

## Problem Set 1

MACS 30100, Dr. Evans

Tong Ju

**Problem 1** Classify a model from a journal

### Part (a-b). Find an article and list the detailed citation

David Jacobs and Jonathan C. Dirlam. (2016). *Politics and Economic Stratification: Power Resources and Income Inequality in the United States*. American Journal of Sociology, 122(2): 469-500  
<http://www.journals.uchicago.edu.proxy.uchicago.edu/doi/abs/10.1086/687744>

### Part (c). Statistical model

This article aims to find out what factors can be used to best explain the U.S. growth in economy stratification since the late 1970s. The authors found the increase in strength of neo-liberal national administrations and skill-biased technical change are the most influential determinants. The income inequality in a specific state in a given year is modeled as follows:

$$\begin{aligned} \text{Inequality}_t = & \mathbf{b}_0 + \mathbf{b}_1 \text{RepublicanGovernor}_{t-2} + \mathbf{b}_2 \text{RepublicanStateLegislature}_{t-2} \\ & + \mathbf{b}_3 \text{RepublicanPresident}_{t-2} + \mathbf{b}_4 \text{PresidentialElection}_{t-1} + \mathbf{b}_5 \% \text{Union}_{t-1} \\ & + \mathbf{b}_6 \% \text{Employed in Finance}_{t-1} + \mathbf{b}_7 \% \text{Black}_{t-1} + \mathbf{b}_8 (\% \text{Black} \times \text{Year})_{t-1} + \mathbf{b}_9 \% \text{Hispanic}_{t-1} \\ & + \mathbf{b}_{10} \text{IncomePerCapita}_{t-1} + \mathbf{b}_{11} \text{IncomePerCapita}^2_{t-1} + \mathbf{b}_{12} \% \text{College}_{t-1} \\ & + \mathbf{b}_{13} (\% \text{College} \times \text{Year})_{t-1} + \mathbf{b}_{14} \text{Density}_{t-1} + \mathbf{b}_{15} \text{Density}^2_{t-1} + \mathbf{b}_{16} \text{Poverty}_{t-1} \\ & + \mathbf{b}_{17} \text{RuralEmployment}_{t-1} + \mathbf{b}_{18} \% \text{WomenEmployment}_{t-1} \\ & + \mathbf{b}_{19} \% \text{Employed in Manufacturing}_{t-1} + \mathbf{b}_{20} \% \text{StockValue}_{t-1} + \mathbf{b}_{21} \% \text{Unemployed}_{t-1} \\ & + \mathbf{b}_{22} (\% \text{Unemployed} \times \text{Year})_{t-1} + \mathbf{b}_{23} \% \text{MarginalTaxRate}_{t-1} \\ & + \mathbf{b}_{24-73} (49 \text{StateTrends}) + \mathbf{b}_{74-123} (49 \text{StateTrends}^2) + \epsilon \end{aligned}$$

### Part (d). Endogenous and Exogenous variables

- Endogenous variables: Income Inequality.
- Exogenous variables: Year, Republican governor (0 or 1), Republican State Legislature (0 or 1), Republican President (0 or 1), Presidential Election Year (0 or 1), % Union, % Employed in Finance, % Black, % Hispanic, Real personal income per capita, % College, Density, % below poverty line, % Women Employment, % Employed in Manufacturing, SP 500 stock value, % Unemployed, Variance marginal tax rate, State trends (for 49 states).

### Part (e). Classify the model

- Dynamic: This model includes Year as a dependent variable, and it describes a dynamic relationship between the income inequality in a given year with all the exogenous variables 1 or 2 years before that year.
- Linear: OLS regression assumes a linear dependent and independent variable relationship.
- Stochastic: In this model, the author assumes an error term following  $N(0, \sigma^2)$ .

### Part (f). Possible variable of feature

The model proposed by this article is rather inclusive, which covers most of the exogenous variables essential for the income inequality. However, Besides these internal factors (*e.g.*, national administration, Skill-biased technological change, race and gender disparities and tax policies), there are still other factors that may contribute to the inequality, such as immigration and corporatism. The *Immigration and Nationality Act* of 1965 increased immigration to America. As a result, since last 4 decades, the foreign-born proportion of America's population grew from 5% to 11%, most of whom had lower education levels and incomes than native-born Americans. We can imagine that the flux

of immigration into U.S. also has certain influence on the increase of income inequality. So I suggest the each year's immigration population might be also taken into consideration. Secondly, If we assume the income inequality is not resulted from a free market, but the rise of corporatism (*MassFlourishing*, Edmund Phelps, 2010). We can measure the labor's share of GDP (labor costs over nominal GDP) and examine the relationship between it with the income inequality.

