How are inequality and growth related?

Analysis of income inequality: a simultaneity problem

Student ID 2102799

Word Count: 4802

Abstract

Based on Panel cross-sectional data from 45 countries from 1974 to 2018, this study provides new techniques, such as Granger-Tests and the use of Empirical Orthogonal Function, to analyse the evidence on the causal relationship between economic growth and income inequality. The aims of the present work are to discover the causality direction between inequality and growth; is economic growth changing inequality or is inequality affecting economic growth. Knowing if low economic growth rates are causing more inequality or more inequality is causing less economic growth is vital to orientate the designee of public policies, allowing them to allocate the efforts to one objective, increase growth or decrease inequality. However, the causality direction stays inconclusive.

Table of Contents

Introduction

Why is it important to talk about inequality?

"We were growing together for the first three decades after World War II, but for the last three decades, we have been growing apart" (Krueger, 2012). Since 1970, the world has experienced an increase in inequality in every possible measure; for example, the median wage for US male workers has not risen since 1973 (Goldin & Katz, 2008). Additionally, in the world, the wealthiest 1 % pays 15 % of total income in taxes, while the 50 % of the population pays 28.4 % (Stiglitz, 2012); moreover, in 2007, "the top 0.1 % of America's households had an income that was 220 times larger than the average of the bottom 90 %" (Stiglitz, 2012). Nevertheless, the reality of inequality goes further than numbers.

The consequence of inequality begins in a macroeconomic perception; the first consequence is a decrease in aggregate demand. The economy seen as a hold; it is cyclical, and the money created returns as an expense; expense that will return as an investment; however, "increased inequality has reduced aggregate demand because the well-off have a lower marginal propensity to consume than everyone else." (Krueger, 2012) Even though this money is expected to return to the economy as an investment for savings, in the short run, the immediate effect on the demand is less than expected because of inequality.

The second consequence of inequality is a stickiness in social-economic mobility over generations; according to the Economic Mobility Project in the United Kingdom, the likelihood that the children of the poorest stay in the same income levels are 40 %, even though it is not clear that this stickiness could make the aggregate demand more inelastic, it has been suggested. (Dabla, et. al, 2015)

Third consequence, inequality affects productivity from two fronts; first, in wage theory, "Ernst Fehr found that raising pay for workers who felt that they were underpaid substantially increased their productivity but raising pay for those who did not feel underpaid had no effect on productivity." (Krueger,2012) Therefore, a fairer distribution of wages would increase productivity. On the other hand, the ones at the bottom are less likely to go to university and have better jobs, which represents a loss of human capital as they will not have the opportunity to get the tools to reach their full potential (Stiglitz, 2012). The problem continues; as the productivity reached from those at the bottom is diminished, the returns archived from the government expenses in public education seem weak, resulting in less motivation to keep spending on education.

The fourth consequence of inequality is that it could be the cause of a financial crisis; Kumhof and Rancière (2011) find that income inequality in the lower deciles leads workers to borrow to maintain consumption, increasing indebtment that increases leverage and eventually causes a shock that leads to a financial crisis (Bordo & Meissner2012). Using a similar approach, in 2010, according to Raghuram Rajan, "rising inequality in the past three decades led to political pressure for redistribution that eventually came in the form of subsidised housing finance", leading to the financial crisis of 2008.

Finally, the increase in violence and political polarisation because of the frustration from the middle class that has not been seeing a growing effect in their income, as well as the moral approach of our longtermism responsibility to the next generation, who will be determined on the income of their parents, over justify the importance to discuss inequality.

Why is it important to talk about economic growth?

There are multiple reasons for economies to incentive their economic growth; namely, "Economic growth increases state capacity and the supply of public goods" (Sen, 2021), the growth in national income increases the tax revenue used to invest in public goods, more money, more public goods. Secondly, when economic growth occurs, the spending on investment increases, leading to more employment (Haldane, 2021); Even though the relationship 'more investment causes more employment´ does not necessarily hold, the evidence suggests that economic growth generates job opportunities.

Literature review

How are inequality and growth related?

Sine 1955, after a paper published by The American Economic Review where Simon Kuznets analysed the trends between inequality and economic growth of Germany, the United Kingdom, and the United States from 1910 to 1950, a positive relationship between equality and growth has been found. "The paper is 5% empirical information and 95% speculation, …, that speculation is an effective way of presenting a broad view of the field; and that so long as it is recognised as a collection of hunches calling for further investigation rather than a set of fully tested conclusions" (Kuznets, 1955, p. 26) However, the World Bank, the International Monetary Fund and most of the countries that follow their recommendations have considered, supported by further analysis, that the relation between growth and equality is not only positive, it is going in one direction, in other words, economy growth produces equality instead of equality produces growth; to the degree that "The World Bank considered the acceleration of economic growth to be a sufficient measure for improving the conditions of all strata within the population" (Lyubimov, 2017). The implications of considering that growth causes equality are relapsing in developing fiscal and monetary policies prioritising growth over solving inequality. This effort may be insufficient to reduce inequality.

In the past 20 years, further investigation has been made. Stiglitz, in 2012 found that the raising in inequality since 1970 has caused lower rates of economic growth, "the bankers and the political leaders have figured out how to design a financial system that could engage in excessive risk-taking, market manipulation, and predatory practices" (Stiglitz,2013, p.XXVI) predatory practices that transfer money from the bottom to the top instead of creating wealth has resulted in a deceleration of the economic growth and an increase in inequality.

Another econometric analysis from 2014, considering the data from 1950 to 2012 of the economies in the OECD, using the Gini coefficient to measure inequality, found that as "the income shares of the 20 % richest increases by 1 %, the GDP growth is 0.08 % lower in the following five years" similarly, "the increase of income by 1 % of the poorest 20 % is associated with 0.38 % of higher economic growth" (Dabla et al., 2015).

Moreover, Piketty in 2014 has found evidence that the low inequality from the 1950s to the 1970s resulted, first, from the high taxes on wealthy individuals in several developed economies. Second, economic growth archived because of technological development. Third, a demographic growth of the population. By the 1970s, Piketty uncovered the increase of inequality due to fiscal policies and shocks that simultaneously deaccelerate economic growth. This conclusion is substantially different from perceiving inequality as an evolution of the market economy described by Kuznets 60 years before. (Lyubimov,2017)

Since Kuznet, the effort to determine whether inequality affects economic growth or economic growth affects inequality has been inconclusive. However, according to Cingano, the new question is to establish the direction in which inequality affects growth; to solve this question, Cingano provides a summary of main cross-country reduced-form studies on inequality and growth from 17 different models where 10 found a statistically significant negative relationship between inequality and growth. However, the use of the Gini coefficient in most studios, the evidence of bias in using cross-sectional time series data, and the inefficacy of controlling for country-specific effects motivate the author to define a new model (2014).

