ECOMOMIC DEVELOPMENT AND POLITICAL LIBERALIZATION

INTRODUCTION

In the past decades, different nations around the world have witnessed instances of economic developments and political liberalizations with most of such studies pegged on economic development. There have emerged questions on whether a country being a democracy results in any form of economic growth or none, with most studies presenting a mix of results (Zhang, 2020). These developments have been attached to growth in trade and liberalization, with more concerns placed on trade openness as being important measures of economic growth. However, the assumption of a positive relation between liberalization and trade has underwent a number of criticisms among different authors. Recently, in examining the joint impacts of economic liberalizations and economic liberalizations, there were conclusions that a number of reforms were quite helpful, especially when countries open up their economies to international trade as causing significant economic growth. Countries that become democracies first before opening up their economies have realized a significant sluggishness in their economic growth measures (Adegboye et.al, 2020). This indicates that countries should undertake liberalizations of their economies before considering political liberalizations.

Economic development measures within countries of the world require sustainment over the years. Most of the developing countries have often underwent challenges from substantial volatility in growth that has a negative effect in economic growth. In this case, policy makers have been advised to consider output volatilities when determining the reform sequences from different global countries (Hummel, 2020). Furthermore, trade openness has also been discussed as aiding economic growth across countries. Nonetheless, it is associated with significant levels of volatility at the macroeconomic levels. Moreover, democratization has also presented ambiguous effects on growth with some positive impacts especially in regards to economic stability from enhanced cohesion and creation of an enabling business environment for economic growth. Available literature in terms of growth volatility suggests the existence of trade-offs in the liberalization sequences over the years.

Literature Review

Economic theory

In economics, economic theory suggests that financial developments have a direct impact on economic growth metrics in the past years. The theory is also linked to political liberalizations since harmonious business environment provided by political stability is essential for business and economic growth (Li & Yao, 2020). These arguments are dated back to Schumpeter in 1912 in which services offered by financial intermediaries were quire helpful in both innovations and relevant developments over the past years. There are five main functions of the financial system across countries that influence economic growth. First, the financial systems enhance mobilization of savings for customers, ensuring that they have appropriate access to their finances across existing commercial banks (Haddad, 2020). The banks across different countries enhance development of the economy through offering savings that customers can use to access credit when in need. Secondly, financial sectors also aid in reducing risks across the economy. This is achieved through the inclusion of insurance corporations, key in terms of offering compensations in the event of any risk. Furthermore, this is also realized through most of customer savings kept within the existing financial systems to assure customers of the funds when needed.

Furthermore, the financial systems across the global economies aid in the facilitation of exchanging goods and services across clients around the world. In this case, governments aid in creating the good-will through political stable landscapes, to ease exchange of goods and services (Khayitboy & Ilhom, 2020). In this case, the aspect of liberalization plays a critical role especially in ensuring that the individuals realize their specific economic targets, hence improving the economic prospects over the years as witnessed across both developed and developing economies around the world. Most of these provisions are quite helpful in terms of promoting checks and balances, in terms of economic milestones in comparison to the previous years. Appropriate financial systems aids in the realization of underlying functions and roles as stipulated from economic perspectives over the years.

RESULTS AND DISCUSSION

The goal of this project is to gain a clearer intuition of the relation of distribution of wealth and perception of corruption through visualization. If the Sokoloff-Engerman hypothesis is true, then we would expect to find a correlation (though rigorously testing whether Sokoloff-Engerman hypothesis’s explanation for such a correlation existing is correct is beyond the scope of this project).

The Data Used Here[¶](file:///C:\Users\User\Downloads\econ323-final%20(1).html#The-Data-Used-Here)

This project uses [this](https://www.kaggle.com/transparencyint/corruption-index) data set from Transparency International, and [this](https://www.kaggle.com/theworldbank/poverty-and-equity-database) data set from the World Bank. The former is representative of perceptions of corruption in 2017. The latter contains data about inequality observed over the course of 1974 through 2018.

