Identifying the Determinants of Income Inequality - 20224885

1. Introduction

The European Union (EU) has the second largest economy in the world, behind the United States (US). In 2019, the World Bank reported the EU had a gross domestic product (GDP) of 15,626,448.48 trillion US dollars. (World Bank 2021) GDP is the most frequent metric for measuring the overall size of a country or region's economy. GDP is calculated using three values, total output of goods and services produced, household and government expenditure and income. The higher the country's GDP, the stronger their economy is while yearly GDP increases suggest the economy is growing. The EU's GDP is approximately 8.5% greater than China and 27% less than the US. However, since 2005, the EU's GDP has greatly fluctuated, driven by the 2008 economic crash and 2015 oil price collapse while the US and China each maintained consistent growth. Despite this fluctuation, the EU's GDP has increased by 31% overall in that period. With this growth, income levels are disproportionally distributed towards the higher earners in society. (Mussini 2017) In 2018, Spanish and German CEOs had an average pay ratio of 143 and 136 to the average worker respectively. (Szmigiera 2021) Income distributions such as this has increased research in recent years into the relationship between income inequality and economic growth and its potential future impacts. (Michálek and Výbošťok 2019) Leow and Tan found economic growth influences income inequality both positively and negatively dependent on the stage of the economy. (Leow and Tan 2019)

Income inequality in the EU has traditionally been measured through averaging country trends or individually per country. (Filauro and Fischer 2021) The EU has maintained significantly lower levels of income inequality among citizens compared to the US, although inequality is higher than Australia and Japan, countries with clearly established welfare models. The EU's cross-country average income inequality level has remained relatively stagnant despite economic growth since 2007. However, these results are different when comparing individual inequality trends, particularly within different economic levels. Inequality levels are lower within EU-member states and higher between member states. (Filauro and Fischer 2021) For example, Bulgaria had a Gini coefficient of 39.6 in 2018, whereas in Ireland it was 28.9 in Ireland and 31.1 in Germany. The EU average in 2018 was 30.08. This is consequential of the EU formation where the Union is made up of countries with heterogeneous income levels joining at different stages. For example, Bulgaria having joined the EU in 2007 have the highest level of income inequality.

While the EU is not a federal state like the US, there is an understanding social and economic progress across member states is dependent on developments in other states and on how EU-level policies and institutions effect each member state. (Filauro and Fischer 2021) The EU single market is a primary example of this. While there are 27 countries officially in the EU, Iceland, Liechtenstein and Norway are part of the EU's single market through the European Economic Area (EEA). (Veld 2019) Although Switzerland is not part of the EU or EEA, they also have access to the single market. The single market establishes an economic space in which individual economic forces impact each country and region with differential implications for salary incomes, capital investments, enterprises. For example, foreign direct investment and trade. (Veld 2019) Therefore, compiling a collection of individual country's data is best suited to identify significant determinants of income inequality within the EU.

This paper aims to identify significant determinants of income inequality. A multiple linear regression model is created using data from 29 developed European countries with access to EU single market. The period studies is 2007 through 2018. Regression diagnostic tests are performed to ensure multicollinearity, heteroskedasticity and autocorrelation if present, are detected and resolved. Two models are created, one using the Gini coefficient as a dependent variable and one using the S80S20 quintile income ratio as a dependent variable. This is to verify the determinants of income inequality are consistent across metrics, but also to demonstrate potential different effects the explanatory variables may have on each metric. The explanatory variables used are related to economic growth, government expenditure, trade openness and foreign investment and education and unemployment.

2. Literature review

Minimising equitable distribution of income is a key aim for world governments. (Dabla-Norris *et al.* 2015) Although, income inequality in itself may not be a problem as it provides incentives for individuals to invest and compete to progress economically. Historically, income inequality begins to widen when an economy is in a transitional or developing stage. (Dabla-Norris *et al.* 2015) For example, rapid investment in education and infrastructure can immediately spur economic activity and capital growth, through encouraging foreign investment. This results in the qualified labour force taking advantage of increased income levels, i.e., increased wages. However, when the country completes this transition into a developed economy, it is critical to implement policies attempting to minimise inequality levels. While inequality provides incentives towards growth, sustained high levels of inequality, particularly inequality of outcomes, measured by income, expenditure or wealth can impose large social costs. (Dabla-Norris *et al.* 2015) Rooted inequality of outcomes undermines individual's choices of education and occupation. Additionally, sustained inequality does not generate the "right" incentives. In that, individuals lose confidence in governing institutions ultimately eroding social cohesion, therefore they are incentivised towards securing favourable treatment and protection. (Dabla-Norris *et al.* 2015) This includes rising nepotism, resource misallocation and potential corruption with attendant adverse long-term economy and social consequences.

The Lorenz Curve is generally used to demonstrate the economic relationship between individual's income and the total share of wealth. (Gastwirth 1971) The Lorenz Curve states the individual's relative share of wealth disproportionately increases as their income increases. This is demonstrated in the following figure:

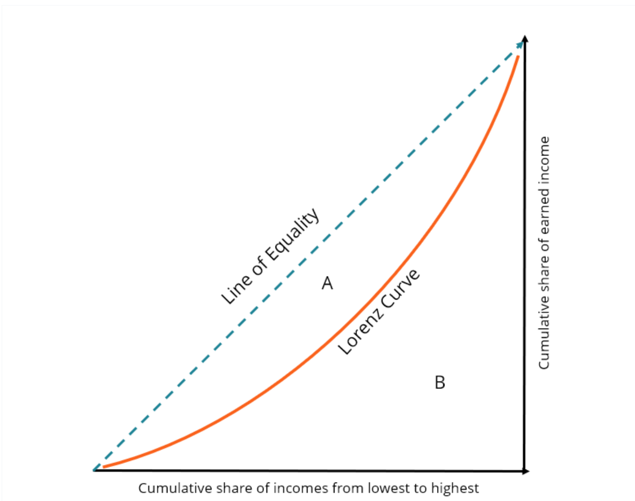


Figure 1 - Lorenz Curve Diagram (Corporate Finance Institute 2021)

For example, the amount of wealth held by the top 10% of the population is disproportionately larger than the wealth of the bottom 10%. One issue of the Lorenz Curve, however, is measuring individuals with a negative net worth. For example, individuals undertaking large debt would be considered poorer than an individual with a reduced salary but no debt. The most common metric used to universally measure income inequality is the Gini coefficient, measuring the distribution of income among all members of the population. (De Battisti *et al.* 2019) Values in the Gini coefficient range from 0 to 1, or 0% to 100%. A Gini coefficient rating of 0 represents perfectly equal distribution of income within the population. Whereas a Gini coefficient rating of 1 represents a perfectly unequal distribution of income within the population. In practice, the Gini coefficient ranges from 0.24 to 0.63 or between 24% and 63%. (Solt 2020) The primary benefit of the Gini coefficient is its independence of the economy's size and overall strength. This allows the results to be compared between differing economies to identify the underlying determinants. For example, while Germany and Bulgaria both possess developed economies, Germany have a significantly higher GDP. One drawback of the Gini coefficient is the sensitivity towards the middle of the distribution, or the middle-class within the economy.

Given the relative stability of the middle-class in developed economies, it is also important to measure and analyse both ends of the distribution. (Cobham and Sumner 2014) The income quintile share ratio, S80/S20 income ratio is calculated as the ratio of the total income of the top quintile, top 20% to the income received by the bottom quintile, bottom 20%. While the Gini coefficient measures income distribution, the S80/S20 ratio measures income concentration. Examining S80/S20 trends enables identification of the relationship between the top quintile and the bottom. (Cobham and Sumner 2014) For example, the influx of highly specialised employment opportunities and its effect on the lower class. The use of multiple metrics for income inequality minimises the risk of bias while ensuring a sufficient understanding of the underlying causation of results. (Trapeznikova 2019) Therefore, both the Gini coefficient and S80/S20 quintile ratio will be considered when examining existing literature to identify the relationship between income inequality and the following variables of economic growth, government expenditure, trade openness and investment and employment and education.

