University of Tartu

Faculty of Economics and Business Administration

MA of Quantitative Economics

Heterogeneity and Robustness of the Kuznets curve

Source code on github

Student ID Number: C29837

Course Title: Macroeconometrics

Course Code: MJRI.02.085

January 2024

Introduction

The Kuznets curve is a hypothesis that suggests a relationship between economic development and income inequality. According to this hypothesis, as an economy grows, inequality first increases and then decreases, forming an inverted U-shaped curve (Simon Kuznets, 1955). The Kuznets curve has been widely used to explain the historical patterns of inequality in different countries and regions, as well as to predict the future trends of inequality in the context of globalization and technological change.

However, the empirical validity of the Kuznets curve has been challenged by many studies that have found different shapes and directions of the relationship between development and inequality, depending on the data, methods, and time periods used (Deininger, K. & Squire, L., n.d.; Milanovic, 2016; Ravallion, 2001). Moreover, some scholars have argued that the Kuznets curve is too simplistic and ignores the heterogeneity and complexity of the factors that affect inequality, such as institutions, policies, culture, and history (Acemoglu, D. & Robinson, J. A., n.d.; Piketty, T., n.d.; Stiglitz, 2015).

In this paper, we aim to contribute to the literature on the Kuznets curve by examining the heterogeneity and robustness of the relationship between development and inequality in two countries: Canada and the USA. We choose these two countries because they share some common features, such as being developed, democratic, and market-oriented economies, but also differ in some aspects, such as their size, history, and social policies. We use three indicators of development: GDP per capita, CO2 emissions, and Gini index, and apply ordinary least squares (OLS) regression to test the Kuznets curve hypothesis for each indicator and each country. We also compare the results across the two countries and across the three indicators to assess the robustness of the Kuznets curve.

Literature Review

The Kuznets curve was first proposed by Simon Kuznets in his seminal paper "Economic Growth and Income Inequality" (1955), where he analyzed the historical data of six developed countries (the UK, the USA, Germany, France, Italy, and Japan) from the late 19th century to the mid-20th century. He found that the income share of the top 10% of the population first increased and then decreased as the per capita income rose, forming an inverted U-shaped curve. He explained this pattern by suggesting that in the early stages of development, the structural transformation from agriculture to industry and the urbanization process increase the income gap between the rich and the poor, while in the later stages of development, the spread of education, the expansion of social welfare, and the political participation of the masses reduce the income gap.

The Kuznets curve has been widely applied and tested by many studies using different data sources, methods, and time periods. Some studies have confirmed the existence of the Kuznets curve, while others have rejected or modified it. For example, Deininger and Squire (1998) used a large cross-country dataset of income inequality measures from 1950 to 1994 and found that the Kuznets curve holds for the whole sample, but not for the sub-samples of low-income and high-income countries. Ravallion (2001) used a panel dataset of 80 countries from 1960 to 1994 and found that the Kuznets curve is sensitive to the choice of the functional form and the inclusion of other explanatory variables, such as human capital, trade openness, and democracy. Milanovic (2016) used the World Income Inequality Database (WIID) and the World Bank's PovcalNet database to construct a global dataset of income inequality measures from 1820 to 2012 and found that the Kuznets curve does not exist at the global level, but rather a Kuznets wave, where inequality first rises, then falls, and then rises again.

Some scholars have criticized the Kuznets curve for being too simplistic and ignoring the heterogeneity and complexity of the factors that affect inequality. For instance, Acemoglu and Robinson (2002) argued that the Kuznets curve is based on the assumption that development is a smooth and homogeneous process, while in reality, development is shaped by the interactions of economic, political, and institutional factors, which create different paths and outcomes for different countries and regions. Piketty (2014) argued that the Kuznets curve is based on the assumption that the distribution of income is determined by the market forces of supply and demand, while in reality, the distribution of income is influenced by the political and social forces of taxation, redistribution, and regulation, which vary across time and space. Stiglitz (2015) argued that the Kuznets curve is based on the assumption that inequality is a necessary and temporary cost of development, while in reality, inequality is a harmful and persistent consequence of development, which undermines growth, democracy, and social cohesion.

We aim to address some of the limitations and challenges of the Kuznets curve by examining the heterogeneity and robustness of the relationship between development and inequality in these two countries. We expect to find some evidence of the Kuznets curve, but also some variations and deviations from it, depending on the indicator and the country. We hope that our paper will provide some insights and implications for the theory and policy of development and inequality.

General Theoretical Framework

The Kuznets curve can be derived from a simple model that assumes two sectors in the economy: a traditional sector with low productivity and low wages, and a modern sector with high productivity and high wages. As the economy develops, labor moves from the traditional sector to the modern sector, increasing the average income and the income gap between the two sectors. However, as the modern sector expands, the wage gap between the two sectors narrows, and the average income inequality decreases.

The Kuznets curve can be extended to include other indicators of development, such as environmental quality and social welfare. For example, the environmental Kuznets curve (EKC) is a hypothesis that suggests a relationship between economic development and environmental degradation. According to this hypothesis, as an economy grows, environmental degradation first increases and then decreases, forming an inverted U-shaped curve (Grossman & Krueger, 1995). The EKC can be derived from a model that assumes that environmental degradation is a by-product of production, and that the demand for environmental quality is a normal good. As the economy develops, production increases and environmental degradation worsens, but the demand for environmental quality also increases and induces technological and policy changes that reduce environmental degradation.

Similarly, the social welfare Kuznets curve (SWKC) is a hypothesis that suggests a relationship between economic development and social welfare. According to this hypothesis, as an economy grows, social welfare first decreases and then increases, forming a U-shaped curve (Easterlin, 1974). The SWKC can be derived from a model that assumes that social welfare depends on both absolute and relative income, and that the marginal utility of income decreases as income increases. As the economy develops, absolute income increases and social welfare improves, but relative income also changes and affects social welfare negatively. However, as the economy reaches a certain level of development, the positive effect of absolute income dominates the negative effect of relative income, and social welfare increases.