This notebook was executed on [kaggle.com](http://kaggle.com), and a version of it will be maintained on [github](https://github.com/keeganland/econ323)

Kaggle Defaults[¶](file:///C:\Users\User\Downloads\econ323-final%20(1).html#Kaggle-Defaults)

In [271]:

# This Python 3 environment comes with many helpful analytics libraries installed

# It is defined by the kaggle/python docker image: https://github.com/kaggle/docker-python

# For example, here's several helpful packages to load in

import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv)

import matplotlib.pyplot as plt

%matplotlib inline

# Input data files are available in the "../input/" directory.

# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os

for dirname, \_, filenames in os.walk('/kaggle/input'):

for filename in filenames:

print(os.path.join(dirname, filename))

# Any results you write to the current directory are saved as output.

/kaggle/input/corruption-index/index.csv

/kaggle/input/corruption-index/history.csv

/kaggle/input/poverty-and-equity-database/povstats-excel-zip-826-kb-/PovStatsEXCEL.xlsx

/kaggle/input/poverty-and-equity-database/povstats-csv-zip-242-kb-/PovStatsCountry.csv

/kaggle/input/poverty-and-equity-database/povstats-csv-zip-242-kb-/PovStatsFootNote.csv

/kaggle/input/poverty-and-equity-database/povstats-csv-zip-242-kb-/PovStatsCountry-Series.csv

/kaggle/input/poverty-and-equity-database/povstats-csv-zip-242-kb-/PovStatsSeries.csv

/kaggle/input/poverty-and-equity-database/povstats-csv-zip-242-kb-/PovStatsData.csv

The Transparency International data on corruption[¶](file:///C:\Users\User\Downloads\econ323-final%20(1).html#The-Transparency-International-data-on-corruption)

This dataset requires very little cleaning. It contains one of the variables of interest (as well as information on measurement error) as well as country codes that can be used for data frame merges. We will do an initial bar-graph visualization of it in order to do an initial intituition check.

In [272]:

corruption\_file = "/kaggle/input/corruption-index/index.csv"

df\_corruption\_index = pd.read\_csv(corruption\_file)

df\_corruption\_index = df\_corruption\_index.iloc[:,:8]

df\_corruption\_index.head()

Out[272]:

|  | CPI Rank | Country | Country Code | Region | Corruption Perceptions Index (CPI) | Standard Error | Lower Confidence Interval | Upper Confidence Interval |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1 | New Zealand | NZL | Asia Pacific | 90 | 2.56 | 86 | 94 |
| 1 | 1 | Denmark | DNK | Europe and Central Asia | 90 | 2.46 | 86 | 94 |
| 2 | 3 | Finland | FIN | Europe and Central Asia | 89 | 1.46 | 87 | 92 |
| 3 | 4 | Sweden | SWE | Europe and Central Asia | 88 | 1.33 | 85 | 90 |
| 4 | 5 | Switzerland | CHE | Europe and Central Asia | 86 | 1.57 | 83 | 89 |

In [273]:

# Used documentation from https://stackabuse.com/python-data-visualization-with-matplotlib/ to

#resize the figure so we can see all the countries listed in a large horizontal bar graph

fig\_size = plt.rcParams["figure.figsize"]