2.1 Gross Domestic Product

Kuznets (1955) argued the relationship between income inequality and economic growth can be demonstrated as an inverted-U shape. As an economy is developing, income inequality begins to widen. When the economy is developed, income inequality is reduced. Brueckner and Lederman (2015) found on average, each 1% increase in the Gini coefficient reduced GDP per capita by 1.1% across a five-year period. The cumulative effect, long-run amounts to a reduction of about 4.5%. These results suggest increasing income inequality reduces transitional GDP growth per capita, with a negative long-term effect on GDP per capita. Brueckner and Lederman (2015) then split their research into the periods, pre and post 1990 to reflect the Cold War ending using Latin American and Caribbean countries. Their overall results concluded in developing economies, increasing income inequality raised GDP per capita, while in developed economies, increasing income inequality decreased GDP per capita. Kolev and Niehues (2016) found the Gini coefficient was significant on GDP per capita across a five-year period at the 90% level when analysing OECD countries. Each unit increase in the Gini coefficient decreased GDP growth by 0.155.

(Bubbico and Freytag 2018) suggest in developed economies, high levels of income inequality are a primary driver of economic stagnation, wealth concentration, erosion of social cohesion and trust, decreasing quality of healthcare services and alienation of traditional politics. Balcilar *et al.* (2021) re-examination of existing income inequality and economic growth evidence concluded increasing income inequality to an average Gini coefficient threshold of 35.92 has a positive impact on economic growth, beyond this threshold, economic growth is negatively impacted. Economic growth until this threshold reduces income inequality through increasing minimum wage faster than or proportionately with average wage, progressive taxing support income redistribution. Ultimately, economic growth reduces unemployment, with less unemployment, governments generate more tax revenue which can be distributed to social protection and education.

Barro (2000) however, suggests that in the short-term, increasing inequality within a developed economy can have a significant positive relationship with subsequent economic growth. This is typically a result of economic growth through technological innovation and globalisation. Technological innovations such as automation have different impacts depending on the environment. For example, automation increases productivity of highly-skilled workers, while completely replaces the role of low-skilled workers. (Bubbico and Freytag 2018) Evidence of this is reflected within the S80/S20 quintile ratio, where the top 20%, primarily skilled workers average income disproportionately increases compared to the bottom 20%, primarily low-skilled workers or unemployed.

2.2 Government Expenditure

Bohn (1990) reiterated Barro (1979) tax smoothing paradigm suggesting government budget deficits would even out over time. However, Larch (2010) found persistent economic deficits have become an entrenched feature of fiscal policy through peacetime deficits accumulating over the years. Budget distribution conflicts can delay necessary reforms. Larch (2010) found countries with below average income inequality maintain lower budget deficits, with a higher share of government expenditure attributed to social protection and education. This result was consistent across analysis from five periods between 1960 and 2008, comparing both the Gini coefficient and quintile income ratios with government deficits.

Marrero and Rodríguez (2012) found within 23 developed European countries, government expenditure on social protection, healthcare and education are all significantly negatively correlated with income inequality measures. Therefore, increasing expenditure within each area reduces income inequality levels. Sánchez and Pérez-Corral (2018) analysed 17 developed economies within the EU between 2005-2014. The rate of government expenditure on social protection was significant at the 99% level, with each unit increase contributing to a 0.391 reduction in the Gini coefficient. Government expenditure on education and healthcare were not significant, although unit increases for each reduced the Gini coefficient by 0.330 and 0.317, respectively. Martinez-Vazquez *et al.* (2012) found government expenditure on social protection to be significant at the 99% level, with each unit increase reducing the Gini coefficient by 0.139. Education was significant at the 95% level, with each unit increase reducing the Gini coefficient by 0.134. Healthcare expenditure was significant at the 99% level, with each unit increase reducing income inequality by 0.695. These results were compiled from panel data of 150 countries in transitional, developing or developed stages between 1970 and 2009. Sylwester (2000) found government expenditure to be significant at the 99% level on GDP growth, with each unit increase in rate of government expenditure on education against GDP to increase GDP growth by 2.2646.

Government expenditure is generally targeted towards the lower-class or bottom deciles within the economy. Therefore, their equivalised disposable income is disproportionately increased compared to higher-income individuals and households receiving government expenditure. (Martinez-Vazquez *et al.* 2012) Larch (2010) found their analysis corroborated with existing research deriving income inequality increases the difficulty of fiscal discipline. Left-leaning governments tend to attribute higher shares of government expenditure to social security, although this also represents special interests such as corporations that benefit most from deficit-financed redistribution. A basic example is the automobile or technology industry, aiming to reduce income inequality through increasing the lower deciles of disposable income therefore increasing sales of more expensive, high-margin goods and services. Larch (2010) suggests the failure to regulate income distribution has unfavourable trade-offs in mounting political pressure to increase government expenditure to reduce income inequality during a period where the economic priority is to reduce spending, instead using tax revenue to minimise deficits.

2.3 Investment and Trade

(Dorn *et al.* 2018) global analysis findings suggest globalisation leads to overall income convergence. It has allowed emerging economies such as China and India to use their productivity advantages to catch up with primarily developed economies such as North America and Europe. However, when reviewing the effects of globalisation features including foreign trade and investment on the aforementioned developed economies, it appears to have widened income inequality. Globalising the workforce disproportionately moves low and medium skilled labour to developing economies has decimating the local workforce. Thus, negative net trade generates higher levels of unemployment and lower minimum wage. Fischer (2001) suggests trade openness in developed economies with a highly skilled workforce has widened income inequality. This is a result of disproportionate increases in highly skilled employment and elimination of low-skilled employment. (Jakobsson *et al.* 2006) found for both the period 1980s, and 1980s and 1990s, total trade as rate of GDP had significant positive impacts on the Gini coefficient at 99% level of significance.

Alderson and Nielsen (1999) found total FDI is the strongest positive influence on income inequality. As total FDI increases, the Gini coefficient increases. Mihaylova (2015) found rate of FDI on GDP was statistically significant at greater than 99%, each unit increase in rate of FDI on GDP increased the Gini coefficient by 0.396. It is important to note this rate of FDI on GDP is based on total FDI inflow, not accounting for the host economies total FDI outflow to calculate net FDI.

(Asteriou *et al.* 2014) suggest increasing total FDI directed towards high-skill sectors such as technology where the high-skill inwards FDI for transitioning or developing countries may be relatively low-skill outwards FDI for developed economies. This will increase demand for skilled labour in both countries, widening inequality within both economies when FDI is skill-based. Asteriou *et al.* (2014) when analysing 27 EU countries since 1995 using a random effects model found rate of FDI inward stock on total GDP was significant at the 95% level, with each unit increase increasing the Gini coefficient by 0.098. Trade openness was found to be insignificant; each unit increase in sum of total exports and total imports by rate of total GDP increased the Gini coefficient by 0.030. However, when breaking EU countries into sub-groups; EU-core, periphery countries and new members, trade openness was significant at the 95% level in reducing income inequality. Each unit increase in trade openness reduced the Gini coefficient by 0.120, 0.012 and 0.123 respectively across each sub-group. FDI was significant at the 95% level in increasing income inequality, each unit increase in FDI stock by rate of GDP increased the Gini coefficient by 0.0849, -0.155 and 0.288 across each sub-group, respectively. Overall, trade openness appears to present an equalising effect, while FDI has had the most influential impact income inequality within the EU. Trade policies such as the EU-single market reduce tariffs on inter-union foreign trade, minimising the income distribution differences between low and high-skilled workers.

2.4 Unemployment and Education

(Checchi and García-Peñalosa (2009) found unemployment rate is statistically significant at the 99% level on Gini coefficient with a positive impact, each unit increase in unemployment rate increases the Gini coefficient by 1.802. This analysis was performed using a panel of OECF countries in the period 1960-2000. (Checchi and García-Peñalosa (2009) suggest increasing minimum wage despite compressing the income distribution has a positive effect on income inequality. However, increasing minimum wage generally results in higher unemployment as smaller businesses do not generate enough revenue. Although, increasing minimum wage should reduce the gap between income distributions offsetting the effects of increasing unemployment rates on income inequality. (Castells-Quintana *et al.* (2015) found unemployment rates are significant at the 99% level on the decile ratio, with each unit increase in unemployment rate increasing the ratio by 4.572. This is the regression output for cross-sectional and panel analysis. Interestingly, unemployment was not significant on the dependent variable for either 1996, 2000 or 2007. This may be due to the recent rapid growth of the 10% income rate compared to lower deciles. Maestri and Roventini (2012) dscoveries support this, finding inequality is generally counter-cyclical and significant positive correlated with unemployment. Unemployment increases income inequality as it is generally lower-skilled, low-income individuals who are unemployed.