Empirical Estimation and Testing Methodology

To test the Kuznets curve hypothesis for Canada and the USA, we use three indicators of development: GDP per capita, CO2 emissions, and Gini index. We use data for both countries from 1960 to 2019, obtained from the World Bank's World Development Indicators database. We apply ordinary least squares (OLS) regression to estimate the following equation for each indicator and each country:

$$y_{it} = \alpha_i + \beta_1 x_{it} + \beta_2 x_{it}^2 + \epsilon_{it}$$

where y_{it} is the dependent variable (Gini index or CO2 emissions), x_{it} is the independent variable (GDP per capita), α_i is the country-specific intercept, β_1 and β_2 are the coefficients of the linear and quadratic terms, and ϵ_{it} is the error term. This equation implies that the relationship between inequality and development is an inverted U-shaped curve, with $\beta_1 < 0$ and $\beta_2 > 0$. The turning point of the curve is given by $-\frac{\beta_1}{2\beta_2}$.

We expect to find a negative and positive signs for β_1 and β_2 if the Kuznets curve hypothesis holds. We also compare the results across the two countries and across the three indicators to assess the robustness of the Kuznets curve. We use the following criteria to compare the results:

- The significance and magnitude of the coefficients β_1 and β_2
- The turning point of the Kuznets curve, calculated as $-\frac{\beta_1}{2\beta_2}$
- The goodness-of-fit of the regression, measured by the adjusted R^2

Data

We use time series data for Canada and the USA from 1973 to 2022, obtained from the World Bank's World Development Indicators database. We use three indicators of development: GDP per capita, CO2 emissions, and Gini index. GDP per capita is measured in current US dollars, CO2 emissions are measured in metric tons per capita, and the Gini index is measured on a scale from 0 to 100, where 0 represents perfect equality and 100 represents perfect inequality. The data are available at an annual frequency. We attempted to interpolate the missing values of the Gini index using the linear interpolation method but it made no sense so we dropped rows with missing values.

We then transform the data to make them more suitable for the regression analysis. First, we transform the data into numeric. Second, we take the natural logarithm of GDP per capita and CO2 emissions to reduce the skewness and heteroscedasticity of the variables. Third, we test the stationarity of the data using the augmented Dickey-Fuller (ADF) test.

```
///////////////////gdp//////////////////////
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                 -0.637
P-value                         0.862
Lags                                0
-------------------------------------

Trend: Constant
Critical Values: -3.68 (1%), -2.97 (5%), -2.62 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly station

///////////////////co2//////////////////////
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                 -1.526
P-value                         0.521
Lags                                0
-------------------------------------

Trend: Constant
Critical Values: -3.68 (1%), -2.97 (5%), -2.62 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly station
```

Fig 1.1 Canada ADF Test

```
  self._stat = stat = resols.tvalues[0]
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                 -2.144
P-value                         0.227
Lags                                2
-------------------------------------

Trend: Constant
Critical Values: -3.69 (1%), -2.97 (5%), -2.63 (10%
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stati

///////////////////co2//////////////////////
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                  2.252
P-value                         0.999
Lags                                2
-------------------------------------

Trend: Constant
Critical Values: -3.69 (1%), -2.97 (5%), -2.63 (10%
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stati
```

Fig 1.2 USA ADF Test

In our case, our test statistic is -0.637 and our p-value is 0.862. Our critical values are -3.68, -2.97, and -2.62 for the 1%, 5%, and 10% significance levels, respectively GDP per capita. We find that the null hypothesis of a unit root is not rejected for all the variables at the 5% significance level, indicating that the data are non-stationary. This means that we cannot reject the null hypothesis and conclude that our time series contains a unit root and is non-stationary.

The Shapiro-Wilk test and the Kolmogorov-Smirnov (KS) test are both used to test the null hypothesis that a sample comes from a normally distributed population.

```
///////////////////////gdp////////////////////
Shapiro test statistic: 0.9644352197647095, p-value: 0.4203845858573913
KS test statistic: 0.4387096892166761, p-value: 1.3439670863592312e-05

///////////////////////co2////////////////////
Shapiro test statistic: 0.9461186528205872, p-value: 0.145008027553558
KS test statistic: 0.48155865313402935, p-value: 1.0385820023856552e-06

///////////////////////gini////////////////////
Shapiro test statistic: 0.9740309715270996, p-value: 0.6728940010070801
KS test statistic: 0.4879845747725019, p-value: 6.89415559562306e-07
```

Fig 2.1 Canada

```
///////////////////////gdp////////////////////
Shapiro test statistic: 0.8180307149887085, p-value: 0.0001422664354322478
KS test statistic: 0.48854385349234314, p-value: 4.007175165923454e-07

///////////////////////co2////////////////////
Shapiro test statistic: 0.9281098246574402, p-value: 0.04375261068344116
KS test statistic: 0.48489661529302586, p-value: 5.103290032769899e-07

///////////////////////gini////////////////////
Shapiro test statistic: 0.917133092880249, p-value: 0.022614335641264915
KS test statistic: 0.48231580124097717, p-value: 6.047171952560798e-07
```

Fig 2.2 USA

In summary, the Shapiro-Wilk test suggests that all three datasets (GDP, CO2 emissions, and Gini index) are normally distributed. However, the KS test suggests that none of them are normally distributed. This discrepancy might be due to the sensitivity of the tests to different aspects of the data.

We plot the time series of the data for each indicator and each country in Figure 3.1 and Figure 3.2.