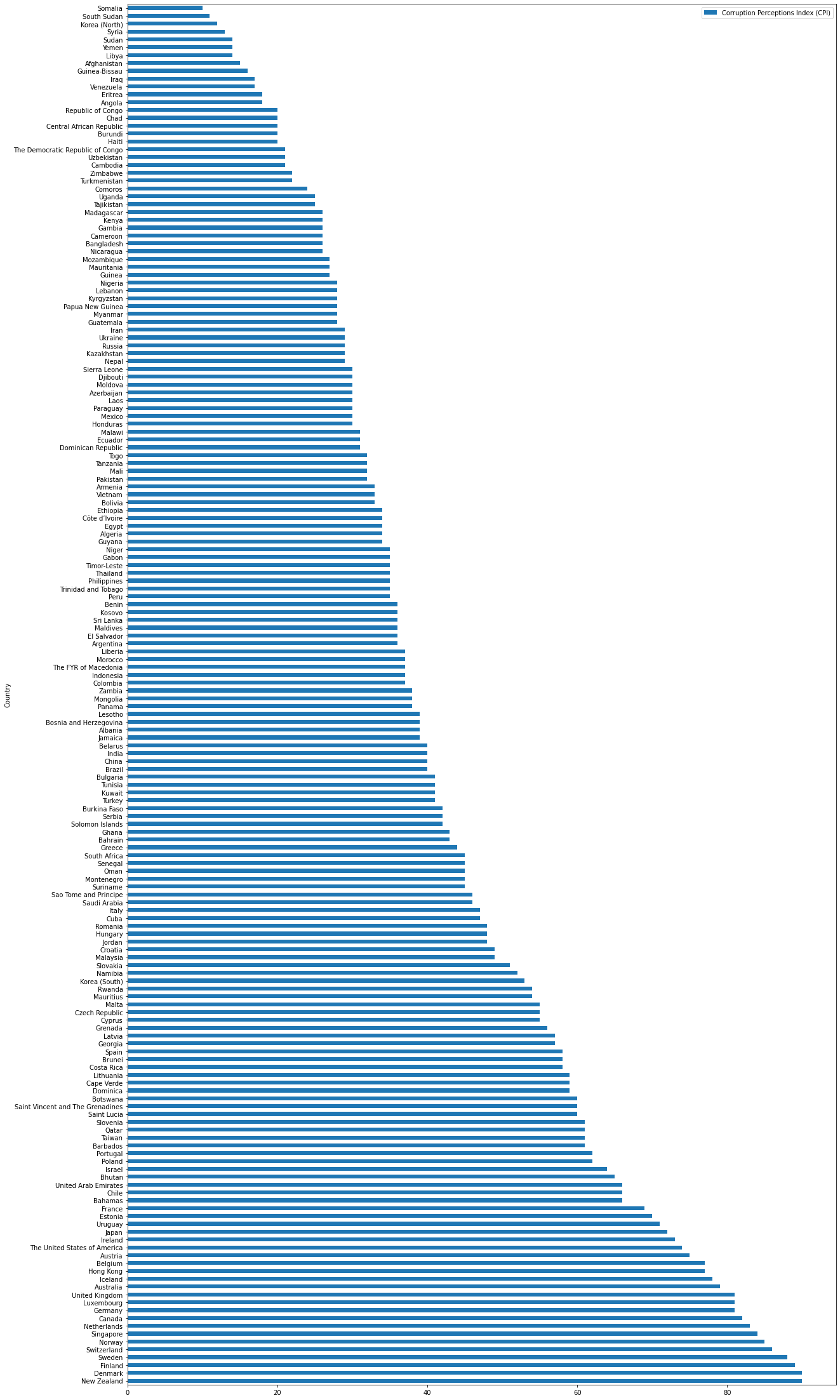
fig\_size[0] = 20

fig\_size[1] = 40

plt.rcParams["figure.figsize"] = fig\_size

#show bar graph

ax\_corruption\_bar\_graph = df\_corruption\_index.plot(x = "Country", y = "Corruption Perceptions Index (CPI)", kind = 'barh')



Gut check[¶](file:///C:\Users\User\Downloads\econ323-final%20(1).html#Gut-check)

The data does not look surprising. Countries that have a reputation as developing countries score poorly, whereas countries that have a reputation as wealthy, developed liberal democracies score highly.

This bar graph helps us get an intuitive feel for what CPI as a variable looks like. It varies from 0 to 100, 0 representing being perceived as most corrupt, 100 being least corrupt (that is, it runs on the "high score = good" intuition). It also serves as a guide, as the CPI varies numerically, which countries that numerical variation actually corresponds to.

In [274]:

# Return the figure size to something more managable for future plotting

fig\_size = plt.rcParams["figure.figsize"]

fig\_size[0] = 12

fig\_size[1] = 10

plt.rcParams["figure.figsize"] = fig\_size

The World Bank's data set[¶](file:///C:\Users\User\Downloads\econ323-final%20(1).html#The-World-Bank's-data-set)

The World Bank's dataset records various economic indicators for multiple regions and countries, with observations recorded by year. However, an observation does not exist for every indicator for every country. Furthermore, we are not necessarily interested in all the indicators in this data set. Because we are interested in seeing the relationship between corruption and inequality, we are most interested in their estimates of countries' Gini index, which is the standard measure of economic inequality. We shall therefore need to slice and clean the data frame.