The results archived by Cingano are that a "1 Gini point reduction in inequality would raise average growth by slightly more than 0.1 % points per year, with a cumulative gain in GDP at the end of the period of around 3%" (2014). A second model used to control for country-specific effects found that changes in top income inequality are found to have no statistically significant impact.

On the other hand, an empirical analysis of 2480 observations on 112 different countries from 1947 to 1994 found a linear regression with a positive relationship between growth and inequality while controlling the effects of education through time (Li & Zou,1998). The relation between fixed and random effects in the panel data used to measure and compare inequality undefine the unbias relation between growth and inequality.

A simultaneity problem

When an econometric model is described, its efficacy in estimating a causal relationship between variables is defined by multiple assumptions; for example, the explanatory variables are not correlated with unobserved factors, the sample data collected should be random, the explanatory variables are not correlated between each other, and the explanatory variables selected should be enough to explain the causality. However, the analysis provided by Cingano of previous models found:

1. Using different inequality estimators instead of Gini, like decil ratio estimator, can change the negative relation between inequality and growth to positive.
2. Dividing the countries into developed or undeveloped economies can change the relationship and effect between inequality and growth.

3. Using cross-sectional or time series data to estimate the model causes bias. This was determined by using a comparation of the prediction capacity in different models.
4. The explanatory variables are highly correlated; therefore, OLS estimator is not accurate and is mainly used.
5. Using Lagged levels of explanatory variables mitigates any simultaneity problem between growth and inequality.

These results encourage the necessity of rewriting the existing models to find unbiased estimators and understand the real effect of the variables. Cingano designs his model using panel lagged level data to mitigate that GDP may feedback inequality (simultaneity problem), choose GMM over OLS as a method for estimation to solve the multicollinearity problem of the explanatory variables and separates the data in developed and undeveloped economies in other to isolate an average effect of low growth for reasons related with develop. The model is the following:

$$lny_{i,t} = (1 + \alpha)lny_{i,t-1} + X_{i,t-1}\beta + \gamma Ineq_{i,t-1} + \mu_i + \epsilon_{i,t}$$

Cingano designed a model with high accuracy for predicting economic growth considering initial economic growth, human capital, inequality and controlling time-invariant variables for each country; however, he chose to use GMM and lagged level data to solve the simultaneity problem between growth and inequality without answering it. The conclusion is clear, the relationship between inequality and economic growth is negative, but the flow of causality stays inconclusive.

Objective

The aims of the present work are to discover the causality direction between inequality and growth; towards this aim, I will compute a Granger test using panel data for multiple countries, from the period 1964 to 2018.

Knowing if low economic growth rates are causing more inequality or more inequality is causing less economic growth is vital to orientate the designee of public policies, allowing them to allocate the efforts to one objective, increase growth or decrease inequality.

Methodology

The data set comprises annual measures of 45 countries of Gini and GDP (see Annex 1 for details). The Gini coefficient is an index of inequality based on comparing cumulative proportions of the population against cumulative proportions of the income they receive. It is defined as 0 for perfect equality and 1 for perfect inequality. The source of the inequality data is the OECD-IDD dataset. This dataset provides a Gini estimation using the equivalised household income, which is, the total income received by the households adjusted for household size with an equivalence scale. Additionally, this dataset considers as income data; only the cash income, excluding imputed components such as home production and imputed rents, and including earnings of the household members, self-employment income, capital income (rents, dividends, and interest), public transfers and household taxes. This dataset was selected to increase the GINI estimator's measurement capacity, avoiding under-reporting, misspecification, and non-consideration of fixed variables like the tax effect on the estimation of inequality.

The Economic growth is measured by the GDP per country, defined as the monetary value of final goods and services produced in a country in a given period of time. The GDP data set is measured using an expenditure approach with the base year 2015 and using millions of US dollars as a unit. The source is the OECD Annual National Accounts.

Unfortunately, the sample period selected from 1964-2018 to capture an external effect that could cause the rise of inequality in 1970, and the chosen sources provide a data set with missing values. The missing value percentage for GDP is 12%, and for Gini, 48%. (See Annex 2 for visualization); therefore, exist a necessity to estimate the missing values to test the causality of GDP and GINI.

How to deal with a high percentage of missing values in data?

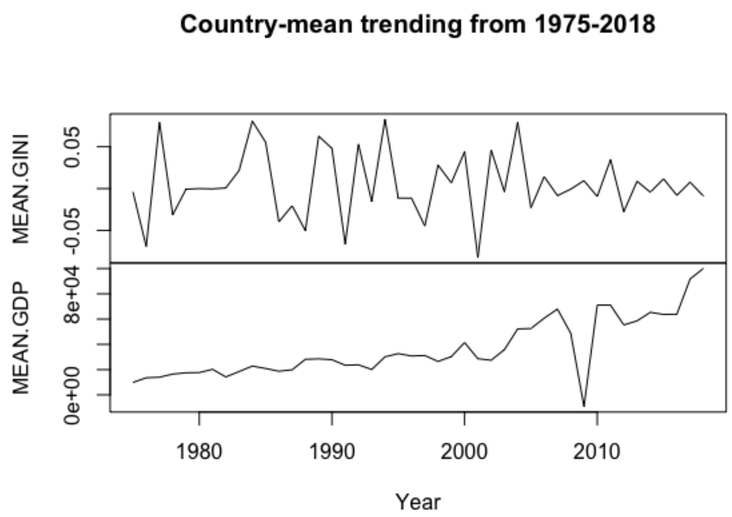
Empirical Orthogonal Function (EOF) is commonly used in the social sciences to describe, reconstruct, and predict highly dimensional data fields when data contain a high percentage of missing values; this method uses a basic principle, preserve the power of explanation, in other words, the estimated data must explain itself (Beckers and Rixen, 2003). Exist three EOF approaches to estimate missing values, via least-squares estimation (LSEOF), via interpolation (DINEOF), and via interpolation but one observation at a time (RSEOF).

Even though LSEOF calculation are relatively easier than DINEOF and RSEOF, in multiple case studies the resulting values are not necessarily positive which overestimates the amount of explained variance (Beckers and Rixen, 2003). Moreover, LSEOF assumes that the variables are linearly correlated which is false for a panel data with Gini coefficients, therefore LSEOF cannot be used for this case.

The other two options, DINEOF and RSEOF, have been used to predict missing values successfully; however, RSEOF causes a loss in the orthogonality property of the data. Orthogonality guarantees that the effect of one interaction can be estimated separately from the effect of any other interaction in the model; in other words, it assures that the independent variables are genuinely independent (Taylot et al., 2014). Therefore, DINEOF is the best model to estimate the missing values in the presented data. DINEOF was performed using a procedure designed by Taylor in 2014, to accurately reconstruct data using interpolation for a converge level of 1e-02, which states for, higher resolved interpolation, returning a complete data set for Gini coefficients and GDP (to visualize see Anex 3).