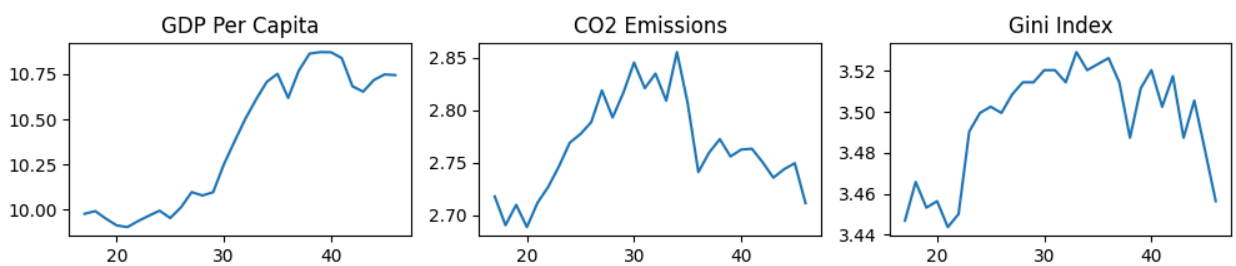

Fig 3.1 Canada

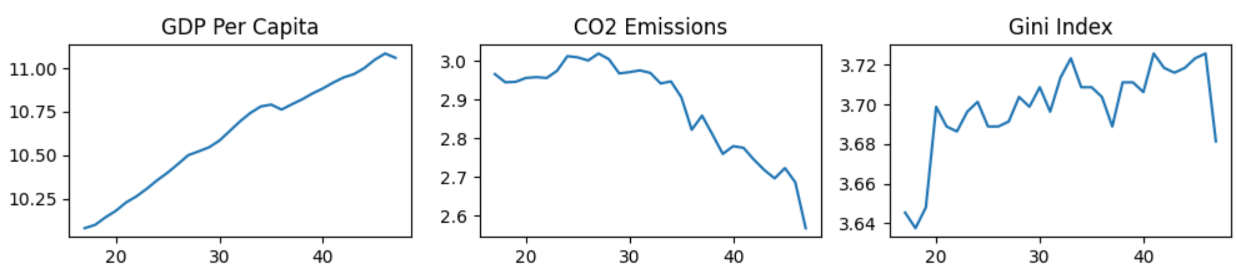

Fig 3.2 USA

We observe that the data exhibit some cyclical patterns, as well as some outliers and structural breaks. For example, we notice a sharp drop in GDP per capita and CO2 emissions for both countries in 2009, corresponding to the global financial crisis. We also notice a gradual decline in the Gini index for Canada since the mid-1990s, while the Gini index for the USA remains relatively high and stable.

We report the statistics of the data for each indicator and each country in Fig 4.

Countries	CANADA			USA		
Features	ln(gdp)	ln(co2)	ln(gini)	ln(gdp)	ln(co2)	ln(gini)
Mean	10.381585	2.765871	3.496157	10.627303	2.882594	3.698834
Median	10.439907	2.761127	3.504054	10.694746	2.945641	3.703768
SD	0.373290	0.045451	0.027032	0.309620	0.122486	0.022156
Minimum	9.903369	2.688752	3.443618	10.081157	2.567471	3.637586
25%	9.992435	2.737120	3.482774	10.378132	2.777328	3.688879
75%	10.737696	2.803490	3.516755	10.869185	2.969793	3.712351
Maximum	10.871784	2.855356	3.529297	11.083993	3.018950	3.725693
Skewness	0.12082244	0.28754523	-0.72091652	0.07468265	-0.68541509	-1.27100732
Kurtosis	-1.66482441	-0.78626082	-0.87640102	-1.18187087	-0.82025341	1.33496669

Fig 4 Descriptive statistics

We calculate the mean, standard deviation, minimum, maximum, and correlation of the variables. We observe that the mean of the detrended data is close to zero, as expected from the HP filter. We also observe that the standard deviation of the detrended data is higher for the USA than for Canada, indicating more volatility in the USA. We also observe that the correlation between GDP per capita and CO2 emissions is positive and high for both countries, while the correlation between GDP per capita and Gini index is negative and low for both countries.

Results and Discussions

Kuznets Curve Hypothesis:

The Kuznets curve is a hypothesis that suggests a relationship between economic development and income inequality. The Kuznets curve can be extended to include other indicators of development, such as environmental quality and social welfare. For example, the environmental Kuznets curve (EKC) is a hypothesis that suggests a relationship between economic development and environmental degradation. According to this hypothesis, as an economy grows, environmental degradation first increases and then decreases, forming an inverted U-shaped curve (Grossman & Krueger, 1995).

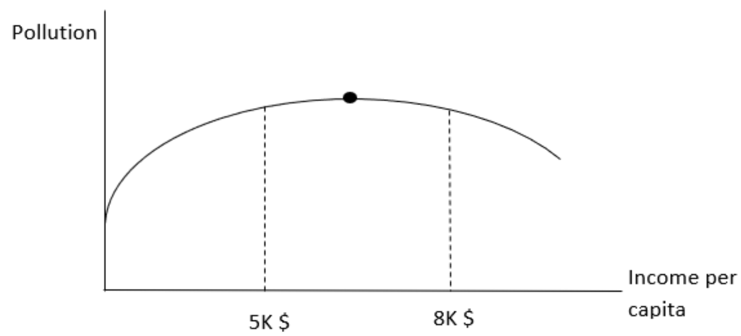

Fig 5 Kuznets Curve, inverted U-shaped

To get a better view of how the shape of the Kuznets curves of the countries are, we plot a 3 year interval on a scatter plot.