In [275]:

poverty\_stats\_series\_file = "/kaggle/input/poverty-and-equity-database/povstats-csv-zip-242-kb-/PovStatsSeries.csv"

poverty\_stats\_country\_file = "/kaggle/input/poverty-and-equity-database/povstats-csv-zip-242-kb-/PovStatsCountry.csv"

poverty\_stats\_country\_series\_file = "/kaggle/input/poverty-and-equity-database/povstats-csv-zip-242-kb-/PovStatsCountry-Series.csv"

poverty\_stats\_data\_file = "/kaggle/input/poverty-and-equity-database/povstats-csv-zip-242-kb-/PovStatsData.csv"

df = pd.read\_csv(poverty\_stats\_data\_file)

df.head()

Out[275]:

|  | Country Name | Country Code | Indicator Name | Indicator Code | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | ... | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Unnamed: 49 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | East Asia & Pacific | EAS | Annualized growth in per capita real survey me... | SI.SPR.PC40.ZG | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 1 | East Asia & Pacific | EAS | Annualized growth in per capita real survey me... | SI.SPR.PT10.ZG | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 2 | East Asia & Pacific | EAS | Annualized growth in per capita real survey me... | SI.SPR.PT60.ZG | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 3 | East Asia & Pacific | EAS | Annualized growth in per capita real survey me... | SI.SPR.PCAP.ZG | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 4 | East Asia & Pacific | EAS | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

5 rows × 50 columns

In [276]:

df\_indexed = df.set\_index(["Country Code", "Indicator Code"])

df\_indexed

Out[276]:

|  |  | Country Name | Indicator Name | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | ... | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Unnamed: 49 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Country Code | Indicator Code |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| EAS | SI.SPR.PC40.ZG | East Asia & Pacific | Annualized growth in per capita real survey me... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| SI.SPR.PT10.ZG | East Asia & Pacific | Annualized growth in per capita real survey me... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| SI.SPR.PT60.ZG | East Asia & Pacific | Annualized growth in per capita real survey me... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| SI.SPR.PCAP.ZG | East Asia & Pacific | Annualized growth in per capita real survey me... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| SI.POV.GINI | East Asia & Pacific | GINI index (World Bank estimate) | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ZWE | SI.SPR.PCAP | Zimbabwe | Survey mean consumption or income per capita, ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| SI.POV.URGP | Zimbabwe | Urban poverty gap at national poverty lines (%) | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| SI.POV.URGP.NC | Zimbabwe | Urban poverty gap at national poverty lines (%... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| SI.POV.URHC | Zimbabwe | Urban poverty headcount ratio at national pove... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| SI.POV.URHC.NC | Zimbabwe | Urban poverty headcount ratio at national pove... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

8140 rows × 48 columns

In [277]:

df\_grouped = df.groupby("Indicator Code")

df\_grouped = df\_grouped.get\_group("SI.POV.GINI")

df\_grouped = df\_grouped.set\_index(["Country Code"])

df\_grouped

Out[277]:

|  | Country Name | Indicator Name | Indicator Code | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | ... | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Unnamed: 49 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Country Code |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| EAS | East Asia & Pacific | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| ECS | Europe & Central Asia | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| FCS | Fragile and conflict affected situations | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| HIC | High income | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| DFS | IDA countries classified as fragile situations | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| VNM | Vietnam | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 39.3 | NaN | 35.6 | NaN | 34.8 | NaN | 35.3 | NaN | NaN | NaN |
| PSE | West Bank and Gaza | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 35.3 | 34.4 | NaN | NaN | NaN | NaN | 33.7 | NaN | NaN | NaN |
| YEM | Yemen, Rep. | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | 36.7 | NaN | NaN | NaN | NaN | NaN |
| ZMB | Zambia | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 55.6 | NaN | NaN | NaN | NaN | 57.1 | NaN | NaN | NaN | NaN |
| ZWE | Zimbabwe | GINI index (World Bank estimate) | SI.POV.GINI | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | 43.2 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

185 rows × 49 columns

Which number is "the" Gini index for our purposes?[¶](file:///C:\Users\User\Downloads\econ323-final%20(1).html#Which-number-is-"the"-Gini-index-for-our-purposes?)

A glance at the above data frame shows us that, as much as we would have liked to have a Gini index number for every country in the world at every year, the numbers we actually have correspond to irregular observations over the course of 1974 through to 2018. Further, as time marches on, economic, social, and political forces will be acting to change the level of inequality in any given country. The irregularity of the observations may disguise interesting trends within a country, to take one example, or patterns that represent causal forces acting on many countries at once, to take another example.

We need some way to summarize these numbers. For the sake of argument, we will assume that there is no systematic biases effecting when a country could be observed for the sake of this data set, meaning the mean of our observations should be a good estimator of the actual mean Gini index for each country over this period.