Kiselakova *et al.* (2020) found tertiary education attainment was statistically significant at the 95% level on Gini coefficient, for each unit increase in rate of population with tertiary education attainment, the Gini coefficient was increased by 0.2844. These results were observed from a fixed effects panel regression analysis on 28 EU countries between 2010 and 2018. Checchi and García-Peñalosa (2009) found the average years of education is not significant on the decile income ratios, in this case, each unit increase in average years of education increased the P90/P10 ratio by 13.951. P90/P10 is the ratio of top 10% income to bottom 10%, like S80/S20. As the years of experience in education increases, the rate of unemployed highly skilled individuals within the labour force increases, this allows organisations to stagnate income rates for skilled positions.

2.5 Expected Explanatory Variable Effects

Variable	Impact on Income Inequality – Gini Coefficient	Impact on Income Inequality – S80/S20 Income Ratio
GDP Growth	Low Significance – Decrease Income Inequality	Low Significance – Decrease income inequality
Social Protection Expenditure	High Significance – Decrease Income Inequality	High Significance – Decrease Income Inequality
Education Expenditure	High Significance – Decrease Income Inequality	High Significance – Decrease Income Inequality
Health Expenditure	High Significance – Decrease Income Inequality	High Significance – Decrease Income Inequality
Government Expenditure (Health, Social Protection, Education)	High Significance – Decrease Income Inequality	High Significance – Decrease Income Inequality
Education Attainment	Low Significance – Increase Income Inequality	Low Significance – Increase Income Inequality
Unemployment	High Significance – Increase Income Inequality	High Significance – Increase Income Inequality
Net FDI Inflow	High Significance – Increase Income Inequality	High Significance – Increase Income Inequality
Net Trade	High Significance – Increase Income Inequality	High Significance – Increase Income Inequality

Table 1 - Expected Explanatory Variable Effects

3. Data

The United Nations have classified 31 countries within the EU-15, EU-13 and other European countries as having developed economies. (United Nations 2021) Based on the research aims, panel data was compiled from the following 29 European developed countries for the years 2007-2018:

Developed European Countries	
1. Austria	2. Belgium
3. Bulgaria	4. Czech Republic
5. Cyprus	6. Estonia
7. Denmark	8. France
9. Finland	10. Hungary
11. Germany	12. Ireland
13. Iceland	14. Latvia
15. Italy	16. Luxembourg
17. Lithuania	18. Netherlands
19. Malta	20. Poland
21. Norway	22. Romania
23. Portugal	24. Slovenia
25. Slovakia	26. Sweden
27. Spain	28. United Kingdom
29. Switzerland	

Table 2 – European Countries with Developed Economies Analysed

Greece and Croatia were excluded from the dataset due to incomplete or inaccurate data. Eurostat and the World Bank were the primary sources of data. World Development Indicators and Education Statistics – All Indicators were accessed using the World Bank Open DataBank API.

As discussed during the literature review, the Gini coefficient and S80/S20 ratio are the primary metrics used for measuring income inequality levels. The Gini coefficient measures the income distribution among the population of a country. It ranges from 0 – 1, with 0 representing no income inequality and 1 representing the highest degree of income inequality. The Gini coefficient of equivalised disposable income was sourced from Eurostat. (Eurostat 2021) Eurostat collected the data through a European Statistics on Income and Living Conditions (EU-SILC) survey. The Gini rate of change was calculated by taking current year minus previous year divided by previous year. The income quintile share ratio (S80/S20) is the ratio of total income received by the top 20% of the population to the total income of the lowest 20% of the population. The equivalised disposable S80S20 ratio was sourced from Eurostat. (Eurostat 2020) The data was based on the EU-SILC. The S80S20 rate of change was calculated by taking current year minus previous year divided by previous year.

Data for government expenditure on education, social protection and healthcare was sourced from the World Bank Open DataBank. For each metric, the government expenditure is percentage of GDP. (World Bank 2021) Government expenditure on education includes expenses towards employee compensation, intermediate consumption, i.e., purchasing goods and services, social benefits, i.e., student transport and accommodation and capital investments, i.e., school facilities. This includes all levels of education. Government expenditure on social protection includes unemployment, pension, housing, child welfare and disability. Government expenditure on healthcare includes medical resources, hospital services, public health services and research and development. Government expenditure for education, social protection and healthcare were added using the gathered from World Bank (2021). It is expressed as rate of government expenditure on those categories on GDP

Gross domestic product (GDP) is the level of household spending, capital investment, government spending and net exports for the country on an annual basis. A higher GDP suggests a stronger economy, therefore annual GDP growth suggests the country's economy is expanding. GDP is the primary metric of economies overall strength. Data for overall GDP in million units and GDP growth in GDP change per year was sourced from the World Bank Open DataBank. (World Bank 2021) Foreign direct investment (FDI) inflow is the net value of foreign country organisations investing within the country minus the country's investments in foreign countries. It is represented in million units valued using the US Dollar. Net trade is the countries' total exports of goods and services divided by total imports of goods and services. It is represented in thousand units valued using the US Dollar. The data for FDI inflow and net trade were sourced from the World Bank Open DataBank. (World Bank 2021) The values were calculated using the International Monetary Fund (IMF) payments database supplemented by data from the UN Conference on Trade and Development.

Unemployment is measured in the rate of the total labour force not currently in paid or self-employment but available for work during the reference time. The total labour force is the number of persons aged between 15 and 64 years. Unemployment data was sourced from the World Bank. (World Bank 2021) The values are a modelled ILO estimate using the International Labour Organisation (ILOSTAT) database. School enrolment in tertiary education and tertiary attainment rates were chosen for this model based on the literature review findings. School enrolment data was sourced from the World Bank education statistics. (DataBank 2020) The values represent the gross rate of school enrolment in tertiary education. The statistics were retrieved from the UNESCO Institute for Statistics. The tertiary education attainment represents the percentage of 30–34-year-olds who have attained a tertiary education. Educational attainment refers to International Standard Classification of Education (ISCED) 1997 for statistics up to 2013 and ISCED 2011 levels 5-8 from 2014 onwards. The data was collected through an EU Labour Force Survey. (Eurostat 2021) Education attainment change was calculated using the education attainment data, by taking current year minus previous year divided by previous year. The following table condenses the descriptions provided here:

Variable Name	Variable	Description	Source
gini	Gini coefficient	Measures the equivalised disposable income distribution among the population of a country. It ranges from 0 – 1, with 0 representing no income inequality and 1 representing the highest degree of income inequality.	Eurostat
GiniChange	Annual rate of change of Gini coefficient	Measures the annual rate of change in Gini coefficient ((current year - previous year) / previous year)	Eurostat
s80s20	S80S20 Income Ratio	The equivalised disposable income quintile share ratio (S80/S20) is the ratio of total income received by the top 20% of the population to the total income of the lowest 20% of the population.	Eurostat
s80s20Change	Annual rate of change of S80S20 Income Ratio	Measures the annual rate of change in S80.S20 ratio ((current year - previous year) / previous year)	Eurostat
gdp	Gross Domestic Product	Gross domestic product (GDP) is the level of household spending, capital investment, government spending and net exports for the country on an annual basis. A higher GDP represents a stronger economy	World Bank
gdpGrowth	Annual Rate of Change in GDP	Measures the annual rate of change in GDP. Positive GDP growth represents increasing economy while negative GDP growth represents shrinking economy ((current year - previous year) / previous year)	World Bank
eduEXP	Government expenditure on education as % of GDP	Government expenditure on education includes expenses towards employee compensation, intermediate consumption, i.e., purchasing goods and services, social benefits, i.e. student transport and accommodation and capital investments, i.e. school facilities. This includes all levels of education.	World Bank
healthEXP	Government expenditure on health as % of GDP	Government expenditure on healthcare includes medical resources, hospital services, public health services and research and development.	World Bank
soprotEXP	Government expenditure on social protection as % of GDP	Government expenditure on social protection includes unemployment, pension, housing, child welfare and disability.	World Bank
govEXP	Government expenditure (health, social, education) as % of GDP	Sum of Social Protection, healthcare and education government expenditure	World Bank
FDIinc	Foreign Direct Investment net inflow	Foreign direct investment (FDI) inflow is the net value of foreign country organisations investing within the country minus the country's investments in foreign countries. It is represented in million units valued using the US Dollar	World Bank
netTrade	Net Trade	Net trade is the countries' total exports of goods and services divided by total imports of goods and services. It is represented in thousand units valued using the US Dollar.	World Bank
unemployment	% of labour force unemployed	Unemployment is measured in the % of the total labour force not currently in paid or self-employment but available for work during the reference time. The total labour force is the number of persons aged between 15 and 64 years.	World Bank
eduENROLL	% of school enrollment in tertiary education	Measures the gross rate of school enrolment in tertiary education	World Bank
eduATTAIN	% of 30–34-year-old with tertiary education attainment	Measures the percentage of the population aged 30-34 who have successfully completed tertiary education	Eurostat
eduATTAINchange	Yearly change in education attainment	Measures the annual rate of change in education attainment ((current year - previous year) / previous year)	Eurostat

Table 3 – Table of Data

3.1 Descriptive Values

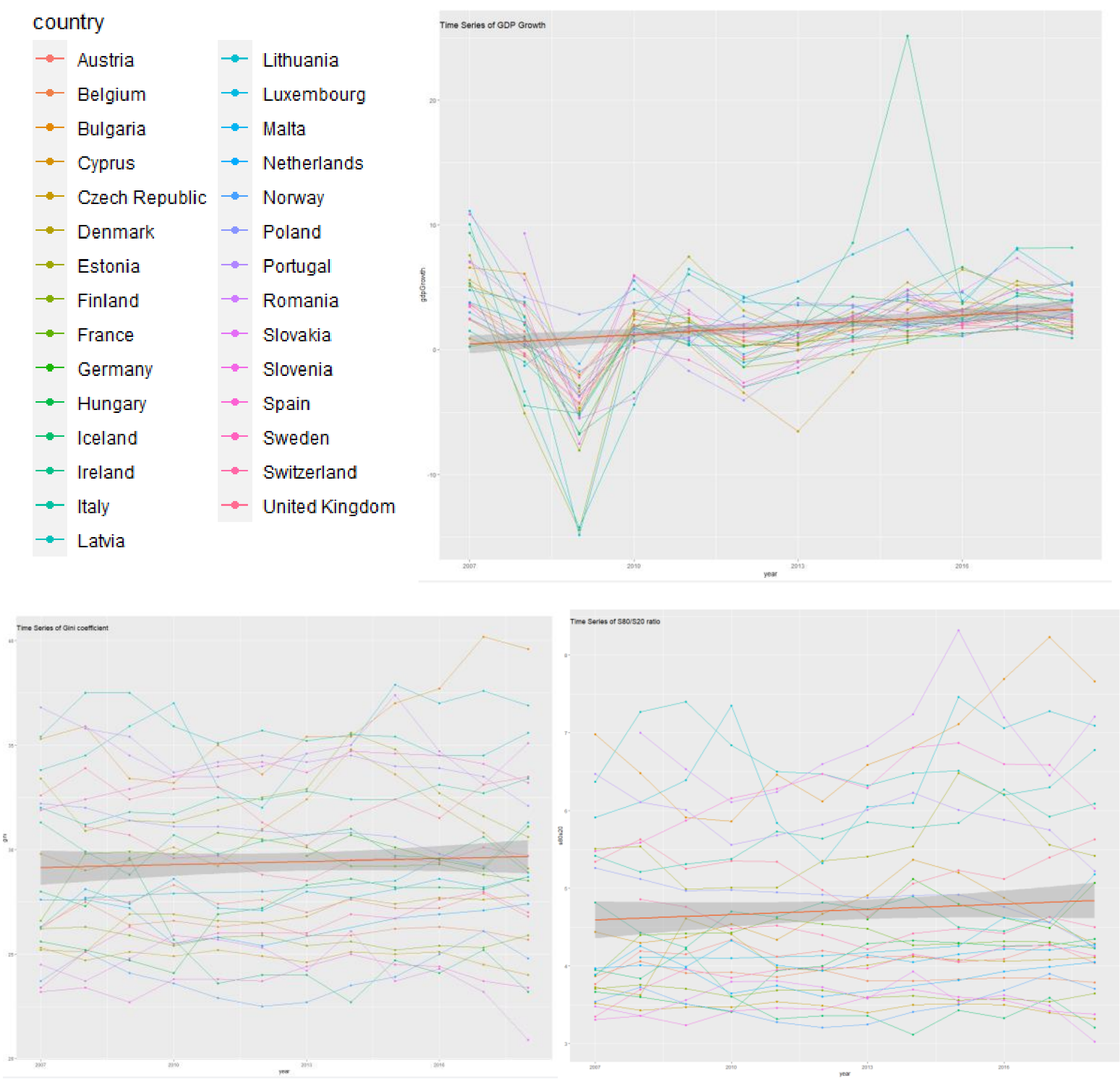
The following table demonstrates descriptive values of the dataset being analysed, highlighting minimum, quantile, median, mean, max and standard deviation values:

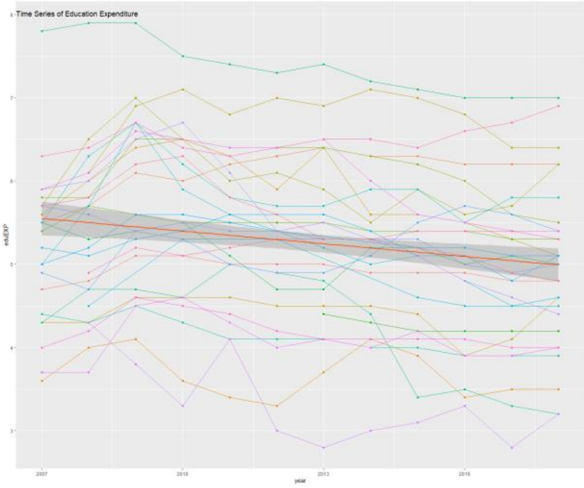
	Gini	Gini Change	s80s20	s80s20 Change	gdp	Gdp Growth	EduEXP	healthEXP
Min	20.9	0.7688	3.03	0.6697	7,925,000,000	-14.8386	2.8	1.6
1st Quantile	26.1	0.09783	3.835	0.9683	51,770,000,000	0.7046	4.55	5
Median	29.1	1	4.34	0.9985	245,000,000,000	2.0246	5.3	6.7
Mean	29.43	0.9991	4.731	1.0006	594,000,000,000	1.8868	5.261	6.167
3rd Quantile	32.65	1.0189	5.525	1.0312	538,000,000,000	3.6278	5.9	7.5
Max	40.2	1.1314	8.32	1.3633	3,960,000,000,000	25.1625	7.9	8.9
Standard Deviation	3.999965	0.04032	1.15835	0.066435	869,579,126,135	3.624832	0.99864	1.662516

