Misty Ann Stone

Exercise 6.1 Sourcing Open Data

Data Source

Dependent Variable

I am planning to use an Economic Freedom index published by the Fraser Institute for my dependent variable. The Fraser Institute (<https://www.fraserinstitute.org>) is a research and educational organization that studies and measures the effects of government policies, entrepreneurship, and choice. The Fraser Institute datasets are used by many academic disciplines including development economists.

The Economic Freedom index measures the economic freedom of individuals – their ability to make their own economic decisions. Economic Freedom involves personal choice, voluntary exchange coordinated by markets, freedom to enter and compete in markets, and protection of persons and their property from aggression by others. In other words, an individual has economic freedom when their property is protected by the government and they can easily and voluntarily participate in markets. In economically free countries, the government’s primary role is protecting personal property from aggression by others. A highly ranked country’s government and systems will protect property, have a fair justice system, have even enforcement of contracts, and a stable monetary environment. Taxes will be relatively low, there will be relatively few barriers to domestic and international trade, and the allocation of goods and resources will primarily be performed by markets rather than government spending and regulation.

Economists are very interested in measures such as the Fraser Institute’s Economic Freedom index because economic freedom is a one of the most important determinants of an economy’s growth. A small (per capita) economy is much, much more likely to grow when people have economic freedom. Many studies have shown that Economic Freedom is the key determinant when a relatively poor country begins to step up the poverty ladder and its GDP per capita begins to grow.

The Fraser Institute studies and indexes several components such as size of government, legal systems and personal property rights, monetary stability, freedom of trade, and market regulations. Each country (or state, or province) receives a weighted rating in each area, thus giving the index. Countries such as Singapore, Hong Kong (SAR – separately from China), Switzerland, and New Zealand tend to have the highest Economic Freedom ratings. Countries such as Venezuela, Zimbabwe, Syria, Sudan, Yemen, and Iran tend to have the lowest Economic Freedom ratings.

The data set includes data from 1970-2021 and includes 165 countries. The Fraser Institute includes indexes for many of the components of Economic Freedom, but I will focus only on the final Economic Freedom Index. This data is readily available for research on the Fraser Institute’s website.

Independent Variables

I will be using several independent variables from the United Nations website (<http://data.un.org>).

* GDP (Gross Domestic Product) refers to the general size of a country’s economy.
* Population variables, including some gender separated.
* CPI (Consumer Price Index) is a measure of inflation.
* Labor Force data includes unemployment as well as labor force participation.
* Education includes enrollment data.

I also need three categorical variables.

* I found a list of countries and their continents on Statistics Times, which cites the UN Statistics Division.
* The World Bank categorizes economies by income: High, Upper-Middle, Lower-Middle, and Low based on GNI (Gross National Income).
* I created a variable that equals 1 if the country is landlocked and 0 otherwise as well as a variable that equals 1 if the country is an LDN (Least Developed Nation, as defined by the UN based on the net income per capita of a nation. (This data is from UNCTAD, UN Conference on Trade and Development, <https://unctad.org/topic/landlocked-developing-countries/list-of-LLDCs>. )
* I created a variable that equals 1 if the country has English as an official language or if it is widely spoken. This data came from the CIA World Factbook. (<https://web.archive.org/web/20070613004519/https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html>)
* I created a variable called “Popularity” to approximate how well-known each country is in the US. To estimate the “popularity” of each country, I created a list of all of the countries in my study and re-sorted them randomly. I then gave my list to my 14 year old daughter and asked her to highlight the names that are countries, without telling her that it was a list of all of the countries. The countries that she knew were actual countries are indexed “1” and the countries that she did not believe were actual countries are indexed “0”. In fairness, this may not be representative of an “average” 14 year old because my family has lived in Abu Dhabi and has traveled extensively in the southern Middle East and in Western Europe. Nonetheless, it is a good approximation of how well-known each country is in the US.

I believe that each of these independent variables may affect the Economic Freedom Index. I did include a surface area variable in the population data set which may not have an effect on Economic Freedom.

Data Cleaning and Exploratory Data Analysis

First, I combined my data into a single excel file, deleting the columns that were unnecessary.

* Economic Freedom columns: year, country, Economic Freedom Summary Index, Rank, Quartile. This data set includes the years 1970, 1975, 1980, 1985, 1990, 1995, and yearly from 2000 up to and including 2021.

\* Note that the UN datasets include aggregate measures of the variables of interest – such as by continent (eg, Europe and North America) as well as by region (eg, Polynesia and Western Africa). I deleted all of these as they are not relevant to my current project.