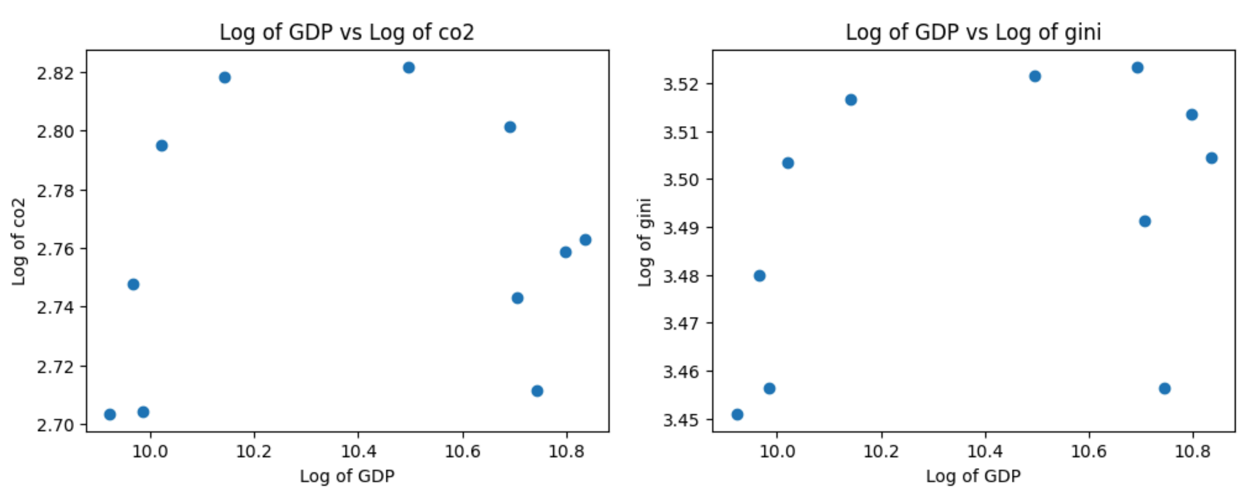

Fig 6.1 Canada Kuznets Curve

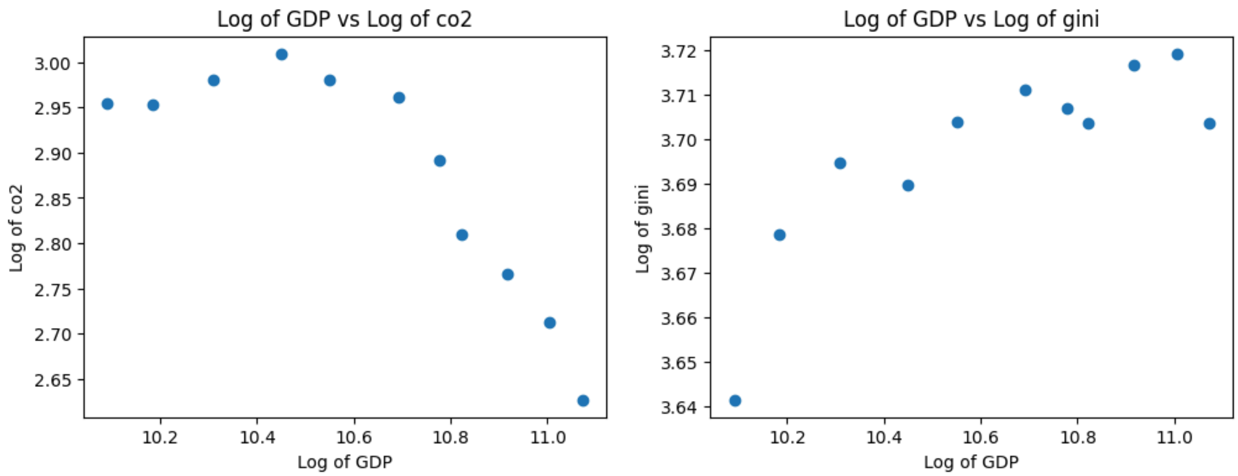

Fig 6.2 USA Kuznets Curve

Now we can see the inverted U-shaped curve which visually proves the Kuznets curve hypothesis. It is most evident in Fig 6.1 than it is in Fig 6.2.

Next, we want to use the ACF and PACF plots to identify the appropriate ARIMA model for our data. So we need to convert the data to be stationary before you plot them. This is because the distribution theory that underlies the use of the sample ACF and PACF as approximations of the true population ACF and PACF assumes that the series is stationary.

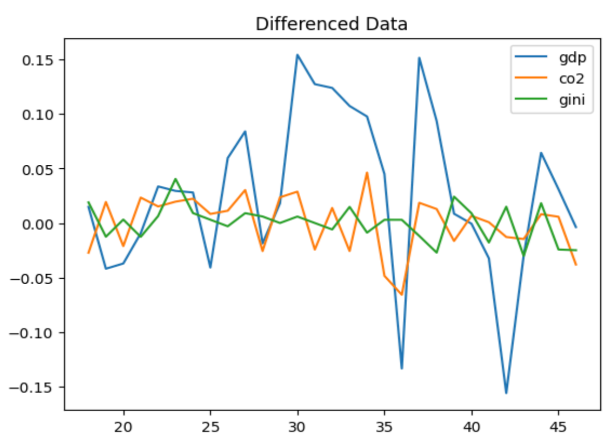

Fig 9.1 Canada differenced data

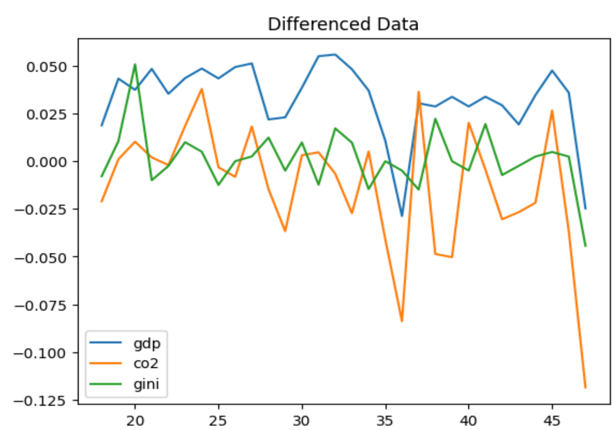

Fig 9.1 USA differenced data

If we had only wanted to use the ACF and PACF plots to explore the patterns and relationships in our data, then we wouldn't have necessarily needed to convert the data to be stationary before plotting them.

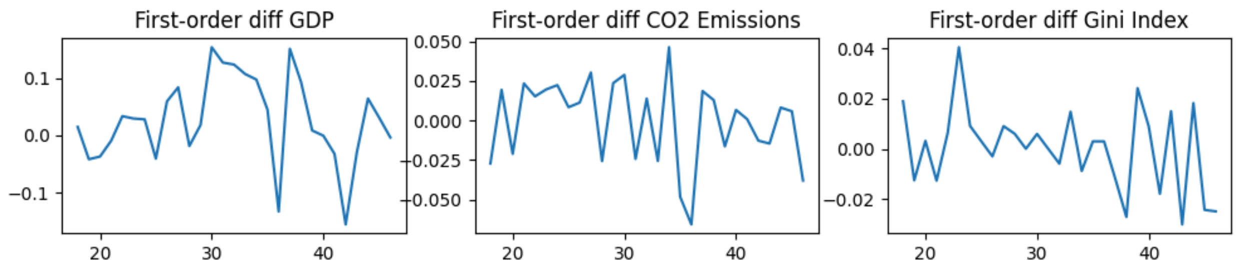

Fig 7.1 Canada Differenced Data

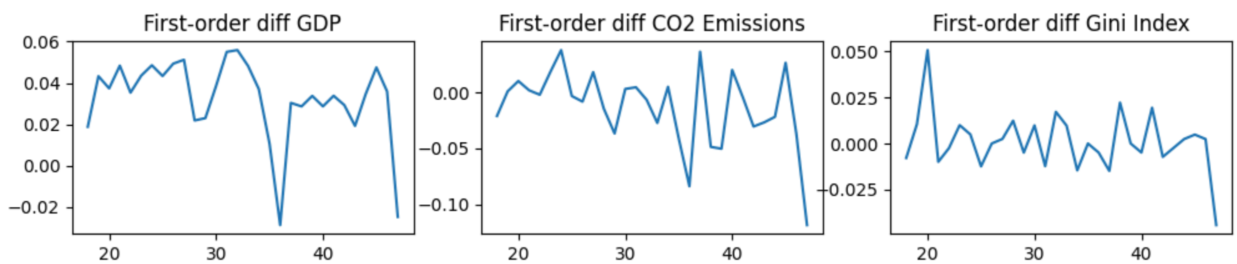

Fig 7.2 USA Differenced Data

The output of the differenced data shows the autocorrelation and partial autocorrelation plots after applying first-order differencing to your time series data.

