Problem Set #1

MACS 30100, Dr. Evans
Dongping Zhang

Problem 1

Part (a). The model I found interesting is from the article *Parenthood and Happiness: Effects of Work-Family Reconciliation Policies in 22 OECD Countries*, published in the *American Journal of Sociology*. This model is constructed to assess the effects of family and child-care public policies on cross-national emotional well-beings of parenthoods by predicting self-reported happiness scales by parents and non-parents in 22 OECD countries.

Part (b). The detailed citation of the article:
Glass, J., Simon, R. W., & Andersson, M. A. (2016). Parenthood and Happiness: Effects of Work-Family Reconciliation Policies in 22 OECD Countries 1. *American Journal of Sociology, 122*(3), 886-929.

Part (c). The model used in the article is a multilevel model. The fixed-effects procedure models country-level differences in the effect of parenthood on happiness and the net of sociodemographic variables. The mixed-effects procedure examines whether parenthood might have differing effects on happiness dependent on country-level policy contexts.
1. Fixed Effect Procedure: dependent variable is regressed on individual sociodemographic control variables with a variable intercept for each country (the U.S. is the reference category):

$$Y_{ij} = \beta_{0j} + \beta_{1j}\mathbf{X}_{1ij} + \beta_{2j}\mathbf{Z}_{ij} + \mathbf{E}_{ij}$$
$$\beta_{0j} = \lambda_0 + \lambda_1\mathbf{X}_{1j}$$
$$\beta_{1j} = \alpha_0 + \alpha_1\mathbf{X}_{1j}$$

2. Mixed Effect Procedure: dependent variable is regressed on policy and economic variables (country level) and sociodemographic control (individual level) using a mixed-effects multilevel estimation procedure:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\mathbf{X}_{1ij} + \beta_{2j}\mathbf{Z}_{ij} + \mathbf{E}_{ij}$$
$$\beta_{0j} = \lambda_0 + \lambda_1\mathbf{X}_{1j}$$
$$\beta_{1j} = \alpha_0 + \alpha_2\mathbf{X}_{2j} + \alpha_3\mathbf{X}_{3j} + \mathbf{U}_j$$

Part (d). Variable types:

- Endogenous Variable:
  - Y_{ij} = happiness of individual i in country j

- Exogenous Variables
  - $\mathbf{X}_{1ij}$ = a binary variable indicating parental status (1 if parent)
  - $\mathbf{Z}_{ij}$ = vectors of individual attributes such as age, gender, education, income decile, and employment status
  - $\mathbf{E}_{ij}$ = the error term of individual i in country j
  - $\mathbf{X}_{1j}$ = vector of variables representing each country in the analysis set with U.S. as referent
  - $\mathbf{X}_{2j}$ = vector of family policy variables
  - $\mathbf{X}_{3j}$ = vector of country-level variables
  - $\mathbf{U}_j$ = the error term in the mixed-effect procedure

Part (e). Model Types: this model is static, linear, and stochastic.

- static: there is no time dimension in the model
- linear: both procedures of the model are in linear format
- stochastic: this model has an error term to account for noise

Part (f). I think the model could potentially be more powerful if it could include a vector of characteristic variables of children if the household has children, or $\mathbf{X}_{1ij} = 1$ of any given i. From my perspective, characteristic variables of children in a household, such as ages, genders, or even the numbers of children, could be valuable regressors to assess the mental well-being of parents.

Problem 2

Part (a). My proposed model of how long popular musicians can live is:

$$Y_i = \beta_0 + \beta_{1j}\mathbf{M}_{ij} + \beta_{2k}\mathbf{P}_{ik} + \beta_{3l}\mathbf{C}_{il} + \epsilon_i$$

- Y_i = predicted lifespan (in years) of a particular musician i
- $\mathbf{M}_{ij}$ = a vector of Career variables of a musician i, in which j could be:
  - music genre (categorical)
  - ratio of debut year over birth year
  - number of song productions
  - number of music videos
  - number of awards nomination(s) and awards winning(s)
  - number of collaborations with other artists
- $\mathbf{P}_{ik}$ = a vector of Popularity variables of a musician i, in which k could be:
  - number of google results
  - number of total views of music videos
  - average number of daily occurrences in news and media since debut
- $\mathbf{C}_{il}$ = a vector of Personal characteristic variables of a musician i, in which l could be:
  - gender (binary)
  - height
  - weight
  - smoking (binary)
  - education attainment (categorical)
  - sexuality (categorical)
  - number of marriages
  - number of children

Part (b). In my proposed model, Y_i, which is the predicted lifespan (in years), is my endogenous variables while all other variables in vectors $\mathbf{M}_{ij}$, $\mathbf{P}_{ik}$, $\mathbf{C}_{il}$ are exogenous.

Part (c). Given all needed parameters, β_{1j} given i, β_{2k} given k, and β_{3l} given l, the model would be able to complete a data generating process producing simulated data of demand.

Part (d). As presented in my model, I consider three aspects to be important to predict the lifespan of a musician, and variables associated with these three aspects are classified into three vectors: Career, Popularity, and Personal characteristics, among which I consider the most influential vector to be Personal characteristic variables $\mathbf{C}_{ij}$.

Part (e). The reason I choose my regressors listed above when constructing my model is because I think it is necessary and crucial to make all my independent variables accessible or observable when collecting data. Realistically, as a researchers, we would not be able to access the musicians' medical records, tax returns, or daily lifestyles. Thus, I have to guarantee all my exogenous variables could be feasibly obtained from my limited data sources: internet and media releases.

Part (f). To test whether my variables are significant in real life, I would randomly select a sample of musicians and to conduct a regression analysis based on data collected so as to assess variable significance. From my perspective, I think my current model could potentially face the issue of multicolinearity. For example, the two variables in my Career vector: the number of song productions and the number of music videos, and the one variable in my Popularity vector: the number of total views of music videos, are highly likely to be linearly correlated to each other. Thus, in order to test whether these variables or some other factors are necessary or redundant, I would implement a Principle Component Analysis (PCA) trying reduce the dimension of my model.