* The UN GDP data set includes several measures of GDP for each country. I kept “GDP in constant 2010 prices” (inflation adjusted) and “GDP per capita”. This data set includes the years 1995, 2005, 2010, 2015, 2018, 2019, and 2020.
* The UN Population data set includes total population, population by gender, percentage of population aged 0-14, percentage of population aged 60+, as well as population density and surface area of the country. This data set includes the years 2010, 2015, 2017, 2020, and 2022.
* Consumer Price Index is a measure of inflation. The UN CPI data set includes a general CPI as well as a food-specific CPI for each country. These data are presented with base year 2010 – so, the price index is 100 in year 2010 for each country. An index greater than 100 indicates higher prices than in 2010, and an index lower than 100 indicates lower prices than in 2010. This data set includes the years 2000, 2005, 2010, 2015, 2019, 2020, and 2021.
* The UN unemployment data set includes the unemployment rate, total as well as separated by male and female; and labor force participation (the percentage of adults who are employed or unemployed – excluding the individuals who are not employed or looking for a job), total as well as separated by male and female. There is data for years 2000-2022, but not every year is included and there is a lot of missing data. I will probably use a weighted average for this data set.
* The UN Education data set includes the Gross Enrollment Ratio for Primary, Secondary, and Upper Secondary for each males and females. This data set includes the years 2000-2021.

I adjusted the name of some of the countries for consistency:

* “Bahamas, The” to “Bahamas”
* “Bolivia (Plurin. State of)” to “Bolivia”
* “Egypt, Arab Rep.” to “Egypt”
* “Gambia, The” to “Gambia”
* “China, Hong Kong SAR” to “Hong Kong”
* “Hong Kong SAR, China” to “Hong Kong”
* “Iran, Islamic Rep.” to “Iran”
* “Iran (Islamic Republic of)” to “Iran”
* “Lao PDR” to “Lao”
* “Lao People's Dem. Rep.” to “Lao”
* “Lao People's Democratic Republic” to “Lao”
* “Kyrgyz Republic” to “Kyrgyzstan”
* “Korea, Rep.” to “Korea”
* “Republic of Korea” to “Korea”
* “Republic of Moldova” to “Moldova”
* “Slovak Republic” to “Slovakia”
* “Syrian Arab Republic” to “Syria”
* “United Rep. of Tanzania” to “Tanzania”
* “United States of America” to “United States”
* “United States of America of America” to “United States”
* “Venezuela, RB” to “Venezuela”
* “Venezuela (Boliv. Rep. of)” to “Venezuela”
* “Viet Nam” to “Vietnam”
* “Yemen, Rep.” to “Yemen”
* “Dem. Rep. of the Congo” to “Congo, Dem. Rep.”
* “ Congo” to “Congo, Rep.”
* “Dem. People's Rep. Korea” to “Korea”
* “United Kingdom of Great Britain and Northern Ireland” to “United Kingdom”
* “Türkiye” to “Turkiye”
* “Turkey” to “Turkiye”
* “United States of America Virgin Islands” to “Virgin Islands”

Column Titles Changes

* GDP per capita (US dollars) to GDP\_PerCapita
* GDP in constant 2010 prices (millions of US dollars) to GDP\_2010Prices\_Millions
* Population mid-year estimates (millions) to Pop\_mill
* Population mid-year estimates for males (millions) to Pop\_male\_mill
* Population mid-year estimates for females (millions) to Pop\_female\_mill
* Population aged 0 to 14 years old (percentage) to Pop\_0\_14\_per
* Population aged 60+ years old (percentage) to Pop\_60\_plus\_per
* Population density to Pop\_density
* Surface area (thousand km2) to Surface\_area\_thou\_km2
* Consumer price index: General to CPI\_General
* Consumer price index: Food to CPI\_Food
* Shortened the unemployment column names to: Unemp\_Male, Unemp\_Female, Unemp\_General
* Shortened Education column names to: Enroll\_Ratio\_Primary\_Female, Enroll\_Ratio\_Primary\_Male, Enroll\_Ratio\_Secondary\_Female, Enroll\_Ratio\_Secondary\_Male, Enroll\_Ratio\_UpperSecondary\_Female, Enroll\_ratio\_UpperSecondary\_Male