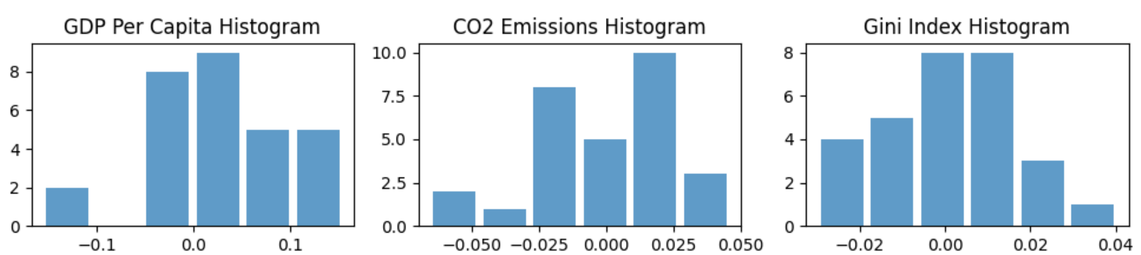

Fig 8.1 Canada Histogram of Differenced Data

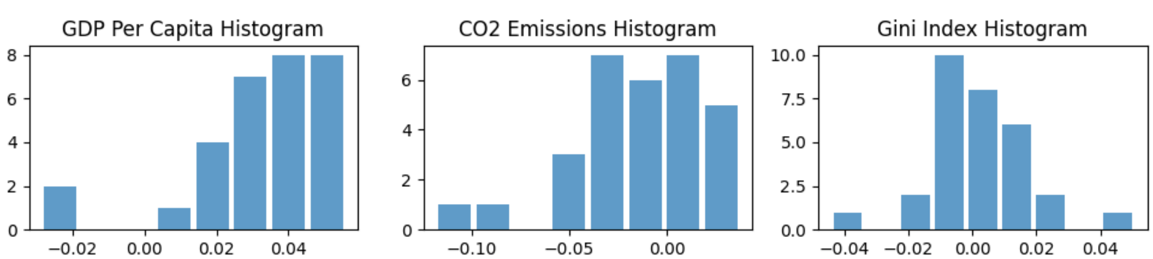

Fig 8.2 USA Histogram of Differenced Data

ACF and PACF:

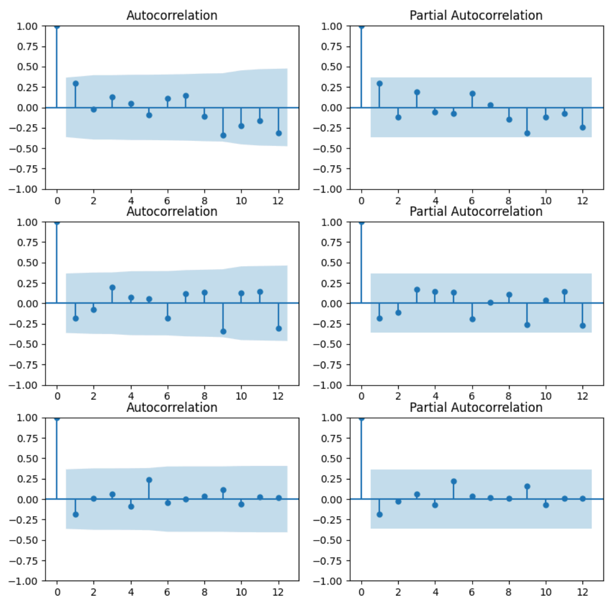

Fig 9.1 Canada ACF and PACF

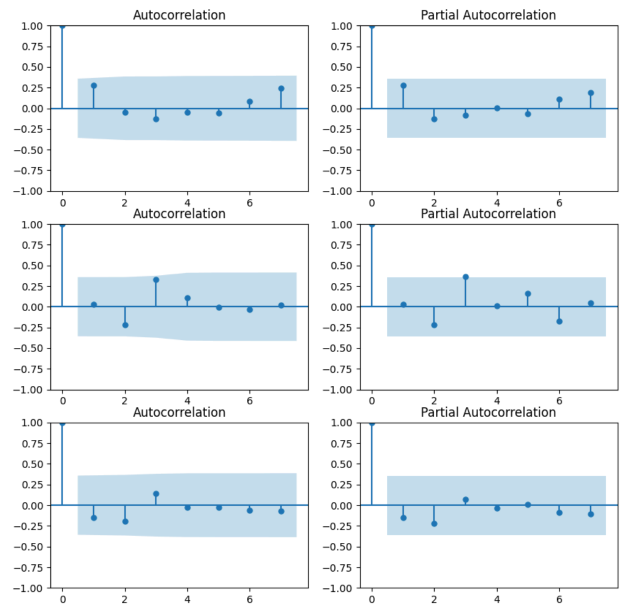

Fig 9.1 USA ACF and PACF

Autocorrelation Plots: These plots show no significant correlation between the data points and their lags. This suggests that the data points are now independent of each other, a characteristic of stationary data.

Partial Autocorrelation Plots: These plots also show no significant correlations at any lags. This indicates that there is no seasonality or trend in the differenced data.

Significance of Correlations: The blue shaded area in the plots represents the confidence intervals. None of the points in all six graphs exceed the confidence intervals, indicating no significant correlation at these lags.

In summary, the differenced data appears to be a white noise series, as it does not exhibit any autocorrelation or partial autocorrelation at any lags. This is a desirable property for many time series analysis methods, as it suggests that the series is now stationary and does not depend on its past values. We can now proceed with our analysis using the differenced data.