While of less intrinsic interest, we also will look at minimum observed Gini indexes and maximum observed Gini indexes. Visualizing these alongside the mean Gini indexes should give a (rough!) intuition of variance.

In [278]:

year\_range = range(1974,2019,1)

country\_codes = df\_grouped.index

minimum\_gini\_series = pd.Series(index=country\_codes, name="Minimum observed GINI index")

maximum\_gini\_series = pd.Series(index=country\_codes, name="Maximum observed GINI index")

mean\_gini\_series = pd.Series(index=country\_codes, name="Mean observed GINI index")

num\_observations\_series = pd.Series(index=country\_codes, name="Number of estimations of GINI index")

for country in country\_codes:

#will be used to compute mean observed gini

successful\_gini\_observations = 0

total\_gini = 0

#conceptually, the Gini index ranges from 0 to 100, these are therefore conceptual extremes of minimum/maximum

minimum\_gini = 100

maximum\_gini = 0

mean\_gini = 0

country\_series = df\_grouped.loc[country]

for year in year\_range:

gini\_this\_year = country\_series.loc[str(year)]

if pd.notna(gini\_this\_year):

successful\_gini\_observations = successful\_gini\_observations + 1

total\_gini = total\_gini + gini\_this\_year

if gini\_this\_year < minimum\_gini:

minimum\_gini = gini\_this\_year

#print(minimum\_gini)

if gini\_this\_year > maximum\_gini:

maximum\_gini = gini\_this\_year

#print(maximum\_gini)

if successful\_gini\_observations > 0:

mean\_gini = total\_gini / successful\_gini\_observations

minimum\_gini\_series.loc[country] = minimum\_gini

maximum\_gini\_series.loc[country] = maximum\_gini

mean\_gini\_series.loc[country] = mean\_gini

num\_observations\_series.loc[country] = int(successful\_gini\_observations)

In [279]:

#simplify the data frame now that we have summary statistics

df\_grouped = df\_grouped.iloc[:,:3]

df\_grouped["Mean GINI"] = mean\_gini\_series

df\_grouped["Min GINI"] = minimum\_gini\_series

df\_grouped["Max GINI"] = maximum\_gini\_series

df\_grouped["Number of observations"] = num\_observations\_series

df\_grouped

Out[279]:

|  | Country Name | Indicator Name | Indicator Code | Mean GINI | Min GINI | Max GINI | Number of observations |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Country Code |  |  |  |  |  |  |  |
| EAS | East Asia & Pacific | GINI index (World Bank estimate) | SI.POV.GINI | 0.000000 | 100.0 | 0.0 | 0.0 |
| ECS | Europe & Central Asia | GINI index (World Bank estimate) | SI.POV.GINI | 0.000000 | 100.0 | 0.0 | 0.0 |
| FCS | Fragile and conflict affected situations | GINI index (World Bank estimate) | SI.POV.GINI | 0.000000 | 100.0 | 0.0 | 0.0 |
| HIC | High income | GINI index (World Bank estimate) | SI.POV.GINI | 0.000000 | 100.0 | 0.0 | 0.0 |
| DFS | IDA countries classified as fragile situations | GINI index (World Bank estimate) | SI.POV.GINI | 0.000000 | 100.0 | 0.0 | 0.0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| VNM | Vietnam | GINI index (World Bank estimate) | SI.POV.GINI | 36.130000 | 34.8 | 39.3 | 10.0 |
| PSE | West Bank and Gaza | GINI index (World Bank estimate) | SI.POV.GINI | 34.525000 | 33.7 | 35.6 | 8.0 |
| YEM | Yemen, Rep. | GINI index (World Bank estimate) | SI.POV.GINI | 35.466667 | 34.7 | 36.7 | 3.0 |
| ZMB | Zambia | GINI index (World Bank estimate) | SI.POV.GINI | 52.688889 | 42.1 | 60.5 | 9.0 |
| ZWE | Zimbabwe | GINI index (World Bank estimate) | SI.POV.GINI | 43.200000 | 43.2 | 43.2 | 1.0 |

185 rows × 7 columns

Some further data cleaning[¶](file:///C:\Users\User\Downloads\econ323-final%20(1).html#Some-further-data-cleaning)

Some observed countries simply do not have an observed Gini index at any point. These are no good to us for our purposes. We remove these from the data frame.