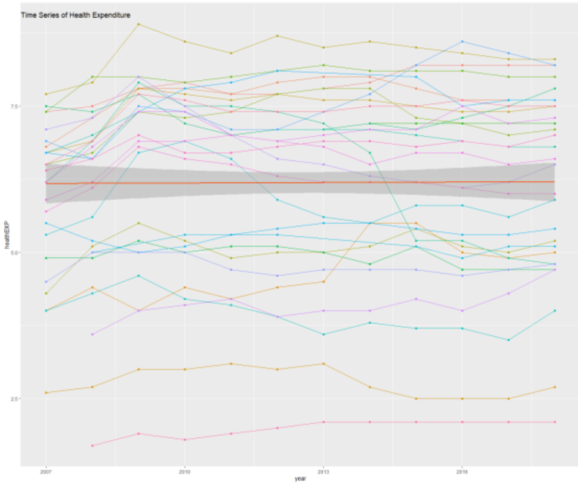
	Soprot EXP	govEXP	FDIinc	netTrade	unemployment	Edu ENROLL	Edu ATTAIN	eduATTAIN change
Min	7.9	17.1	-361000000000	-85470000000	2.24	10.61	13.3	-0.1139
1st Quantile	12.9	23.4	1285000000	-1123000000	5.28	58.36	30.9	0.00464
Median	16.1	27.3	7804000000	1742000000	7.09	67.67	41.3	0.03052
Mean	16.17	27.65	26210000000	13640000000	8.012	66.61	38.61	0.03549
3rd Quantile	19.3	31.6	27460000000	20890000000	9.635	75.88	46	0.5576
Max	25.5	40.5	734000000000	258000000000	26.09	94.92	58.7	0.47601
Standard Deviation	4.0708	5.550	69416457839	42632148713	3.954279	14.22968	10.29561	0.05318

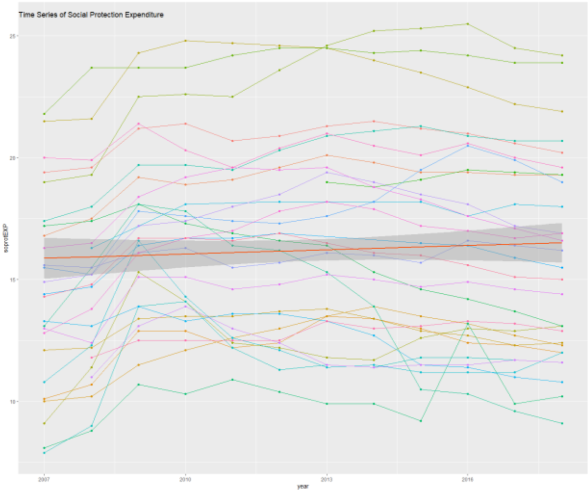
Table 4 - Descriptive Values

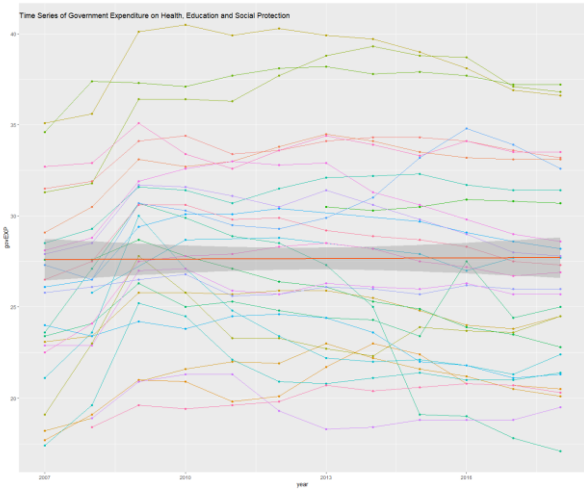
The following set of figures represent time series line graphs for each variable between 2007 and 2018, while the initial figure demonstrates the shared colour scheme for each plot. A trend line was calculated and plotted using linear regression in R to demonstrate variable movement across years.

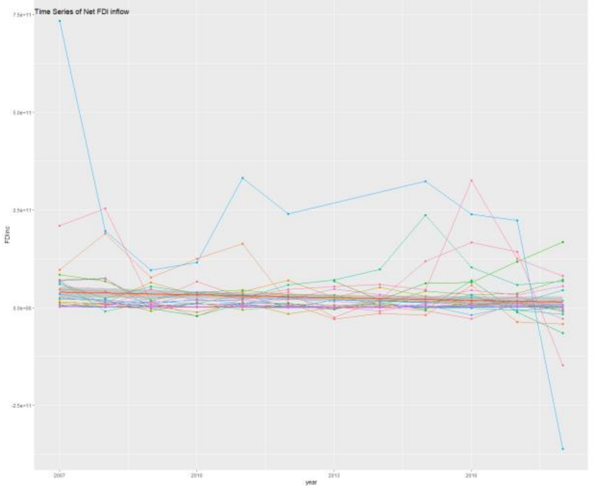

country
Austria
Belgium
Bulgaria
Cyprus
Czech Republic
Denmark
Estonia
Finland
France
Germany
Hungary
Iceland
Ireland
Italy
Latvia
Lithuania
Luxembourg
Malta
Netherlands
Norway
Poland
Portugal
Romania
Slovakia
Slovenia
Spain
Sweden
Switzerland
United Kingdom
Time Series of GDP Growth
20
10
0
-10
2007
2010
2013
2016
year
Time Series of Gini coefficient
2007
2010
2013
2016
year
Time Series of S80/S20 ratio
2007
2010
2013
2016
year

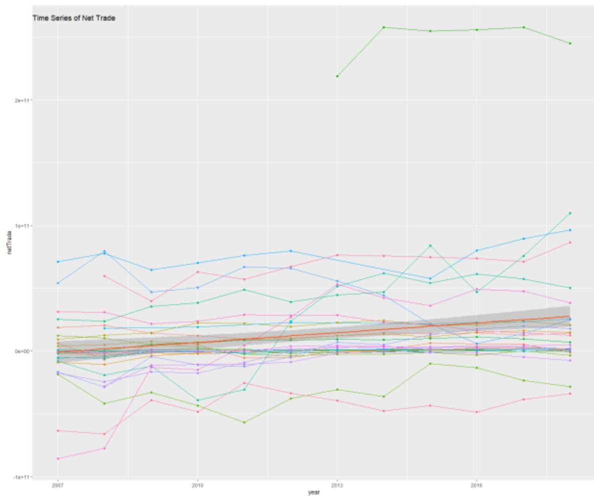

Time Series of Education Expenditure
eduEXP
2007
2010
2013
2016
year


Time Series of Health Expenditure
healthEXP
2007
2010
2013
2016
year


Time Series of Social Protection Expenditure
2007
2010
2013
2016
year


Time Series of Government Expenditure on Health, Education and Social Protection
govEXP
2007
2010
2013
2016
year