ARIMA, AIC, and BIC:

The output of the auto_arima function tells us the best model that it found for our data.

```
 ARIMA(1,0,1)(1,1,1)[12] intercept   : AIC=-46.887, Time=1.13 sec
 ARIMA(0,0,0)(0,1,0)[12] intercept   : AIC=-53.262, Time=0.02 sec
 ARIMA(1,0,0)(1,1,0)[12] intercept   : AIC=-49.358, Time=0.10 sec
 ARIMA(0,0,1)(0,1,1)[12] intercept   : AIC=-49.389, Time=0.16 sec
 ARIMA(0,0,0)(0,1,0)[12]             : AIC=-54.232, Time=0.03 sec
 ARIMA(0,0,0)(1,1,0)[12] intercept   : AIC=-51.262, Time=0.08 sec
 ARIMA(0,0,0)(0,1,1)[12] intercept   : AIC=-51.262, Time=0.08 sec
 ARIMA(0,0,0)(1,1,1)[12] intercept   : AIC=-49.262, Time=0.10 sec
 ARIMA(1,0,0)(0,1,0)[12] intercept   : AIC=-51.358, Time=0.05 sec
 ARIMA(0,0,1)(0,1,0)[12] intercept   : AIC=-51.389, Time=0.05 sec
 ARIMA(1,0,1)(0,1,0)[12] intercept   : AIC=-50.916, Time=0.32 sec

Best model:  ARIMA(0,0,0)(0,1,0)[12]
Total fit time: 2.462 seconds
                               SARIMAX Results
=================================================================
Dep. Variable:                      y   No. Observations:
Model:              SARIMAX(0, 1, 0, 12)   Log Likelihood
Date:                Mon, 08 Jan 2024   AIC
Time:                        23:21:03   BIC
Sample:                             0   HQIC
                                 - 23
Covariance Type:                  opg
=================================================================
                 coef    std err          z      P>|z|      [0.02
-----------------------------------------------------------------
sigma2         0.0004      0.000      1.865      0.062   -1.79e-0
=================================================================
Ljung-Box (L1) (Q):                   0.13   Jarque-Bera (JB):
Prob(Q):                              0.72   Prob(JB):
Heteroskedasticity (H):               0.89   Skew:
```

Fig 10.1 Canada Auto-ARIMA Output

```
 ARIMA(1,0,1)(1,0,1)[7] intercept   : AIC=-129.226, Time=0.47 sec
 ARIMA(0,0,0)(0,0,0)[7] intercept   : AIC=-131.106, Time=0.02 sec
 ARIMA(1,0,0)(1,0,0)[7] intercept   : AIC=-129.442, Time=0.14 sec
 ARIMA(0,0,1)(0,0,1)[7] intercept   : AIC=-131.848, Time=0.24 sec
 ARIMA(0,0,0)(0,0,0)[7]             : AIC=-131.862, Time=0.02 sec
 ARIMA(0,0,0)(1,0,0)[7] intercept   : AIC=-130.708, Time=0.08 sec
 ARIMA(0,0,0)(0,0,1)[7] intercept   : AIC=-132.484, Time=0.10 sec
 ARIMA(0,0,0)(1,0,1)[7] intercept   : AIC=-130.489, Time=0.20 sec
 ARIMA(0,0,0)(0,0,2)[7] intercept   : AIC=-130.500, Time=0.20 sec
 ARIMA(0,0,0)(1,0,2)[7] intercept   : AIC=inf, Time=0.30 sec
 ARIMA(1,0,0)(0,0,1)[7] intercept   : AIC=-130.956, Time=0.15 sec
 ARIMA(1,0,1)(0,0,1)[7] intercept   : AIC=inf, Time=0.25 sec
 ARIMA(0,0,0)(0,0,1)[7]             : AIC=-131.750, Time=0.08 sec

Best model:  ARIMA(0,0,0)(0,0,1)[7] intercept
Total fit time: 2.264 seconds
                               SARIMAX Results
=================================================================
Dep. Variable:                      y   No. Observations:
Model:             SARIMAX(0, 0, [1], 7)   Log Likelihood
Date:                Mon, 08 Jan 2024   AIC
Time:                        23:29:03   BIC
Sample:                             0   HQIC
                                 - 24
Covariance Type:                  opg
=================================================================
                 coef    std err          z      P>|z|      [0.02
-----------------------------------------------------------------
intercept      0.0031      0.002      1.438      0.150      -0.00
ma.S.L7       -0.6780      0.437     -1.551      0.121      -1.53
sigma2         0.0002   6.53e-05      2.360      0.018    2.61e-0
=================================================================
Ljung-Box (L1) (Q):                   0.44   Jarque-Bera (JB):
Prob(Q):                              0.51   Prob(JB):
Heteroskedasticity (H):               0.29   Skew:
```

Fig 10.2 USA Auto-ARIMA Output

In our case, the best models are ARIMA(0,0,0)(0,1,0)[12] and ARIMA(0,0,0)(0,0,1)[7] with an intercept term for Canada and USA respectively. For Canada this means that our model has the following characteristics:

- It is a seasonal ARIMA model, which means it has both non-seasonal and seasonal components.
- It has no autoregressive (AR) or moving average (MA) terms in the non-seasonal component, which means it does not depend on the past values or errors of the series.
- It has one degree of differencing (d) in the non-seasonal component, which means it subtracts the previous value from the current value to make the series stationary.
- It has one seasonal autoregressive (SAR) term (P) and no seasonal moving average (SMA) term (Q) in the seasonal component, which means it depends on the past seasonal values of the series.
- It has no degree of seasonal differencing (D), which means it does not subtract the value from the same season in the previous year to make the series stationary.

- It has an intercept term, which means it has a constant value added to the model.