First impressions of causation

According to philosopher David Hume, “Causation is a relation between phenomenon that we employ in our reasoning to yield less that demonstrative knowledge beyond our immediate impressions” (Buehner, 2014). The first impression between Gini and GDP is that GDP causes Gini to change because Gini depends on overall income, and GDP changes the income in an economy; hence, if we can observe the change of Gini and GDP over time in the same graph, we have to be able to see a tendency between them, how they change together, but to integrate 45 countries in one graph with two variables will result in a misleading of our impressions; therefore, a simple way of archive this graph will be to get the average change of Gini and GDP per year.



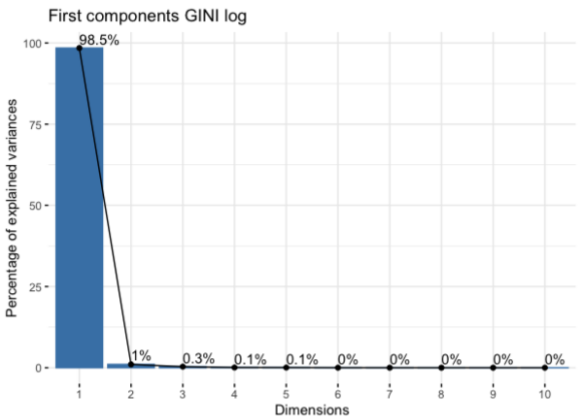
Source: Author using R (R-code is provided in Annex 5)

Following Hume’s definition of causation, this result fails to archive demonstrative knowledge because there is no apparent tendency between them. Consequently, it is necessary to find another way to compact the variance of Gini and GDP in different countries over the years; the answer is using Principal component analysis (PCA).

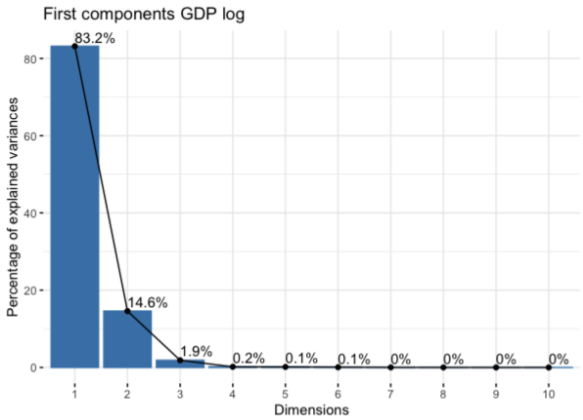
PCA is a technique for reducing the dimensionality of large datasets, minimizing information loss. It does so by computing new uncorrelated linear variables called components, solving an eigenvalue/eigenvector problem (Jolliffe & Cadima, 2016). In order to find the components in the Gini and GDP data set, we first need to adjust the measurement between the variables using a logarithmic and differential function, calculate a correlation matrix and solve the eigenvalue/eigenvector problem. As a result, we get a new variable that we can use to follow the change over time of Gini and GDP in our 45 countries.

It is important to mention that these new components on their own do not have a significant meaning; for example, Gini for Australia in 2016 was 0.33, which means that the better-off are, on average, 1.98 times richer than the less well-off (to see this calculation see Annex 4). On the other hand, the first component of the Gini for Australia is 1.3962 without any meaning for income distribution; it only compacts the linear variances between 1974 and 2018 of the Gini in Australia to be able to see a time tendency.

The percentage of explained variance archived in the calculation of this components it is the following using logarithmic normalization.

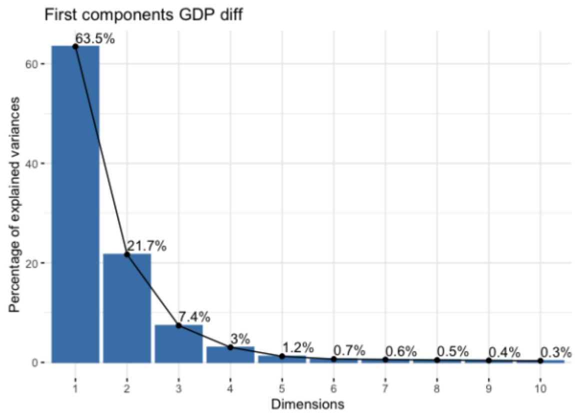


Source: Author using R

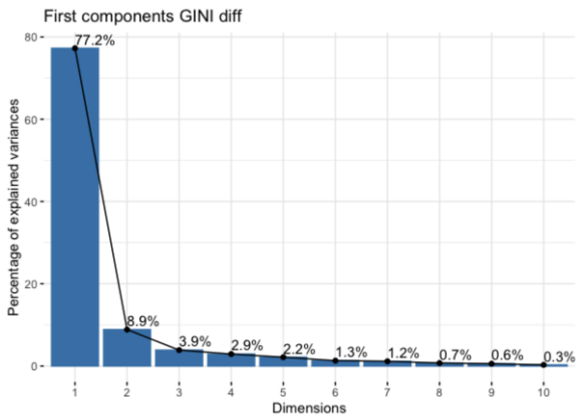


Source: Author using R

The percentage of explained variance archived in the calculation of this components it is the following using differentiation as normalization.

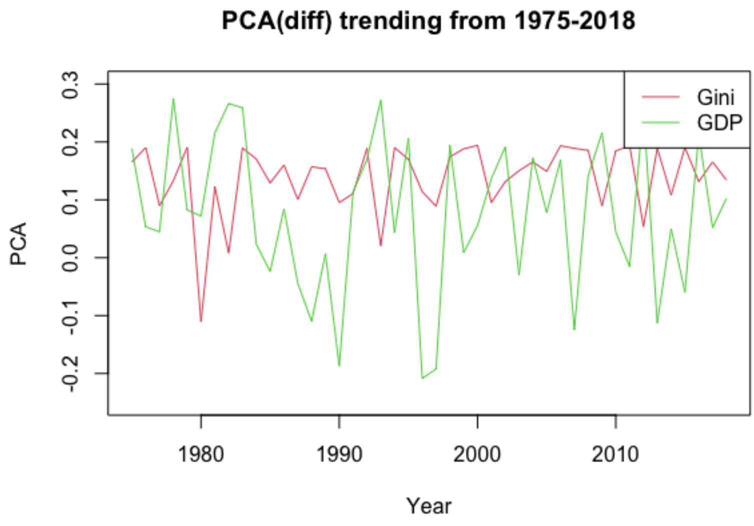


Source: Author using R

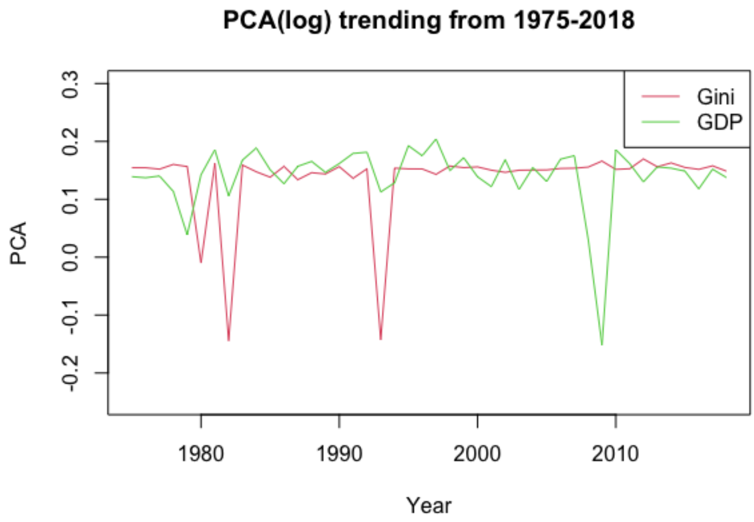


Source: Author using R

If we graph this first components in a same graph to perceive the tendency between Gini and GDP we get:



Source: Author using R



Source: Author using R

This time the graphs are more demonstrative. If we use differentiation to normalize our data, we can notice that Gini tends to follow GDP changes over time, and the variable Gini is not enough to explain all the changes in GDP because of the disparities over the years. On the other hand, using logarithmic to normalize our data, we recognize the opposite causality direction; this time, GDP follows more times the changes in Gini than the other way around, and surprisingly, logarithmic normalization tends to be the most used technique to normalize data in econometric models that conclude that a change in the Gini coefficient affects GDP directly, using previous statement made by Cingano "1 Gini point reduction in inequality would raise average growth by slightly more than 0.1 % points per year, with a cumulative gain in GDP at the end of the period of around 3%" (2014) is now in doubt.

Finding demonstrative knowledge

The next step in Hume's definition of causation is to find demonstrative knowledge; according to Hume, a phenomenon can be called 'causal' if it meets the following requirements: First, Cause precedes effect in time; Second, Cause includes information about the effect that is not available in a broad group of other variables (Maziarz, 2015). Granger, in 1980 used this assumption to find a statistics method using the following axioms:

(1) The past and the present may cause the future, but the future cannot cause the past.

(2) All the knowledge available contains no redundant information, so that if some variable is functionally related to one or more other variables, in a deterministic fashion, then the variable should be excluded from the universe.

(3) All causal relationships remain constant in direction throughout time." (Granger 1980)

Since then, the Granger-Causality test has been one of the most used approaches to understanding causality in economics; however, using an infinite set of information proposed by this axiom seems challenging. Therefore, many economists like Clive W. J have adopted the method using Probabilistic Theory; the reasoning is the following: if the conditional probability of B given A is greater than B alone, and A occurs before B over time, then A causes B (Maziarz, 2015). In this sense, the Granger-Causality is an opportunity to test if Gini causes GDP or GDP causes Gini.

Granger test

In order to use the classical Granger test formulation, it was used the first component of Gini and GDP with logarithmic and differentiation normalization, but not both at the same time, because, according to Kónya, Granger test overestimate the causality relation between variables if both methods are used in non-stationary variables as GDP and Gini (2006). A non-stationary variable is a data which means, variances, and covariances change over time, like trends or cycles.

Results

The first two granger test computed, to name it, Granger-Test PCA log and Granger-Test PCA diff, use the following null and alternative hypotheses: Null Hypothesis, Time series Gini does not Granger-cause time series GDP. Alternative Hypothesis, Time series Gini Granger-cause time series GDP.

The procedure uses two models per test; the first model for the Granger-Test PCA log attempts to predict the logarithmic GDP using the logarithmic GDP in the previous year and the logarithmic of Gini in the previous year as predictor variables. The second model attempts to predict the logarithmic GDP using only the logarithmic Gini in the previous year. The results are the following:

```
Granger causality test

Model 1: GDP ~ Lags(GDP, 1:1) + Lags(GINI, 1:1)
Model 2: GDP ~ Lags(GDP, 1:1)
  Res.Df Df      F Pr(>F)
1     40
2     41 -1 0.4025 0.5294
```

Source: Author using R

The F test statistic is 0.4025, and P-value is 0.5294; since P-value is more than 0.5, we fail to reject the null hypothesis of the test and conclude that logarithmic previous year's Gini does not Granger-cause logarithmic GDP, however, if we keep doing the same procedure but changing the previous year for 2, 3, 4 and 5 years the summary of the results is the following:

Year	F test statistic	P-value	Ho result
Previous year Gini	0.4025	0.5294	Fail to reject
Two years ago Gini	0.6265	0.5401	Fail to reject
Three years ago Gini	0.4961	0.6874	Fail to reject
Four years ago Gini	0.5311	0.7138	Fail to reject
Five years ago Gini	0.361	0.8707	Fail to reject

Source: Author

It becomes evident that when the first component of logarithmic Gini is used to explain GDP, no significant casualisation is found. If we run the same test but in the opposite direction to answer, is Logarithmic GDP causing Logarithmic Gini to change over the years; we get the following results, for the new Null Hypothesis, Time series GDP does not Granger-cause time series Gini and Alternative Hypothesis, Time series GDP Granger-cause time series Gini:

Year	F test statistic	P-value	Ho result
Previous year GDP	0.1575	0.6936	Fail to reject
Two years ago GDP	0.1149	0.8917	Fail to reject
Three years ago GDP	0.3207	0.8103	Fail to reject
Four years ago GDP	0.2465	0.9096	Fail to reject
Five years ago GDP	0.2128	0.9542	Fail to reject

Source: Author

From these results, two outcomes are possible; the first option, the logarithmic PCA of Gini and GDP, is useless to explain the causality relation between Gini and GDP, and the second option does not exist any causality relation between Gini and GDP. More tests are need it to find any conclusion; this time, we will use differentiation normalization for our PCA calculation, as suggested by Maziarz (2015).

For Null Hypothesis, Time series Gini does not Granger-cause time series GDP and Alternative Hypothesis, Time series Gini Granger-cause time series GDP, the summary of the results is the following:

Year	F test statistic	P-value	Ho result
Previous year Gini	0.382	0.54	Fail to reject
Two years ago Gini	0.3885	0.6808	Fail to reject
Three years ago Gini	0.2297	0.8751	Fail to reject
Four years ago Gini	0.4052	0.8034	Fail to reject
Five years ago Gini	0.4011	0.8439	Fail to reject

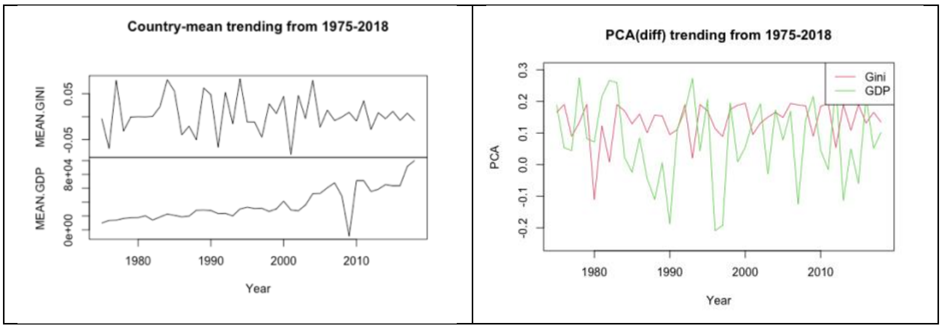
Source: Author

In the other direction, the Null Hypothesis is Time series GDP does not Granger-cause time series Gini and Alternative Hypothesis, Time series GDP Granger-cause time series Gini, the summary of the results is the following:

Year	F test statistic	P-value	Ho result
Previous year GDP	0.6287	0.4325	Reject
Two years ago GDP	1.5175	0.2326	Reject
Three years ago GDP	0.9113	0.4458	Reject
Four years ago GDP	0.7914	0.5397	Fail to reject
Five years ago GDP	0.9399	0.4706	Reject

Source: Author

From these results, two outcomes are possible: the first option, GDP causes Gini to change, and the second option using PCA to analyse the causal relationship between GDP and Gini is useless, given that previous researchers, like Piketty, T. (2014) found a prove of Gini causing changes in GDP. Therefore, more tests need it to set a conclusion. Even though PCA was perceived as more trending than Mean in the visuals timelines between PCA and Mean, we will test for Mean, too, to confirm our visual perception.



For Null Hypothesis, Time series Gini does not Granger-cause time series GDP and Alternative Hypothesis, Time series Gini Granger-cause time series GDP, the summary of the results is the following:

Year	F test statistic	P-value	Ho result
Previous year Gini	16.523	0.0002122	Reject
Two years ago Gini	6.7888	0.003014	Reject
Three years ago Gini	37.449	5.048e-11	Reject
Four years ago Gini	5.5782	0.001597	Reject
Five years ago Gini	3.6651	0.01083	Reject

Source: Author

In the other direction, the Null Hypothesis is Time series GDP does not Granger-cause time series Gini and Alternative Hypothesis, Time series GDP Granger-cause time series Gini, the summary of the results is the following:

Year	F test statistic	P-value	Ho result
Previous year GDP	4.317	0.04404	Reject
Two years ago GDP	1.5829	0.2186	Reject
Three years ago GDP	2.0153	0.1297	Reject
Four years ago GDP	1.5638	0.2078	Reject
Five years ago GDP	1.9257	0.1205	Reject

Source: Author

This time, we have found more possible outcomes. First, Gini and GDP simultaneity explain each other. Second, given the level of statistical significance found, the causal direction follows more the direction Gini causes GDP than GDP causes Gini. Third, the Mean is not useful to explain the causality between Gini and GDP, which is unlikely because previous research has found a simultaneity causal relation between Gini and GDP, like Cingano (2014). Fourth, trying to compact the variation of Gini and GDP through time from different countries with PCA and Mean leads to wrong results. Consequently, it is necessary to compute more tests, this time using a Panel version of the Granger test, intending to analyse the causality direction country by country.

Following Lopez and Weber extension of granger test for panel data, the results are the following:

Year	P-value	Null Hypothesis (Ho)	Ho result
logarithmic Gini	8.164e-05	Gini does not Granger cause GDP for all countries	Reject
Differential Gini	0.003699	Gini does not Granger cause GDP for all countries	Reject
logarithmic GDP	2.2e-16	GDP does not Granger cause Gini for all countries	Reject
Differential GDP	0.0008154	GDP does not Granger cause Gini for all countries	Reject

Source: Author

Conclusions

After running the Granger test, giving different specifications, we can only say one thing for certain, Gini and GDP are simultaneously causing changes in each other. However, there are some matters to consider:

- Because the Gini data used has a high percentage of missing values (48%), and an interpolation method was used to predict the missing values, we are giving more explanatory power to the variable to explain itself than the one that could truly have. Therefore, our causality granger test finds a weaker relationship between the variables.
- The findings of bi-directional or simultaneity causality in the granger test can also mean that an additional variable causing changes in GDP and Gini is more relevant (Maziarz, 2015).
- Using logarithmic and differentiation transformations is not enough to normalise the data and make them stationary. In this case, all the founded results will be false.

In consequence, further research is needed to find the proper conclusions. One of the best options will be to use a more powerful approach to re-construct data for Gini, like the techniques developed by the Pooled Mean Group (PMG). This approach allows us to deal with parameter heterogeneity and to separately estimate short- and long-run coefficients for each growth determinant and compute the Granger test again. A second option is to use a different data set like the one provided by the Luxembourg Income Survey. However, only 20 countries could be analysed with this option, A third option is to use a different transformation than logarithmic, and differentiation should be used to detrending the variables and compute the Granger test again. Finally, following Piketty's line of research the effect of public policies in each economy could be added as a third variable to explains the changes in Gini and GDP.

Because the results are considered inconclusive, I encourage public policies creators to develop strategies for decreasing inequality at the same time that economic growth is archived, mainly because exist the possibility that from 1974 to 2018, the levels of inequality on average were not high enough to cause a persistent negative effect on economic growth, however, given the levels of rapid increase of inequality this could change, and the efforts of increased economic growth will not be enough given the new levels of inequality as has been suggested by Stiglitz (2013).

Annex 1

List of countries used in data set:

- Australia
- Austria
- Belgium
- Brazil
- Bulgaria
- Canada
- Chile
- China
- Costa Rica
- Czech Republic
- Denmark
- Estonia
- Finland
- France
- Germany
- Greece
- Hungary
- Iceland
- India
- Ireland
- Israel
- Italy
- Japan
- Korea
- Latvia
- Lithuania
- Luxembourg
- Mexico
- Netherlands
- New Zealand
- Norway
- Poland
- Portugal
- Romania
- Russia
- Slovak Republic
- Slovenia
- South Africa
- Spain
- Sweden
- Switzerland
- Türkiye
- United Kingdom
- United States

Annex 2

GDP original data-set visualization of the first rows:

Year	Australia	Austria	Belgium	Brazil	Bulgaria	Canada	Chile	China	Costa Rica	Czech Republic	Denmark
1974	53532.1	53036.0	60030.8			123692.3		45406.7			30946.4
1975	62617.1	56255.6	66436.3			139138.2		48807.8			34568.6
1976	72274.6	62140.4	75509.0			161079.3		47990.5			40051.3
1977	78926.3	68691.7	81663.8			177982.1		52188.0			44421.7
1978	89213.3	72655.1	87699.4		1648.1	197120.5		59072.0			49477.1
1979	101128.4	79696.4	93796.1			224907.0		65845.2			55024.1
1980	114460.3	85600.9	102007.0			252412.8		73666.9	1682.0		59582.8
1981	132114.3	91218.6	106949.8			288052.8		79258.3			65515.2
1982	142296.9	98036.5	115730.9			303554.3		86285.2			74820.3
1983	160533.1	104457.1	122607.9			329465.6		96682.6			82432.3
1984	176916.6	109683.7	132464.6			361269.1		116877.0			90983.6
1985	195976.2	115755.4	140884.0	147464.0		391016.9		146110.3		29725.4	98686.7