Time Series of Net FDI inflow
FDIinc
2007
2010
2013
2016
year


Time Series of Net Trade
netTrade
2007
2010
2013
2016
year

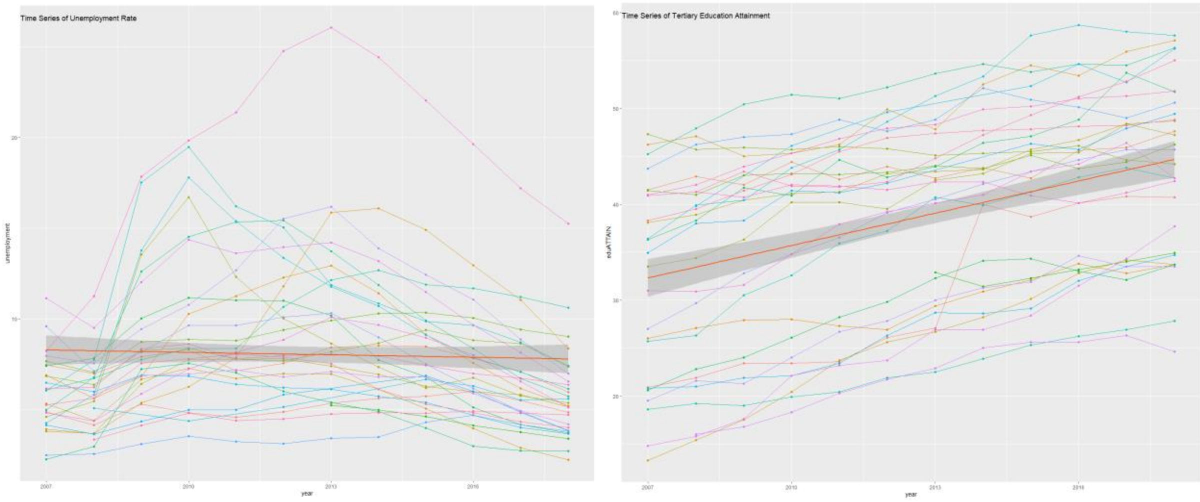


Figure 2 - Time Series Plots

GDP Growth appears to be slowly increasing between 1-3% annually. This is aligned with contemporary economics ideas of maintaining consistent growth. (Dreger and Reimers 2013) The plots for income inequality metrics; Gini coefficient and S80/S20 ratio suggest income inequality has marginally widened during this period. Bulgaria appears to have the highest levels of income inequality while Slovakia and Slovenia have the lowest. Education expenditure has decreased during this period, while health expenditure has been relatively stable. Social protection expenditure appears to be increasing, however, overall, government expenditure in these three areas has remained stable during the period. Nordic countries such as Denmark, Finland and Norway have the highest levels of government expenditure while, Ireland, Bulgaria and Romania appear to have lower levels of government expenditure. Net FDI inflow has gradually decreased. Notably, the Netherlands had the highest Net FDI inflow in 2007 and the lowest in 2018. Net trade appears to be increasing. Germany’s net trade is significantly the highest since 2013, although data is missing for prior years. France and the United Kingdom appear to have the lowest levels of net trade. The unemployment rate appears to have decreased in accordance with GDP growth, although there are outliers such as Spain who have significantly higher levels of unemployment compared to other EU countries. Lastly, tertiary education attainment has strongly increased since 2007 for all countries within this dataset. Lithuania, Cyprus and Ireland appear to be among the highest, while Romania and Italy are the lowest.

3.2 Correlation Matrix

The following table and figure represent the correlation matrix and demonstrate of the correlation heatmap calculated for each variable used in the dataset using R:

	gini	Gini Change	s80s20	s80s20 Change	gdp	Gdp Growth	eduEXP	healthEXP
gini	1.00000	0.16522	0.96498	0.16387	0.13484	-0.02315	-0.30828	-0.43770
giniChange	0.16522	1.00000	0.13670	0.89502	0.07456	-0.09962	0.00280	0.03198
s80s20	0.96498	0.13670	1.00000	0.17444	0.08208	-0.02185	-0.37305	-0.40305
s80s20Change	0.16387	0.89502	0.17444	1.00000	0.06539	-0.11243	-0.02175	0.03333
gdp	0.13484	0.07456	0.08208	0.06539	1.00000	-0.11413	-0.16053	0.31776
gdpGrowth	-0.02315	-0.09962	-0.02185	-0.11243	-0.11413	1.00000	-0.22744	-0.16723
eduEXP	-0.30828	0.00280	-0.37305	-0.02175	-0.16053	-0.22744	1.00000	0.21074
healthEXP	-0.43770	0.03198	-0.40305	0.03333	0.31776	-0.16723	0.21074	1.00000
soprotEXP	-0.31765	0.06474	-0.31448	0.04845	0.39591	-0.29714	0.21140	0.61487
govEXP	-0.41746	0.05735	-0.41654	0.04141	0.35490	-0.30788	0.39698	0.78383
FDIinc	-0.00710	0.05093	-0.05472	0.02118	0.29126	0.05483	-0.06987	0.07445
netTrade	-0.13779	0.02220	-0.12884	0.02134	0.22349	0.04103	-0.18811	-0.00423
unemployment	0.39301	0.03312	0.40366	0.04800	0.01434	-0.22380	-0.11693	-0.05787
eduENROLL	-0.10023	-0.05771	-0.04883	-0.06506	-0.03876	-0.06463	0.30734	0.34640
eduATTAIN	-0.08693	0.04998	-0.15033	0.04189	0.01742	0.08608	0.36607	0.00902
eduATTAINchange	0.01979	-0.04891	0.01850	-0.05978	-0.12895	-0.04554	-0.06704	-0.06174

	soprotEXP	govEXP	FDIinc	netTrade	unemployment	eduENROLL	eduATTAIN	EduATTAIN change
gini	-0.31765	-0.41746	-0.00710	-0.13779	0.39301	-0.10023	-0.08693	0.01979
giniChange	0.06474	0.05735	0.05093	0.02220	0.03312	-0.05771	0.04998	-0.04891
s80s20	-0.31448	-0.41654	-0.05472	-0.12884	0.40366	-0.04883	-0.15033	0.01850
s80s20Change	0.04845	0.04141	0.02118	0.02134	0.04800	-0.06506	0.04189	-0.05978
gdp	0.39591	0.35490	0.29126	0.22349	0.01434	-0.03876	0.01742	-0.12895
gdpGrowth	-0.29714	-0.30788	0.05483	0.04103	-0.22380	-0.06463	0.08608	-0.04554
eduEXP	0.21140	0.39698	-0.06987	-0.18811	-0.11693	0.30734	0.36607	-0.06704
healthEXP	0.61487	0.78383	0.07445	-0.00423	-0.05787	0.34640	0.00902	-0.06174
soprotEXP	1.00000	0.95185	0.00205	0.06653	0.08579	0.25042	0.10044	-0.14927
govEXP	0.95185	1.00000	0.01096	0.01360	0.02462	0.34107	0.14204	-0.13957
FDIinc	0.00205	0.01096	1.00000	0.17177	-0.06923	-0.05491	0.09234	-0.07721
netTrade	0.06653	0.01360	0.17177	1.00000	-0.14708	0.07782	0.10020	-0.06675
unemployment	0.08579	0.02462	-0.06923	-0.14708	1.00000	0.14562	-0.03707	0.05656
eduENROLL	0.25042	0.34107	-0.05491	0.07782	0.14562	1.00000	0.23636	-0.10062
eduATTAIN	0.10044	0.14204	0.09234	0.10020	-0.03707	0.23636	1.00000	-0.28981
eduATTAINchange	-0.14927	-0.13957	-0.07721	-0.06675	0.05656	-0.10062	-0.28981	1.00000