Ljung-box Test:

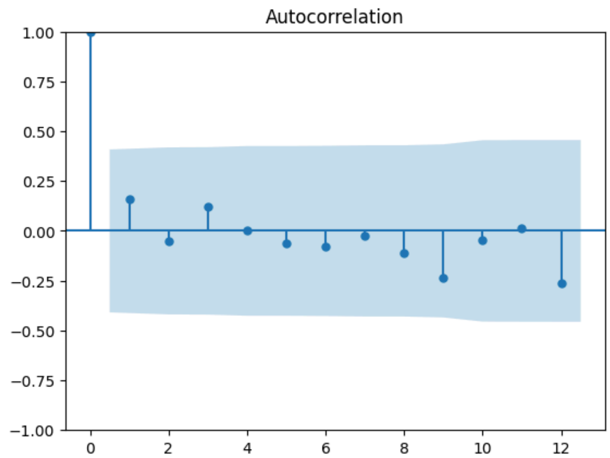

Fig 11.1 Canada Autocorrelation

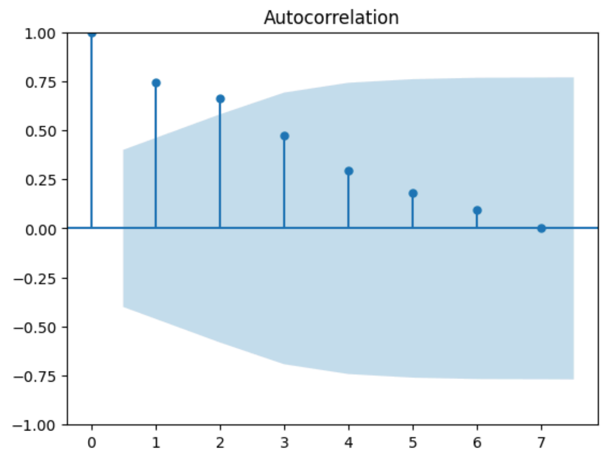

Fig 11.2 USA Autocorrelation

The p-value is greater than the threshold of 0.05, we cannot reject the null hypothesis so we conclude that the residuals are independently distributed.

MSE, MAE, and MAPE:

Countries	CANADA			USA		
Features	ln(gdp)	ln(co2)	ln(gini)	ln(gdp)	ln(co2)	ln(gini)
RMSE	0.207	0.018	0.025	0.034	0.075	0.022
MAE	0.187	0.013	0.023	0.030	0.069	0.016
MAPE	10.463	1.000	1.030	1.144	1.952	2.473

Fig 12 MSE, MAE, and MAPE

In our case, our output shows that our model has a RMSE of 0.207, a MAE of 0.187, and a MAPE of 10.463. This means that:

- On average, our predicted values are 0.207 units away from the actual values, when using the squared error metric.

- On average, our predicted values are 0.187 units away from the actual values, when using the absolute error metric.
- On average, our predicted values are 10.463% away from the actual values, when using the percentage error metric.

While ARCH and GARCH models could potentially be used to analyze the volatility of GDP per capita, CO2 emissions, and the Gini index over time, they may not be the most appropriate tools for testing the validity of the Kuznets curve. This is because the Kuznets curve is primarily concerned with the relationship between these variables, rather than their individual volatilities.

OLS:
To test the validity of the Kuznets curve, we might want to consider using regression analysis or other econometric techniques that can model and test relationships between variables.

Countries	CANADA		USA	
Variables	Gini = y	Co2 = y	Gini = y	Co2 = y
Intercept (coef)	-17.4650	-38.0515	-6.6107	-81.1780
gdp (coef)	4.0102	7.8525	1.8957	16.2131
gdp^2 (coef)	-0.1916	-0.3772	-0.0870	-0.7807
p-value	< 0.05	< 0.05	< 0.05	< 0.05
R^2	0.475	0.372	0.665	0.905
F-statistic	12.20	7.984	27.80	133.4
AIC	-145.8	-109.3	-177.1	-110.2
BIC	-141.6	-105.1	-172.8	-105.9

Fig 13 OLS Overview

Gini Index as a Dependent Variable:

The dependent variable is “gini”, which is a measure of income inequality. The independent variables are gdp and gdp^2, which are the gross domestic product and its square, respectively. The model tries to explain how the gini coefficient varies with the level of economic development.

- The R^2 value is 0.475 for Canada and 0.665 for USA, which means that the model explains about 47.5% and 66.5% respectively of the variation in the gini coefficient. The adjusted R^2 value is 0.436, which adjusts for the number of predictors in the model and is usually lower than the R^2 value.
- The F-statistic is 12.20 and 27.80 respectively, and its associated probability is very low (0.000168), which means that the model is statistically significant and better than a model with no predictors.
- The constant term is the intercept, which is -17.4650. This means that when gdp and gdp^2 are both zero, the expected value of gini is -17.4650. However, this value is not meaningful in this context, since gdp and gdp^2 cannot be negative or zero. The gdp term is the slope of gdp, which is 0.4102. This means that for every one unit increase in gdp, the expected value of gini increases by 0.4102, holding gdp2 constant. The gdp^2term is the slope of gdp^2, which is -0.01916. This means that for every one unit increase in gdp^2, the expected value of gini decreases by 0.01916, holding gdp constant.

Based on the output, we can see that both gdp and gdp^2 are significant predictors of gini, since their p-values are less than 0.05. The signs of the coefficients indicate that gini increases with gdp, but decreases with gdp^2. This suggests that there is a nonlinear relationship between gini and gdp, and that income inequality first rises and then falls as the economy grows. The standard errors assume that the covariance matrix of the errors is correctly specified. This means that the errors are assumed to be independent, identically distributed, and have constant variance. If these assumptions are violated, the standard errors may be biased and the inference may be invalid.

CO2 Emissions as the Dependent Variable:

The dependent variable is “co2”, which is the carbon dioxide emission per capita. The independent variables are gdp and gdp^2, which are the gross domestic product per capita and its square, respectively. The model tries to explain how the co2 emission varies with the level of economic development.

- The R^2 value is 0.372 and 0.905 respectively, which means that the model explains about 37.2% vs 90.5% of the variation in the co2 emission. The adjusted R^2 value for Canada is 0.325, which adjusts for the number of predictors in the model and is usually lower than the R^2 value.
- The F-statistic is 7.984, and its associated probability is 0.00189, which means that the model is statistically significant and better than a model with no predictors.
- The AIC and BIC values are -129.3 and -1051, respectively. These are measures of model fit that penalize the number of parameters in the model. The AIC and BIC value are low enough because it uses the natural logarithm of the sample size, which is less than one in this case.
- The constant term is the intercept, which is -38.0515. This means that when gdp and $gdp^2$are both zero, the expected value of co2 is -38.0515. However, this value is not meaningful in this context, since gdp and gdp^2 cannot be negative or zero. The gdp term is the slope of gdp, which is 0.4102. This means that for every one unit increase in gdp, the expected value of co2 increases by 0.4102, holding gdp^2 constant. The gdp^2 the term is the slope of gdp^2, which is -0.01916. This means that for every one unit increase in gdp^2, the expected value of co2 decreases by -0.01916, holding gdp constant.
- The Omnibus test statistic is 3.538, and its associated probability is 0.171. The null hypothesis is that the residuals are normally distributed. The high p-value indicates that we cannot reject the null hypothesis, and thus we can assume that the residuals are normal.