Data Profile

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Economic Freedom Variables | | | | |
|  | Year | Country | Economic Freedom Summary Index | Rank | Quartile |
| Minimum | 1970 |  | 1.87 | 1 | 1 |
| Maximum | 2021 |  | 9.28 | 165 | 4 |
| Mean | 2004.5 |  | 6.53 | 72.32 | 2.48 |
| Mode | 2021 |  | 8.11 | 25 | 2 |
| Number of Observations | 4620 | 165 countries | 3941 | 3941 | 3941 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | GDP | | | |
|  | Country | Year | GDP in 2010 prices | GDP per capita |
| Minimum |  | 1995 | 0 | $91 |
| Maximum |  | 2020 | $19,925,439 million | $189,507 |
| Mean |  | 2012 | $323,124.53 million | $14,880 |
| Mode |  | 2010 | n/a | n/a |
| Number of Observations | 613 countries | 1995, 2005, 2010, 2015, 2018, 2019, 2020 | 1491 | 1485 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Population Dataset | | | | | | | | |
|  | Countries | Years | Population Mid Year Estimates | Population Mid Year Estimates (Males) | Population Mid Year Estimates (Females) | % of population aged 0 to 14 years | % of population aged 60+ | Population density | Surface Area |
| Minimum |  |  | 0 | 0 | 0 | 11% | 0.30% | 0.1 | 0 |
| Maximum |  |  | 1425.89 million | 731.18 million | 698.24 million | 49.90% | 43.70% | 24,779.90 | 17,098 |
| Mean |  |  | 32.56 million | 16.38 million | 16.18 million | 26.85% | 13.14% | 443.25 | 587.25 |
| Number of Observ. | 234 countries | 2010, 2015, 2017, 2020, 2022 | 926 | 926 | 926 | 926 | 926 | 930 | 457 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | CPI Dataset | | | |
|  | Countries | Years | CPI: General | CPI: Food |
| Minimum |  |  | 0 | 19.6 |
| Maximum |  |  | 93,584.3 (Venezuela) | 6268.9 (Also Venezuela) |
| Mean |  |  | 268.14 | 141.32 |
| Mode |  | 2015 |  |  |
| Number of Observations | 202 Countries | 2000, 2005, 2010, 2015, 2019, 2020, 2021, total 2049 observations | 1238 | 811 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Unemployment Dataset | | | | | | | |
|  | Countries | Years | Labor Force Participation (Female) | Labor Force Participation (Male) | Labor Force Participation | Unemployment Rate (Female | Unemployment Rate (Male) | Unemployment Rate |
| Minimum |  |  | 6 | 32.8 | 26.4 | 0.3 | 0.1 | 0.1 |
| Maximum |  |  | 86.6 | 96.1 | 88.3 | 60.5 | 56.4 | 45.4 |
| Mean |  |  | 50.87 | 71.18 | 61.17 | 10.08 | 7.93 | 8.59 |
| Number of Observations | 216 countries | 2000, 2001, 2002, 2003, 2005, 2006, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2022  The years 2005, 2010, 2015, and 2022 have the most observations | 807 | 807 | 814 | 813 | 813 | 821 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Education | | | | | | | |
|  | Countries | Years | Gross enrollment ratio - Primary (female) | Gross enrollment ratio - Primary (male) | Gross enrollment ratio - Secondary (female) | Gross enrollment ratio - Secondary (male) | Gross enrollment ratio - Upper secondary level (female) | Gross enrollment ratio - Upper secondary level (male) |
| Minimum |  |  | 16.6 | 30.1 | 3.7 | 8.1 | 2.9 | 5.3 |
| Maximum |  |  | 163.4 | 164.2 | 180.4 | 173.5 | 210.2 | 239.8 |
| Mean |  |  | 102.21 | 105.29 | 81.74 | 81.53 | 74.69 | 72.95 |
| Mode |  |  |  |  |  |  |  |  |
| Number of Observations | 205 Countries | Yearly, 2000-2021 | 818 | 818 | 711 | 711 | 709 | 711 |

|  |  |
| --- | --- |
| Continent | |
| 248 countries | |
| Africa | 60 |
| Antarctica | 1 |
| Asia | 49 |
| Europe | 52 |
| North America | 41 |
| Oceania | 29 |
| South America | 16 |

|  |  |
| --- | --- |
| Income Group | |
| 218 countries | |
| High income | 83 |
| Low income | 26 |
| Lower middle income | 54 |
| Upper middle income | 54 |
| Blank (Venezuela) | 1 |

|  |  |
| --- | --- |
| Indicators | |
| 165 countries | |
| Number of Landlocked countries | 29 |
| Number of LDCs | 39 |
| Number of English speaking countries | 65 |
| Number of Popular countries | 92 |

Data Limitations

All of the data (with one exception, discussed below) included in this study are sourced from government or reliable international organizations. Each dataset has a slightly different list of countries. I have not deleted any countries, but the countries that are not included in the “Economic Freedom” will not be included with the principal statistical analysis.

The obvious (small, in terms of the volume of the data set) limitation is the “Popularity” indicator. I give my permission, and with her permission, to include this variable. I’m interested to see whether this variable is statistically significant.

Defining Questions

The principal question that we are going to explore is: Which of the independent variables determine the Economic Freedom of a country?

In the process, we are going to ask:

* How has each country’s Economic Freedom changed over time?
* Which are the top countries in terms of changes in Economic Freedom over time?
* Is there a particular independent variable that seems to have been a catalyst of change in a country’s Economic Freedom?
* Which are the top (and bottom) countries in terms of changes in GDP? Population? CPI? Unemployment? Enrollment?