Table 5 - Correlation Matrix

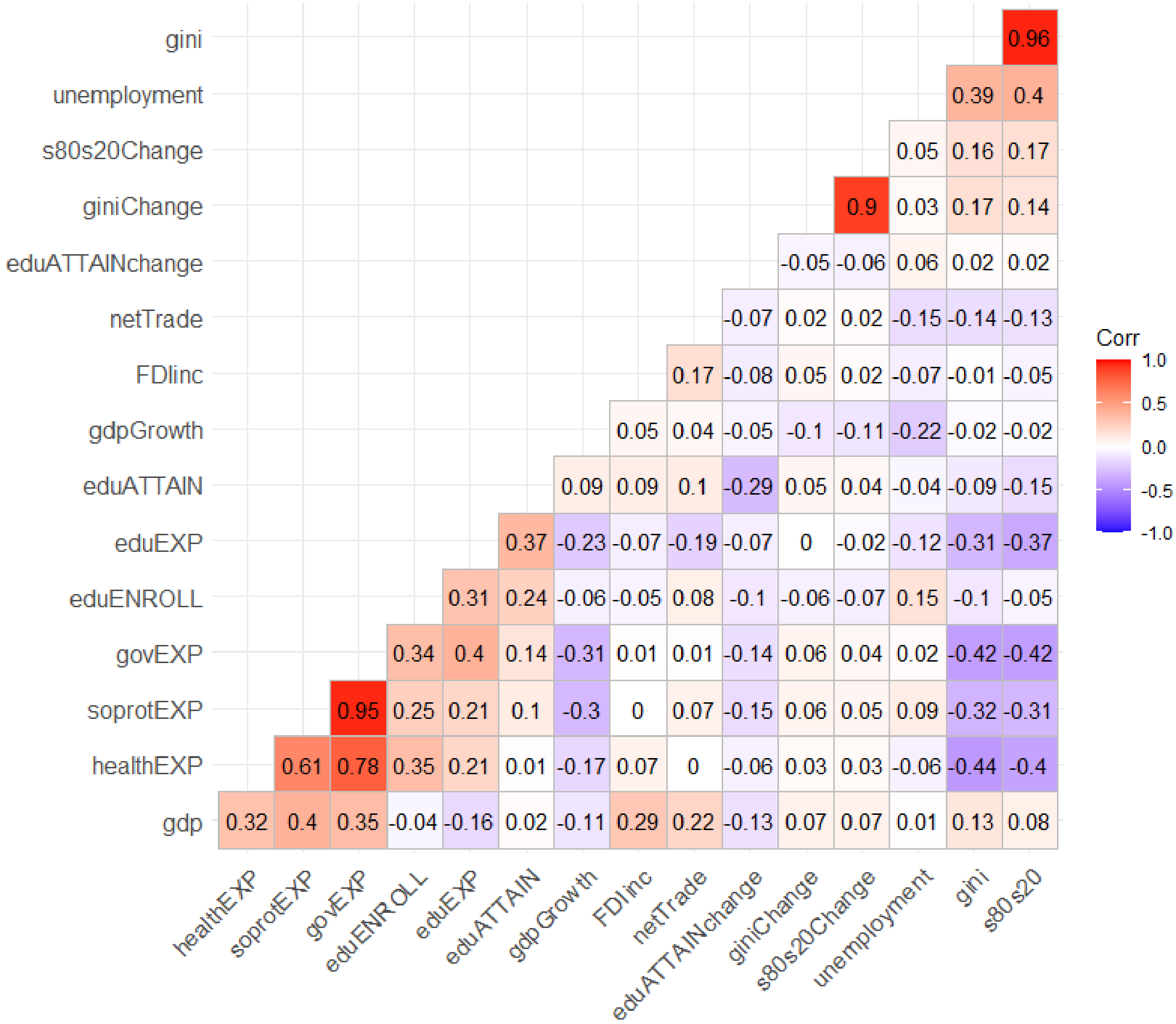


Figure 3 - Correlation Heatmap

The Gini coefficient and S80/S20 quintile ratios have a strong positive linear relationship at 0.96. This is as expected as both are different metrics of income inequality. The variance is due to Gini measuring the overall distribution while S80/S20 uses only the top 20% and bottom 20% quintiles. As discussed before, government expenditure is the sum on health expenditure, social protection expenditure and education expenditure as rate of GDP. Social protection has the highest correlation with government expenditure as it is a larger value compared to health and education expenditure. Healthcare is second and education is lowest of the three. Based on this correlation matrix, each of the government expenditure metrics are negatively linearly correlated with income inequality. Therefore, as each increase, income inequality is reduced. Unemployment has the strongest positive linear relationship with income inequality, with 0.39 and 0.4 correlations with Gini and S80/S20 respectively. Total GDP is positively linearly related with each of the government expenditure metrics as they are rates of GDP. Net FDI inflow and tertiary education attainment do not appear to have linear relationships with the Gini coefficient. However, education attainment has a slight negative linear relationship with the S80/S20 quintile ratio. This may be due to tertiary education attainment increasing income distribution towards the top 20%. Overall GDP appears to have a slight positive linear relationship with both income inequality metrics, while GDP growth does not appear to be linearly related. Overall, it appears there is no multicollinearity present when analysing the correlations between each income inequality metric and the remaining explanatory variables.

4. The Model

This paper aims to identify the determinants of income inequality using data compiled from 29 European countries with access to the EU single market, in the period 2007 – 2018. As the goal is to determine the relationship and coefficients between explanatory variables and dependent variables, multiple linear regression (MLR) was chosen as the ideal regression technique. Ordinary Least Squares (OLS) is used for regression. Alternatively, fixed-effects or random-effects panel linear models (PLM) as produced by Kiselakova *et al.* (2020) and Checchi and García-Peñalosa (2009) respectively. The following table represents the dependent and explanatory variables:

Dependent Variables	Explanatory Variables
Gini coefficient	GDP Growth
S80/S20 quintile ratio	Social Protection Exp
	Health Exp
	Education Exp
	Government Exp (Sum(Social Protection, Health, Education)
	Tertiary education attainment
	Unemployment
	Net FDI inflow
	Net trade

Table 6 - Dependent and Explanatory Variables

Both income inequality metrics, Gini coefficient and S80/S20 will be used in separate models. This is to ensure all potential explanatory variable relationships are accounted for, as well as identifying potential differences in significance levels depending on which income inequality metric is used and whether the effect is positive or negative. The explanatory variables include economic growth, government expenditure, trade openness foreign investment, unemployment and tertiary education. The data for these models are pooled because the aim is to detect the cross-country significant determinants of income inequality.

4.1 Initial Model

Based on literature review findings, the following models will be created:

Gimodel11:

$$\begin{aligned} gini_i = b_0 + b_1(gdpGrowth_i) + b_2(soprotEXP_i) + b_3(healthEXP_i) + b_4(eduEXP_i) \\ + b_5(eduATTAIN_i) + b_6(unemployment_i) + b_7(\log(FDIinc_i)) \\ + b_8(\ln(netTrade_i) + \ u_t \end{aligned}$$

S80s20model1:

$$\begin{aligned} s80s20_i = b_0 + b_1(gdpGrowth_i) + b_2(soprotEXP_i) + b_3(healthEXP_i) + b_4(eduEXP_i) \\ + b_5(eduATTAIN_i) + b_6(unemployment_i) + b_7(\log(FDIinc_i)) \\ + b_8(\ln(netTrade_i) + \ u_t \end{aligned}$$

Initially, the same dependent variables are used for modelling each income inequality metric. This is to enable identification of the varying impacts each explanatory variable has on the dependents. It is expected, GDP growth will reduce income inequality based on existing literature of similar studies. As the data analysed consists of EU developed countries, economic growth should contribute to reducing income inequality based on Kuznets (1955) inverted-U curve. It is expected each of the government expenditure explanatory variables will be significant in reducing income inequality. (Marrero and Rodríguez 2012) However, education expenditure in particular is expected to have a high significance on the S80/S20 quintile ratio due to tertiary education attainment expectedly rate of population in skilled employment. Unemployment is expected to have a significant positive effect on income inequality metrics, where each unit increase in unemployment will widen income inequality. This dataset is using net FDI inflow and net trade metrics, whereas existing research primarily focuses on total FDI inflow and total trade. For each metric, the total amount has been demonstrated to widen income inequality. (Asteriou *et al.* 2014) Therefore, it is interesting to use the net value for each metric to identify whether there are underlying causes which influence its effect on income inequality. It is important to note, as the net FDI inflow and net trade variables are in million units, a constant value is applied to each to remove negative values allowing the variable to be logged in the regression model.

4.2 Hypotheses

With the initial models established, the hypotheses can be defined before explanatory analysis of the initial models. As discussed, explanatory variables for this model are based on the following categories, economic growth, government expenditure, trade openness and foreign investment and education and unemployment. The explanatory variable coefficients' t-values are analysed to verify if each is statistically significant in determining income inequality. The following hypotheses are created:

Hypothesis 1: GDP Growth Significance

$$h_0: b_1 = 0 - gdpGrowth\ is\ statistically\ significant$$
$$h_1: b_1\ != 0 -\ gdpGrowth\ is\ not\ statistically\ significant$$

Hypothesis 2: Government Social Protection Expenditure Significance

$$h_0: b_2 = 0 -\ Social\ Protection\ Expenditure\ is\ statistically\ significant$$
$$h_1: b_2\ != 0 -\ Social\ Protection\ Expenditure\ is\ not\ statistically\ significant$$

Hypothesis 3: Government Healthcare Expenditure Significance

$$h_0: b_3 = 0 -\ \text{Healthcare}\ Expenditure\ is\ statistically\ significant$$
$$h_1: b_3\ != 0 -\ \text{Healthcare}\ Expenditure\ is\ not\ statistically\ significant$$

Hypothesis 4: Government Education Expenditure Significance

$$h_0: b_4 = 0 -\ Education\ Expenditure\ is\ statistically\ significant$$
$$h_1: b_4\ != 0 -\ Education\ Expenditure\ is\ not\ statistically\ significant$$

Hypothesis 5: Tertiary Education Attainment Significance

$$h_0: b_5 = 0 -\ Tertiary\ Education\ Attainment\ is\ statistically\ significant$$
$$h_1: b_5\ != 0 -\ Tertiary\ Education\ Attainment\ is\ not\ statistically\ significant$$

Hypothesis 6: Unemployment Significance

$$h_0: b_6 = 0 -\ Unemployment\ is\ statistically\ significant$$
$$h_1: b_6\ != 0 -\ Unemployment\ is\ not\ statistically\ significant$$