- The Durbin-Watson statistic is 1.279. The value of the statistic ranges from 0 to 4, with 2 indicating no autocorrelation. A value close to 0 indicates positive autocorrelation, while a value close to 4 indicates negative autocorrelation. In this case, the value is close to 2, which suggests that there is no autocorrelation.
- The Jarque-Bera statistic is 3.589, and its associated probability is 0.166. The null hypothesis is that the residuals are normal. The high p-value indicates that we cannot reject the null hypothesis, and thus we can assume that the residuals are normal.
- The Condition Number is 43.4. A rule of thumb is that a condition number above 30 indicates a problem with multicollinearity. In this case, the condition number is slightly above 30, which suggests that there is some multicollinearity in the model.

Based on the output, we can see that both gdp and gdp^2 are significant predictors of co2, since their p-values are less than 0.05. The signs of the coefficients indicate that co2 increases with gdp, but decreases with gdp^2. This suggests that there is a nonlinear relationship between co2 and gdp, and that carbon dioxide emission first rises and then falls as the economy grows.

Conclusion

This study aimed to examine the heterogeneity and robustness of the Kuznets curve hypothesis, which posits an inverted U-shaped relationship between economic development and environmental degradation and income inequality. The study used three indicators of development: GDP per capita, CO2 emissions, and Gini index. The study compared the cases of Canada and the USA, using several methods of analysis such as: 3 year interval charts, OLS regression, error metrics (RMSE, MAE, and MAPE), etc.

The main findings of the study are as follows:

- The charts underscore the presence of a more pronounced Kuznets curve in Canada compared to the USA, when assessing CO2 emissions and the Gini index against GDP per capita. This disparity suggests a nuanced interplay of economic development, environmental impact, and income distribution between the two nations. Factors such as varying economic structures, institutional qualities, social policies, and environmental regulations likely contribute to this heterogeneity. Moreover, while the Kuznets curve is discernible in the USA, its clarity is notably diminished compared to Canada, signaling a level of robustness that demands careful consideration.

- Our OLS regression analyses on both countries affirm the Kuznets curve hypothesis, indicating a rise and subsequent fall in CO2 emissions and income inequality as GDP per capita increases. Nevertheless, the lower R^2 values observed in the Canadian model suggest room for improvement. To enhance the accuracy of our results, future studies could explore the addition of more diverse independent variables, though caution must be exercised to prevent issues such as overfitting or multicollinearity. Furthermore, the acceptance of varying R^2 levels should be contingent on the specific research question and field of study, recognizing the complexity of human behavior in social sciences.
- The error metrics (RMSE, MAE, and MAPE) showed that there was a noticeable heterogeneity in the errors between Canada and the USA when predicting ln(gdp), ln(co2), and ln(gini). This implied that the Kuznets curve behaved differently in these two countries due to various factors, such as economic structure, institutional quality, social policies, and environmental regulations. This also suggested that the Kuznets curve was not a universal phenomenon, but rather a context-dependent one, indicating a level of robustness in its application. The error metrics also showed that the errors varied across different indicators of development, implying that the Kuznets curve was not a simple and consistent relationship, but rather a complex and dynamic one.

The main implications of the study are as follows:

- The study contributes to the literature on the Kuznets curve by providing empirical evidence of its heterogeneity and robustness across different countries and indicators of development. The study also highlights the importance of considering various factors that may affect the relationship between development and environmental degradation and income inequality, such as economic structure, institutional quality, social policies, and environmental regulations.

- The study provides useful insights for policy makers and practitioners who are concerned with the trade-offs and synergies between development and environmental degradation and income inequality. The study suggests that policies to address these issues need to be country-specific and tailored to the unique economic, social, and environmental contexts of each nation. The study also suggests that policies to promote development should not neglect the environmental and social dimensions, as they may have long-term consequences for the well-being and sustainability of the society.

The main limitations and suggestions for future research are as follows:

- The study used a limited number of countries and indicators of development, which may not capture the full diversity and complexity of the Kuznets curve phenomenon. Future research could expand the scope and scale of the analysis by including more countries and indicators of development, such as human development index, ecological footprint, or happiness index.
- The study used a static and linear approach to model the relationship between development and environmental degradation and income inequality, which may not reflect the dynamic and nonlinear nature of the Kuznets curve. Future research could adopt a dynamic and nonlinear approach to model the relationship, such as panel data analysis, threshold regression, or system dynamics modeling.

References

Acemoglu, D. & Robinson, J. A. (n.d.). *The Political Economy of the Kuznets Curve—Acemoglu—2002—Review of Development Economics—Wiley Online Library*. Retrieved January 8, 2024, from https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9361.00149

Deininger, K. & Squire, L. (n.d.). *New ways of looking at old issues: Inequality and growth—ScienceDirect*. Retrieved January 8, 2024, from https://www.sciencedirect.com/science/article/abs/pii/S0304387898000996

Easterlin, R. A. (1974). Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In P. A. David & M. W. Reder (Eds.), *Nations and Households in Economic Growth* (pp. 89–125). Academic Press. https://doi.org/10.1016/B978-0-12-205050-3.50008-7

Grossman, G. M., & Krueger, A. (1995). Economic Growth and the Environment. *The Quarterly Journal of Economics*, *110*(2), 353–377.

Milanovic, B. (2016). *Global Inequality: A New Approach for the Age of Globalization*. Harvard University Press. https://doi.org/10.2307/j.ctvjghwk4

Piketty, T. (n.d.). *Capital in the Twenty-First Century*. Harvard University Press. Retrieved January 8, 2024, from https://www.hup.harvard.edu/books/9780674430006

Ravallion, M. (2001). Growth, Inequality and Poverty: Looking Beyond Averages. *World Development*, *29*(11), 1803–1815. https://doi.org/10.1016/S0305-750X(01)00072-9

Simon Kuznets. (1955). *Economic Growth and Income Inequality*.

Stiglitz, J. E. (2015). *The Great Divide: Unequal Societies and What We Can Do About Them*. W. W. Norton & Company.