1972-2005, from the United Kingdom Office of National Statistics. The regression will be estimated with ordinary least squares with robust standard errors.

However, there is multicollinearity in the regression as income will also allow individuals to have better diets. To tackle the problem of multicollinearity of the income variable, I perform instrumental variable (IV) regression. I intend to use the proportion of times the musician has entered the weekly billboard top hits over the length of their music career as a proxy of income. This instrument is relevant to income as high billboard rankings often means more album sales and higher popularity. The instrument is also exogenous as the musician does not decide if he or she enters the billboard.