STAT3503A – Course Assignment

Estimating Canada's Yearly Labor Force based on current Population Measurements and Economic Growth Factors

Proposal Stage

Vee Nguyen

101093925

November 6th, 2020

School of Mathematics and Statistics

Carleton University

Motivation

According to the standard definition employed by Statistics Canada, the employed are persons having a job or business, whereas the unemployed are without work, are available for work, and are actively seeking work. Together the unemployed and the employed constitute the labor force, which reflects the country's economy growth and can be used to explain various social movements and issues. Statistics Canada believes the ratio of people in the labor force and people who is economically inactive can be used as a planning tool for decision making. Some noble uses of this number include developing government's policy to create balance of some programs such as public pension and changing the distribution of fund to social programs which focus on knowledge transfer, immigrant integration and employment equity.

The OpenGov has published and ongoing updated various datasets regarding to labor force survey, marital status distribution, and Gross Domestic Product (GDP) from past years. The first two datasets reflect how population is distributed across provinces and provide insight into the characteristics of population like gender and age distribution, educational levels, and employment statuses. On another hand, by definition, GDP is the final value of the goods and services produced within the geographic boundaries of a country within a period, which is normally a year. GDP growth rate is an important indicator of the economic performance of a country. The assumption I make for this project is the economic growth positively proportional to the number of job availability of the country. The main goal of this project is to examine the number using various statistical analysis to answer question: "How can we estimate labor force efficiently?"

Outline

I believe exploring that information will be helpful in extracting the attributes to predict Canada labor force and examining the relationship between attributes. There are many models can be built based on different attribute combination, and our goal is to find the best model that fit whatever we are exploring. In this project, we will define the best model to be the one with minimal Residual Sum of Square (SSE)

My hypothesis at this point is to the model of labor force estimator includes variable related to age, gender, education level distribution, change in marital status and change in GDP per year, where certain group age, marital status, education level plays a significant role in the final model.

Dataset Description

Each unprocessed dataset provides enormous auxiliary information, so it will require a lot of work to extract the features. Hence, the data cleaning process will be not discussed in this proposal.

a. **Unprocessed Dataset**. This dataset includes Number of persons in the labor force (employment and unemployment) and not in the labor force, unemployment rate, participation rate, and employment rate, by educational degree, sex, and age group, last 5 years. Each row contains the year, geography, labor force characteristics (Population, Employment either Full Time and Part Time, Unemployment…), educational degree (no degree/diploma, without/with high school graduation, post-secondary level and higher than post-secondary level), gender (male and female), age group (15-24, 25 and over, 25-54, 55-64), and the number within population corresponded to the declare statuses. The dataset can be retrieved using:

```
library(tidyverse)
temp <- tempfile()
download.file("https://www150.statcan.gc.ca/n1/tbl/csv/14100118-eng.zip",temp)
# download as a temporary file
(file_list <- as.character(unzip (temp, list = TRUE) $Name)) # unzip the file
population <- read_csv(unz(temp, "14100118.csv"))
unlink(temp) # Delete temporary file
population %>% glimpse
```

b. **Unprocessed Raw Martial Status Dataset**. This table contains 43650 series, with data for years 1971 - 2007 (not all combinations necessarily have data for all years). This table contains data described by the following dimensions (Not all combinations are available): Geography (15 items: Canada; Nova Scotia; Prince Edward Island; Newfoundland and Labrador ...) Type of marital status (2 items: Marital status; Legal marital status ...) Marital status (5 items: Married; Widowed; All marital statuses; Single; never married ...) Sex (3 items: Both sexes; Males; Females ...) Age group (97 items: All ages; U 15 years; 0 to 14 years; 15 years and over ...). The dataset can be retrieve using:

```
temp <- tempfile()
download.file("https://www150.statcan.gc.ca/n1/tbl/csv/17100013-eng.zip",temp)
# download as a temporary file
(file_list <- as.character(unzip(temp, list = TRUE)$Name)) # unzip the file
martial <- read_csv(unz(temp, "17100013.csv"))
unlink(temp) # Delete temporary file
martial %>% glimpse
```

c. **Unprocessed GDP**. Gross Domestic Product (GDP) at basic prices, by North American Industry Classification System (NAICS) aggregates, by Industry, volume measures, monthly, 5 most recent time periods. The dataset can be using:

```
temp <- tempfile()
download.file("https://www150.statcan.gc.ca/n1/tbl/csv/36100434-eng.zip",temp)
# download as a temporary file
(file_list <- as.character(unzip(temp, list = TRUE)$Name)) # unzip the file
GDP <- read_csv(unz(temp, "36100434.csv"))
unlink(temp) # Delete temporary file
GDP %>% glimpse
```

Proposed Variables

Independent Variable

1. Year
2. Geography (by province)
3. Change in population
4. Number of males within the population
5. Number of females within the population

6. Number of people without any diploma
7. Number of people with high school diploma
8. Number of people with post-education or higher degree
9. Number of young adult (15-24 year old)
10. Number of mature adult (25 – 55 year old)
11. Number of senior (55 and up)
12. Number single people/never marriage
13. Number of marriage people
14. Number of divorced people without re-marriage
15. Change in GDP

Dependent Variable

Number of people in labor force