Source: OECD Annual National Accounts US$ millions 2015

Gini original data-set visualization of the first rows:

Year	Australia	Austria	Belgium	Brazil	Bulgaria	Canada	Chile	China	Costa Rica	Czech Republic	Denmark
1974											
1975											
1976						0.304					
1977						0.288					
1978						0.293					
1979						0.289					
1980						0.289					
1981						0.287					
1982						0.291					
1983						0.299					
1984						0.296					
1985						0.293					0.221
1986						0.292					0.224

Source: OECD-IDD

Annex 3

Gini complete using DINEOF, data-set visualization of the first rows:

Year	Australia	Austria	Belgium	Brazil	Bulgaria	Canada	Chile	China	Costa Rica	Czech Republic	Denmark
1974	0.337	0.274	0.13457143	0.40150682	0.369	0.312	0.38592521	0.161155219	0.485	0.257	0.256
1975	0.33	0.284	0.13708645	0.481	0.402	0.307	0.3931894	0.166535966	0.484	0.253	0.261
1976	0.326	0.276	0.13159325	0.39290319	0.356	0.317	0.37735365	0.156178914	0.483	0.254	0.249
1977	0.325	0.28	0.258	0.39977126	0.408	0.304	0.3843876	0.161117599	0.479	0.249	0.263
1978	0.3235834	0.281	0.14874893	0.482	0.34	0.313	0.471	0.514	0.48	0.257	0.251
1979	0.317	0.23173435	0.11224432	0.33639661	0.29860643	0.315	0.3217315	0.126895872	0.38286806	0.22659133	0.227
1980	0.31676445	0.275	0.14439058	0.42954678	0.395	0.31	0.46	0.179193918	0.48	0.249	0.264
1981	0.315	0.26699458	0.287	0.3924466	0.34431228	0.321	0.3767322	0.155074329	0.44337824	0.267	0.26617955
1982	0.31499206	0.275	0.14318515	0.4262132	0.377	0.318	0.454	0.176436085	0.479	0.258	0.263
1983	0.309	0.20925116	0.10094096	0.30278296	0.26958094	0.289	0.28930351	0.112804633	0.34526949	0.20563194	0.215
1984	0.30601943	0.28	0.13791759	0.469	0.354	0.319	0.465	0.166160967	0.494	0.259	0.254
1985	0.29703697	0.289	0.13336495	0.485	0.33	0.316	0.48	0.159056621	0.44886795	0.257	0.26438186

Source: Author using R

GDP complete using DINEOF, data-set visualization of the first rows:

Year	Australia	Austria	Belgium	Brazil	Bulgaria	Canada	Chile	China	Costa Rica	Czech Republic	Denmark
1974	53532.06	53035.99	60030.79	24213.88	-6024.44625	123692.3	-6742.2612	45406.75	-726.7053	-9002.8006	30946.4
1975	62617.09	56255.56	66436.31	20186.51	-5994.29018	139138.2	-5767.862	48807.8	-895.8369	-6653.5888	34568.56
1976	72274.59	62140.4	75508.99	21301.72	-6382.63641	161079.3	-5628.5426	47990.45	-983.2474	-6014.0141	40051.28
1977	78926.28	68691.73	81663.76	25524.86	-6770.0456	177982.1	-5551.0004	52187.97	-967.1231	-5684.2556	44421.74
1978	89213.33	72655.06	87699.41	36543.38	-6915.29483	197120.5	-5067.6491	59071.97	-709.5719	-5058.2253	49477.12
1979	101128.38	79696.4	93796.08	38148.79	-7007.39332	224907	-4604.2374	65845.16	-768.949	-3415.621	55024.07
1980	114460.31	85600.9	102007.03	42198.9	-6775.45438	252412.8	-2889.7579	73666.93	-740.9332	-438.0358	59582.82
1981	132114.26	91218.62	106949.84	55153.15	-6512.96063	288052.8	-1910.0271	79258.27	-372.7723	1364.0216	65515.24
1982	142296.89	98036.53	115730.9	63988.8	-6086.65054	303554.3	126.3909	86285.18	-161.3616	3938.1398	74820.3
1983	160533.07	104457.06	122607.86	76764.28	-5397.28343	329465.5	2071.0575	96682.63	221.9316	6980.3446	82432.26
1984	176916.63	109683.66	132464.6	92825.99	-4933.41066	361269.1	3427.1093	116876.97	694.869	9283.2913	90983.62
1985	195976.19	115755.43	140884.02	103799.52	-4585.43376	391016.9	5388.3932	146110.26	973.2042	12181.5453	98686.75

Source: Author using R

Annex 4

ANOTHER WAY TO INTERPRET THE GINI COEFFICIENT

The Gini coefficient (g) is a measure of inequality, calculated as one-half times the average of the income differences between people in the population, divided by the average income of the population:

$$g = \frac{1}{2}\left(\frac{\text{Average difference in income}}{\text{Average income}}\right)$$

The Gini coefficient ranges from 0 (perfect equality) to 1 (extreme inequality), where higher numbers indicate more-unequal distributions.

But there is a more interesting interpretation of the Gini coefficient, which we can derive mathematically.

Considering all possible pairs in a population of size n, the mean income of that population ($\overline{y}$) can be rewritten as

$$\begin{aligned}\bar{y} &= \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n} y_i^r + \frac{1}{n}\sum_{i=1}^{n} y_i^p\right) \\ &= \frac{1}{2}(\bar{y}^r + \bar{y}^p)\end{aligned}$$

where y^r is the income of the richer of the pair and y^p is the income of the poorer of the pair. So we can now rewrite the expression for the Gini coefficient in terms of the average incomes of the richer and poorer of each of the pairs, $\overline{y}^r$ and $\overline{y}^p$:

$$g = \left(\frac{1}{2}\right)2\left(\frac{\bar{y}^r - \bar{y}^p}{\bar{y}^r + \bar{y}^p}\right)$$

We can rearrange this expression to get a formula in terms of the ratio of average incomes of the richer and poorer of the pair ($\frac{\bar{y}^r}{\bar{y}^p}$), as shown below:

$$\begin{aligned} g(\bar{y}^r + \bar{y}^p) &= \bar{y}^r - \bar{y}^p \\ g\left(\frac{\bar{y}^r}{\bar{y}^p} + 1\right) &= \frac{\bar{y}^r}{\bar{y}^p} - 1 \\ g\frac{\bar{y}^r}{\bar{y}^p} + g &= \frac{\bar{y}^r}{\bar{y}^p} - 1 \\ g + 1 &= \frac{\bar{y}^r}{\bar{y}^p}(1 - g) \\ \frac{g+1}{1-g} &= \frac{\bar{y}^r}{\bar{y}^p} \end{aligned}$$

Using the final expression, if the Gini coefficient (g) is 0.62, then

$$\frac{\bar{y}^r}{\bar{y}^p} = \frac{g+1}{1-g} = \frac{0.62+1}{1-0.62} = 4.26$$

Source: Tipoe, Eileen 2020.