In [280]:

#For some countries, we simply lack any helpful data about inequality. We can pick these out because Mean GINI is still 0.

for country in country\_codes:

row = df\_grouped.loc[country]

if row.loc["Number of observations"] == 0:

df\_grouped = df\_grouped.drop([country])

df\_grouped

Out[280]:

|  | Country Name | Indicator Name | Indicator Code | Mean GINI | Min GINI | Max GINI | Number of observations |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Country Code |  |  |  |  |  |  |  |
| ALB | Albania | GINI index (World Bank estimate) | SI.POV.GINI | 29.660000 | 27.0 | 31.7 | 5.0 |
| DZA | Algeria | GINI index (World Bank estimate) | SI.POV.GINI | 34.366667 | 27.6 | 40.2 | 3.0 |
| AGO | Angola | GINI index (World Bank estimate) | SI.POV.GINI | 47.350000 | 42.7 | 52.0 | 2.0 |
| ARG | Argentina | GINI index (World Bank estimate) | SI.POV.GINI | 46.141379 | 40.6 | 53.8 | 29.0 |
| ARM | Armenia | GINI index (World Bank estimate) | SI.POV.GINI | 32.255556 | 28.0 | 37.5 | 18.0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| VNM | Vietnam | GINI index (World Bank estimate) | SI.POV.GINI | 36.130000 | 34.8 | 39.3 | 10.0 |
| PSE | West Bank and Gaza | GINI index (World Bank estimate) | SI.POV.GINI | 34.525000 | 33.7 | 35.6 | 8.0 |
| YEM | Yemen, Rep. | GINI index (World Bank estimate) | SI.POV.GINI | 35.466667 | 34.7 | 36.7 | 3.0 |
| ZMB | Zambia | GINI index (World Bank estimate) | SI.POV.GINI | 52.688889 | 42.1 | 60.5 | 9.0 |
| ZWE | Zimbabwe | GINI index (World Bank estimate) | SI.POV.GINI | 43.200000 | 43.2 | 43.2 | 1.0 |

164 rows × 7 columns

Merging the data sets[¶](file:///C:\Users\User\Downloads\econ323-final%20(1).html#Merging-the-data-sets)

'Country Code' is common to both data sets, allowing us to easily perform a merge. We do an inner merge here because that automatically excludes countries (or regions) that are only in one data set or the other. The merged data frame still contains a majority of the countries in the world.

In [281]:

df\_merged = df\_grouped.merge(right=df\_corruption\_index,how='inner',on='Country Code')

df\_merged

Out[281]:

|  | Country Code | Country Name | Indicator Name | Indicator Code | Mean GINI | Min GINI | Max GINI | Number of observations | CPI Rank | Country | Region | Corruption Perceptions Index (CPI) | Standard Error | Lower Confidence Interval | Upper Confidence Interval |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | ALB | Albania | GINI index (World Bank estimate) | SI.POV.GINI | 29.660000 | 27.0 | 31.7 | 5.0 | 83 | Albania | Europe and Central Asia | 39 | 1.99 | 36 | 42 |
| 1 | DZA | Algeria | GINI index (World Bank estimate) | SI.POV.GINI | 34.366667 | 27.6 | 40.2 | 3.0 | 108 | Algeria | Middle East and North Africa | 34 | 2.94 | 29 | 39 |
| 2 | AGO | Angola | GINI index (World Bank estimate) | SI.POV.GINI | 47.350000 | 42.7 | 52.0 | 2.0 | 164 | Angola | Sub-Saharan Africa | 18 | 1.68 | 15 | 21 |
| 3 | ARG | Argentina | GINI index (World Bank estimate) | SI.POV.GINI | 46.141379 | 40.6 | 53.8 | 29.0 | 95 | Argentina | Americas | 36 | 1.76 | 33 | 39 |
| 4 | ARM | Armenia | GINI index (World Bank estimate) | SI.POV.GINI | 32.255556 | 28.0 | 37.5 | 18.0 | 113 | Armenia | Europe and Central Asia | 33 | 4.01 | 26 | 40 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 144 | VEN | Venezuela, RB | GINI index (World Bank estimate) | SI.POV.GINI | 49.307692 | 42.5 | 55.6 | 13.0 | 166 | Venezuela | Americas | 17 | 1.41 | 15 | 20 |
| 145 | VNM | Vietnam | GINI index (World Bank estimate) | SI.POV.GINI | 36.130000 | 34.8 | 39.3 | 10.0 | 113 | Vietnam | Asia Pacific | 33 | 2.46 | 29 | 38 |
| 146 | YEM | Yemen, Rep. | GINI index (World Bank estimate) | SI.POV.GINI | 35.466667 | 34.7 | 36.7 | 3.0 | 170 | Yemen | Middle East and North Africa | 14 | 3.05 | 9 | 19 |
| 147 | ZMB | Zambia | GINI index (World Bank estimate) | SI.POV.GINI | 52.688889 | 42.1 | 60.5 | 9.0 | 87 | Zambia | Sub-Saharan Africa | 38 | 2.91 | 34 | 43 |
| 148 | ZWE | Zimbabwe | GINI index (World Bank estimate) | SI.POV.GINI | 43.200000 | 43.2 | 43.2 | 1.0 | 154 | Zimbabwe | Sub-Saharan Africa | 22 | 2.59 | 18 | 26 |