## **Problem 2** Model for life span of popular musicians

### **Part (a-c). Model for popular musicians live**

This is my model estimating popular musicians life span:

$$\text{PredictedLifespan} = \text{Average}_{\text{BirthYear}} + \alpha_1(\text{Substance}) + \alpha_2(\text{Mental}) + \alpha_3(\text{Health}) + \alpha_4(\text{Genre}) + \alpha_5(\text{Marital}) + \alpha_6(\text{Education}) + \alpha_7(\text{Gender}) + \alpha_8(\text{Ethnic}) + \epsilon$$

Defining the variables:

- *Average<sub>BirthYear</sub>* is the average life span of the population of those who were born on the same year of the popular musician. This variable serves as a reference to predict the life span of the musician;
- *Health Factors* include variable *Substance* (binary variable, 0 or 1, indicating whether the musician is a drug consumer in his or her life), *Mental* (binary variable, 0 or 1, indicating whether the musician suffers any mental disease, such as depression, bipolar or other metal disorder, reported during his or her life time), and *Heath* (binary variable, 0 or 1, indicating whether the musician suffers any physical disease, such as diabetes, cancer, AIDS or others, reported in his or her life);
- *Genre Variable* is a categorical variable, which denotes the major music genre that the musician engaged in. *Genre* has the following categories: Blues, World, Country, Gospel, Hip-Hop, Jazz, Metal, Pop, Punk, RB, Rock and Others;
- *Social Factors* includes *Marital* (categorical variable, which indicates marital status of the musician: single, married, divorced, and Others), and *Education* (continuous variable, which is the education year that the musician have taken over his or her life) ;
- *Demographic Factors* includes *Ethnic* (categorical variable: Black, White, Hispanic and Asian), and *Gender* (binary variable. 0 for male, 1 for female).

### **Part (d). Key factors**

First, I hypothesize that Health Factors (mental or physical health as well as the consumption of drugs) may have a heavy impact on the life span of the popular musicians. Besides that, I conjecture the music genre also has certain influence on the life style and vulnerability of musicians.

### **Part (e). Rationale of Choosing Independent Variables**

In this model, I consider several factors into consideration. First and most obvious one, the health factor will have significant effect on the life span of musicians. Especially, it is not abnormal to find lots of news reports on the death of young musicians due to the excessive drug use or mental illness. Secondly, according to Dianna Theadora Kenny's study ([Life expectancy and cause of death in popular musicians, Medical problems of performing artists 3\(1\):37-44, 2016](#) ), there is a strong association between the music genre and life span of musicians. She found that musicians who are dying youngest belong to newer genres (electronic, punk, metal, rap, hip-hop) that have not existed as long as genres such as jazz, country, gospel and blues. Thirdly, besides health factor and music genre, I suppose the social factor, such as education years and marital status, may be relevant in the longevity of musicians. Since, the marital status and education level may be considered to reflect the social integration and formality of a person, I hypothesize that the musician with a stable martial status and longer education year will have longer life span than other musicians.

#### Part (f). Preliminary Test for Hypotheses

- *Data Collection* Besides Wikipedia, there are several other available sources, such as [The Dead Rock Star's Club](#), [Voices from Dark Side](#), [Dead Punk Star](#), Gerard Herzhaft's *Encyclopedia of the Blues*, and so on, which list the life profile and death cause of the passed-away popular musicians.
- *Creating Dataframe* by scraping the on-line information of above resources, we can establish a data frame including the age of the dead musicians, cause of death, health profile, music genre, gender, education year, marital status and so on.
- *Data Exploration* In the third step, we can examine the hypothesis proposed above, to test the relationship between the life span with the health status, social and demographic factors. Since I propose a linear model, OLS regression can be deployed to analyze the data.
- *Optimization of models* Finally, according to the initial analysis of data, we can figure out which factors have strong and significant association with the life span, and thus optimize the model by reducing the irrelevant factors.