Annex 5

```r
library(sinkr)
library(FactoMineR)
library(ggcorrplot)
library(corrr)
library(usethis)
library(devtools)
library(tidyverse)
library(ggplot2)
#install_github("kassambara/factoextra")
library(factoextra)
library(lmtest)

Gini <- as.matrix(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/GINI.csv",na.strings = ""))
Gini <- Gini[,-1]

#The dineof interpolated field
set.seed(1)
GiniDineof <- dineof(Gini, delta.rms = 1e-02)# lower 'delta.rms' for higher resolved interpolation
Ginicomplete <- (GiniDineof$Xa)
GDP <- as.matrix(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/GDP.csv",na.strings = ""))
GDP <- GDP[,-1]
class(GDP)
set.seed(2)
GdpDineof <- dineof(GDP, delta.rms = 1e-02) # lower 'delta.rms' for higher resolved interpolation
Gdpcomplete <-(GdpDineof$Xa)

#PCA

#log for non-stationary
GDPDINEOFLOG <- as.matrix(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/GDPDINEOFLOG.csv",na.strings = ""))
GINIDINEOFLOG <- as.matrix(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/GINIDINEOFLOG.csv",na.strings = ""))

#Correlation matrix
corr_matrixGinilog <- cor(GINIDINEOFLOG)
corr_matrixGDPlog <- cor(GDPDINEOFLOG)
Ginilog.pca <- princomp(corr_matrixGinilog) #Creates Gini PCA with DINEOF estimated Gini
GDPlog.pca <- princomp(corr_matrixGDPlog)   #Creates GDP PCA with DINEOF estimated GDP

#Visualization of the principal components
Ginilog.pca$loadings[, 1:2]
GDPlog.pca$loadings[, 1:2]
#Percentage of explained variances
fviz_eig(Ginilog.pca, addlabels = TRUE, main = "First components GINI log")
fviz_eig(GDPlog.pca, addlabels = TRUE, main = "First components GDP log")

#first difference for a a stochastic trend Maziarz

GDPDINEOFDIFF <- as.matrix(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/GDPDINEOFDIFF.csv",na.strings = ""))
GINIDINEOFDIFF <- as.matrix(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/GINIDINEOFDIFF.csv",na.strings = ""))

#Correlation matrix
corr_matrixGinidiff <- cor(GINIDINEOFDIFF)
corr_matrixGDPdiff <- cor(GDPDINEOFDIFF)
Ginidiff.pca <- princomp(corr_matrixGinidiff) #Creates Gini PCA with DINEOF estimated Gini
GDPdiff.pca <- princomp(corr_matrixGDPdiff)   #Creates GDP PCA with DINEOF estimated GDP

#Visualization of the principal components
Ginidiff.pca$loadings[, 1:2]
GDPdiff.pca$loadings[, 1:2]
#Percentage of explained variances
fviz_eig(Ginidiff.pca, addlabels = TRUE, main = "First components GINI diff")
fviz_eig(GDPdiff.pca, addlabels = TRUE, main = "First components GDP diff")

#Granger test

#PCA log

PCAlog <- as.data.frame(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/PCAlog.csv",na.strings = "")) #Outside R this data was created using GINI 1stPCA and GDP 1stPCA

grangertest(GINI ~ GDP, order = 1, data = PCAlog)
grangertest(GINI ~ GDP, order = 2, data = PCAlog)
grangertest(GINI ~ GDP, order = 3, data = PCAlog)
grangertest(GINI ~ GDP, order = 4, data = PCAlog)
grangertest(GINI ~ GDP, order = 5, data = PCAlog) #The results change significantly

grangertest(GDP ~ GINI, order = 1, data = PCAlog)
grangertest(GDP ~ GINI, order = 2, data = PCAlog)
grangertest(GDP ~ GINI, order = 3, data = PCAlog)
grangertest(GDP ~ GINI, order = 4, data = PCAlog)
grangertest(GDP ~ GINI, order = 5, data = PCAlog)

#PCA diff

PCAdiff <- as.data.frame(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/PCAdiff.csv",na.strings = "")) #Outside R this data was created using GINI 1stPCA and GDP 1stPCA

grangertest(GINI ~ GDP, order = 1, data = PCAdiff)
grangertest(GINI ~ GDP, order = 2, data = PCAdiff)
grangertest(GINI ~ GDP, order = 3, data = PCAdiff)
grangertest(GINI ~ GDP, order = 4, data = PCAdiff)
grangertest(GINI ~ GDP, order = 5, data = PCAdiff) #The results change significantly

grangertest(GDP ~ GINI, order = 1, data = PCAdiff)
grangertest(GDP ~ GINI, order = 2, data = PCAdiff)
grangertest(GDP ~ GINI, order = 3, data = PCAdiff)
grangertest(GDP ~ GINI, order = 4, data = PCAdiff)
grangertest(GDP ~ GINI, order = 5, data = PCAdiff)
```

```
#Mean log
Meanlog <- as.data.frame(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/Meanlog.csv",na.strings = ""))

grangertest(MEAN.GINI ~ MEAN.GDP, order = 1, data = Meanlog)
grangertest(MEAN.GINI ~ MEAN.GDP, order = 2, data = Meanlog)
grangertest(MEAN.GINI ~ MEAN.GDP, order = 3, data = Meanlog)
grangertest(MEAN.GINI ~ MEAN.GDP, order = 4, data = Meanlog)
grangertest(MEAN.GINI ~ MEAN.GDP, order = 5, data = Meanlog)

grangertest(MEAN.GDP ~ MEAN.GINI, order = 1, data = Meanlog)
grangertest(MEAN.GDP ~ MEAN.GINI, order = 2, data = Meanlog)
grangertest(MEAN.GDP ~ MEAN.GINI, order = 3, data = Meanlog)
grangertest(MEAN.GDP ~ MEAN.GINI, order = 4, data = Meanlog)
grangertest(MEAN.GDP ~ MEAN.GINI, order = 5, data = Meanlog)

#Mean diff
Meandiff <- as.data.frame(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/Meandiff.csv",na.strings = ""))

grangertest(MEAN.GINI ~ MEAN.GDP, order = 1, data = Meandiff)
grangertest(MEAN.GINI ~ MEAN.GDP, order = 2, data = Meandiff)
grangertest(MEAN.GINI ~ MEAN.GDP, order = 3, data = Meandiff)
grangertest(MEAN.GINI ~ MEAN.GDP, order = 4, data = Meandiff)
grangertest(MEAN.GINI ~ MEAN.GDP, order = 5, data = Meandiff)

grangertest(MEAN.GDP ~ MEAN.GINI, order = 1, data = Meandiff)
grangertest(MEAN.GDP ~ MEAN.GINI, order = 2, data = Meandiff)
grangertest(MEAN.GDP ~ MEAN.GINI, order = 3, data = Meandiff)
grangertest(MEAN.GDP ~ MEAN.GINI, order = 4, data = Meandiff)
grangertest(MEAN.GDP ~ MEAN.GINI, order = 5, data = Meandiff)

#PANEL

#PANEL DINEOF LOG
PANELDINEOFLOG <- as.data.frame(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/PANELDINEOFLOG.csv",na.strings = ""))
library(plm)
pgrangertest(GINI ~ GDP, data = PANELDINEOFLOG)
pgrangertest(GDP ~ GINI, data = PANELDINEOFLOG)
pgrangertest(GINI ~ GDP, data = PANELDINEOFLOG, order = 2L)
pgrangertest(GDP ~ GINI, data = PANELDINEOFLOG, order = 2L)
pgrangertest(GINI ~ GDP, data = PANELDINEOFLOG, order = 2L, test = "Zbar")
pgrangertest(GDP ~ GINI, data = PANELDINEOFLOG, order = 2L, test = "Zbar")

#PANEL DINEOF DIFF
PANELDINEOFDIFF <- as.data.frame(read.csv("/Users/majohernandezdelprado/Desktop/Essex/Third year/DISSERTATION/PANELDINEOFDIFF.csv",na.strings = ""))

pgrangertest(GINI ~ GDP, data = PANELDINEOFDIFF)
pgrangertest(GDP ~ GINI, data = PANELDINEOFDIFF)
pgrangertest(GINI ~ GDP, data = PANELDINEOFDIFF, order = 2L)
pgrangertest(GDP ~ GINI, data = PANELDINEOFDIFF, order = 2L)
pgrangertest(GINI ~ GDP, data = PANELDINEOFDIFF, order = 2L, test = "Zbar")
pgrangertest(GDP ~ GINI, data = PANELDINEOFDIFF, order = 2L, test = "Zbar")

#Visualizing trendings

#PCAdiff

plot(PCAdiff$YEAR,main = "PCA(diff) trending from 1975-2018",
     PCAdiff$GINI,
     type = "l",
     col = 2,
     ylim = c(-0.25, 0.3),
     xlab = "Year",
     ylab = "PCA")
lines(PCAdiff$YEAR,                              # Draw second time series
      PCAdiff$GDP,
      type = "l",
      col = 3)
legend("topright",
       c("Gini", "GDP"),
       lty = 1,
       col = 2:3)

#PCAlog

plot(PCAlog$YEAR,main = "PCA(log) trending from 1975-2018",
     PCAlog$GINI,
     type = "l",
     col = 2,
     ylim = c(-0.25, 0.3),
     xlab = "Year",
     ylab = "PCA")
lines(PCAlog$YEAR,                               # Draw second time series
      PCAlog$GDP,
      type = "l",
      col = 3)
legend("topright",
       c("Gini", "GDP"),
       lty = 1,
       col = 2:3)
```