149 rows × 15 columns

Linear Regressions and Visualizations[¶](file:///C:\Users\User\Downloads\econ323-final%20(1).html#Linear-Regressions-and-Visualizations)

In [282]:

import seaborn as sns

from sklearn import linear\_model

linear\_regressor = LinearRegression()

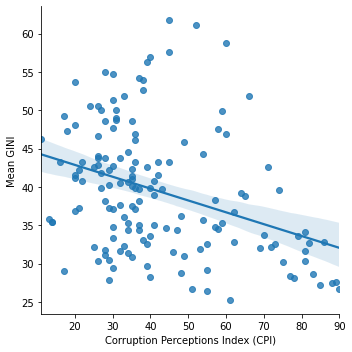
x\_corruption = df\_merged["Corruption Perceptions Index (CPI)"]

y\_gini = df\_merged["Mean GINI"]

sns.lmplot(data = df\_merged, x = "Corruption Perceptions Index (CPI)", y = "Mean GINI")

Out[282]:

<seaborn.axisgrid.FacetGrid at 0x7f671308eac8>

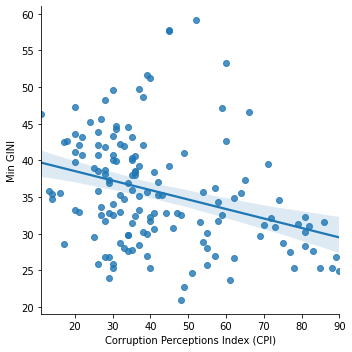


In [283]:

sns.lmplot(data = df\_merged, x = "Corruption Perceptions Index (CPI)", y = "Min GINI")

Out[283]:

<seaborn.axisgrid.FacetGrid at 0x7f671304d908>

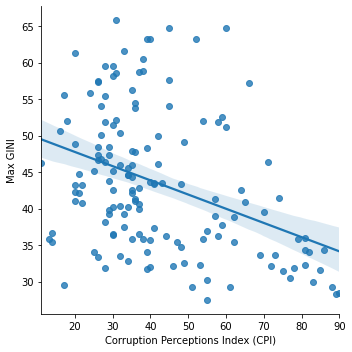


In [284]:

sns.lmplot(data = df\_merged, x = "Corruption Perceptions Index (CPI)", y = "Max GINI")

Out[284]:

<seaborn.axisgrid.FacetGrid at 0x7f6712f73ef0>



What do these graphs mean?[¶](file:///C:\Users\User\Downloads\econ323-final%20(1).html#What-do-these-graphs-mean?)

CONCLUSION

The Gini index is often described as ranging from a score of 0, which represents a perfectly egalitarian economy with the income or wealth of every person in the economy is exactly equal, to a score of 100, which represents an economy where all the income or wealth goes to a single person and none goes to anyone else. Thus, a lower score is indicative of a more egalitarian economy, and a higher score is indicative of a less egalitarian. The Corruption Perception Index, however, works on the "high score is good" intuition. Low scorers are perceived as corrupt, high scorers are perceived as not corrupt.

Therefore, given the empirical theories mentioned in the introduction about corruption and rent seeking causing inequality, we would predict there to be a negative relationship - which is exactly what our linear regression says we do predict. The slope (though not the intercept, obviously) is even roughly the same regardless of which representative Gini index number we use.

This is of minimum value for confirming the causal hypotheses discussed at the beginning of this report, but it is consistent with those hypotheses in such a way so as to suggest to us they are on the right track.