References

Alan B.Krueger(2012)."The Rise and Consequences of Inequality in the United States,"Remarks at theCenter for American Progress, January 12, 2012. Available at http://www.americanprogress.org/we content/uploads/events/2012/01/pdf/krueger.pdf.

Beckers, J-M, and M. Rixen. "EOF Calculations and Data Filling from "Incomplete Oceanographic Datasets." Journal of Atmospheric and Oceanic Technology 20.12 (2003): 1839-1856.

Bordo, M. D., & Meissner, C. M. (2012). Does inequality lead to a financial crisis? Journal of International Money and Finance, 31(8), 2147-2161.

Buehner, M. J. (2014). Time and causality. Frontiers in Psychology, 5, 228.

Cingano, F. (2014), "Trends in Income Inequality and its Impact on Economic Growth", OECD Social, Employment and Migration Working Papers, No. 163, OECD Publishing, Paris, https://doi.org/10.1787/5jxrjncwxv6j-en.

Dabla-Norris, M. E., Kochhar, M. K., Suphaphiphat, M. N., Ricka, M. F., & Tsounta, M. E. (2015). Causes and consequences of income inequality: A global perspective. International Monetary Fund.

Granger, Clive W. J. (1980), 'Testing for causality.A personal viewpoint', Journal of Economic Dynamic and Control, 2 (4), 329-352.

Haldane, Andy. (2019). Why does economic growth matter? Bankofengland.co.uk. https://www.bankofengland.co.uk/knowledgebank/why-does-economic-growth-matter

Hongyi Li & Heng-fu Zou, 1998. "Income Inequality Is Not Harmful for Growth: Theory and Evidence," CEMA Working Papers 74, China Economics and Management Academy, Central University of Finance and Economics.

Jolliffe Ian T. and Cadima, Jorge. (2016). Principal component analysis: a review and recent developmentsPhil. Trans. R. Soc. A.374:2015020220150202 http://doi.org/10.1098/rsta.2015.0202

Kónya, L. (2006). Exports and growth: Granger causality analysis on OECD countries with a panel data approach. Economic Modelling, 23(6), 978-992.

Kuznets, S. (1955). Economic Growth and Income Inequality. The American Economic Review, 45(1), 1–28. http://www.jstor.org/stable/1811581

Lopez, L., & Weber, S. (2017). Testing for Granger causality in panel data. The Stata Journal, 17(4), 972-984.

Lyubimov, I. (2017). Income inequality revisited 60 years later: Piketty vs Kuznets. Russian Journal of Economics, [online] 3(1), pp.42–53. doi:10.1016/j.ruje.2017.02.003.

Maziarz, M. (2015). A review of the Granger-causality fallacy. The journal of philosophical economics: Reflections on economic and social issues, 8(2), 86-105.

Piketty, T. (2014). El capital en el siglo XX1. Fondo de cultura económica. México, primera edición.

Sen, K. (2021, January 26). Why should I care about economic growth? UNU. Retrieved December 13, 2022, from https://www.wider.unu.edu/publication/why-should-i-care-about-economic-growth

Stigilitz, J. E. (2013). The price of inequality. Norton & Company, Inc. United States of America, second edition.

Taylor, M. H., Losch, M., Wenzel, M., & Schroter, J. (2014). *On the sensitivity of field reconstruction and prediction using Empirical Orthogonal Functions derived from gappy data*. Heitstrasse. https://epic.awi.de/id/eprint/33458/1/taylor-etal_gappy.pdf

Taylor, M. (2014, January 1). *Multivariate methods and handling of geographic datasets*. GitHub. Retrieved April 15, 2023, from https://github.com/marchtaylor/sinkr

Tipoe, E. (2020, December 14). *ANOTHER WAY TO INTERPRET THE GINI COEFFICIENT*. Harvard Online, LabXchange. https://www.labxchange.org/library/pathway/lx-pathway:947bd402-9d3a-40f0-bcac-31c660f34459/items/lx-pb:947bd402-9d3a-40f0-bcac-31c660f34459:html:f5176875