Hypothesis 7: Net FDI Inflow Significance

$$h_0: b_7 = 0 -\ Net\ FDI\ inflow\ is\ statistically\ significant$$
$$h_1: b_7\ != 0 -\ Net\ FDI\ inflow\ is\ not\ statistically\ significant$$

Hypothesis 8: Net Trade Significance

$$h_0: b_8 = 0 -\ Net\ Trade\ is\ statistically\ significant$$
$$h_1: b_8\ != 0 -\ Net\ Trade\ is\ not\ statistically\ significant$$

4.3 Initial Findings

The models above were created in R producing the following equations, the attached figures refer to the summary output from each regression:

Gimodel1 equation:

$$gini_i = 80.4324 - 0.10627(gdpGrowth_i) - 0.11823(soprotEXP_i) - 0.77828(healthEXP_i) - 0.99203(eduEXP_i) + 0.022364(eduATTAIN_i) + 0.31624(unemployment_i) + 0.86256(\log(FDIinc_i)) - 2.58246(\ln(netTrade_i) + u_t$$

```
Call:
lm(formula = gini ~ gdpGrowth + soprotEXP + healthEXP + eduEXP +
    eduATTAIN + unemployment + log(FDIinc) + log(netTrade), data = Portfoliodata)

Residuals:
   Min     1Q Median     3Q    Max
-8.547 -1.828 -0.271  2.087  8.459

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   80.43240   39.16699   2.054   0.0408 *
gdpGrowth     -0.10627    0.05220  -2.036   0.0426 *
soprotEXP     -0.11823    0.05558  -2.127   0.0342 *
healthEXP     -0.77828    0.13364  -5.824 1.39e-08 ***
eduEXP        -0.99203    0.20054  -4.947 1.21e-06 ***
eduATTAIN      0.02364    0.01848   1.279   0.2017
unemployment   0.31624    0.04535   6.973 1.75e-11 ***
log(FDIinc)    0.86256    1.35737   0.635   0.5256
log(netTrade) -2.58246    0.53217  -4.853 1.90e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.072 on 324 degrees of freedom
Multiple R-squared:  0.4193,    Adjusted R-squared:  0.405
F-statistic: 29.25 on 8 and 324 DF,  p-value: < 2.2e-16
```

Figure 4 - Gimodel1 Summary

Under the coefficients section, the explanatory variables health expenditure, education expenditure, unemployment and ln(net trade) are significant at the 100% level. GDP growth and social protection expenditure are significant at the 99% level. Tertiary education attainment and ln(net FDI inflow) are found not to be significant. A one unit increase in each explanatory variable will increase the Gini value by the coefficient. For example, each one unit increase in health expenditure per rate of GDP will decrease the Gini coefficient value by 0.77828. Whereas each unit increase in unemployment will increase the Gini coefficient value by 0.31624. The trade openness and foreign investment values were logged due to their values measured in million units creating greater variance the model must accommodate. Therefore, for each, the coefficients are divided by 100 as this is a linear-log model. For example, a 1% increase in Net Trade, the Gini coefficient is reduced by 0.0258246. For net FDI inflow, a 1% increase will increase the Gini coefficient by 0.0086256.

R-squared, R^2 or the coefficient of determination determines the proportion of variance in the dependent variable that can be explained by the explanatory variables. (Frost 2017) It is a measure of the goodness of fit of a regression model. It is a formal method of analysing how well the predicted line matches with the data. This model has an R-squared of 0.0.4193, this is good for an initial model aiming to identify relationships rather than predict outputs. An R-squared over 0.7 generally suggests heteroscedasticity is present. The residual standard error is 30.72 on 324 degrees of freedom. This implies model residuals vary by 30.72 between the real values and predicted values.

The F-statistic tests whether the explanatory variables in the model are significant. This model has an F-statistic of 29.25 on 8 explanatory variables and 324 degrees of freedom. The p value 2.2e-16 is less than the level of statistical significance, 0.05 meaning the 8 explanatory variables in the model are a better fit to the dataset than an intercept only model.

S80s20model2 equation:

$$s80s20_i = 5.57745 - 0.00582(gdpGrowth_i) - 0.00563(soprotEXP_i) - 0.04161(healthEXP_i) - 0.06868(eduEXP_i) + 0.00069(eduATTAIN_i) + 0.01899(unemployment_i) - 0.00623(\ln(FDIinc_i)) - 0.13159(\ln(netTrade_i) + u_t$$

```
Call:
lm(formula = log(s80s20) ~ gdpGrowth + soprotEXP + healthEXP +
    eduEXP + eduATTAIN + unemployment + log(FDIinc) + log(netTrade),
    data = Portfoliodata)

Residuals:
    Min      1Q  Median      3Q     Max
-0.4356 -0.1102 -0.0131  0.1297  0.4290

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.5774519  2.2581296   2.470   0.0140 *
gdpGrowth     -0.0058215  0.0030093  -1.935   0.0539 .
soprotEXP     -0.0056307  0.0032044  -1.757   0.0798 .
healthEXP     -0.0416107  0.0077049  -5.401 1.29e-07 ***
eduEXP        -0.0686860  0.0115621  -5.941 7.32e-09 ***
eduATTAIN      0.0006908  0.0010654   0.648   0.5172
unemployment   0.0189963  0.0026149   7.265 2.81e-12 ***
log(FDIinc)   -0.0062254  0.0782579  -0.080   0.9366
log(netTrade) -0.1315987  0.0306816  -4.289 2.37e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1771 on 324 degrees of freedom
Multiple R-squared:  0.4243,    Adjusted R-squared:  0.4101
F-statistic: 29.85 on 8 and 324 DF,  p-value: < 2.2e-16
```

Figure 5 - S80s20model1 Summary

Under the coefficients section, the explanatory variables health expenditure, education expenditure, unemployment and ln(net trade) are significant at the 100% level. GDP growth and social protection expenditure are significant at the 90% level. Tertiary education attainment and ln(net FDI inflow) are found not to be significant. Each one unit increase in health expenditure per rate of GDP will decrease the S80/S20 quintile ratio by 0.0416107. Whereas each unit increase in unemployment will increase the S80/S20 quintile ratio by 0.0189963. The trade openness and foreign investment values were logged due to their values measured in million units creating greater variance the model must accommodate. Therefore, for each, the coefficients are divided by 100 as this is a linear-log model. For example, a 1% increase in Net Trade, the S80/S20 quintile ratio is reduced by 0.001315987. While for net FDI inflow, there is essentially zero impact on the S80/S20 quintile ratio.

This model has an R-squared of 0.0.4243, this is similar to the Gimodel2, which is useful for an initial model. Heteroscedasticity does not appear to be present in this model either. The residual standard error is 0.1771 on 324 degrees of freedom. This implies model residuals vary by 0.1771 between the real values and predicted values.

This model has an F-statistic of 29.85 on 8 explanatory variables and 324 degrees of freedom. The p value 2.2e-16 is less than the level of statistical significance, 0.05 meaning the 8 explanatory variables in the model are a better fit to the dataset than an intercept only model. The previous model had an F-statistic of 29.25 therefore these results are very similar.

Both of these models demonstrate consistency in which explanatory variables are found to be significant, and at the level of significance. Health expenditure, education expenditure, unemployment and net trade were each significant at the 100% level for each model. GDP growth and social protection were significant at the 99% level in the Gini model and significant at the 90% level on the S80S20 model. For both models, neither tertiary education attainment or net FDI inflow were statistically significant. The explanatory variables aside from those found unsignificant are consistent with literature review findings. Therefore, the model is deemed acceptable and can progress to the next section for regression diagnostics and empirical results analysis.

5. Empirical results

5.1 Regression Diagnostics

MLR is a very effective statistical method to identify relationships between dependent and explanatory variables, however, for it to be applied, the model must satisfy the following assumptions: a linear relationship between independent and explanatory variables. (Altman and Krzywinski 2016) Multicollinearity is not present, where the explanatory variables are not strongly correlated. Residual independence or autocorrelation, where there is no correlation between consecutive residuals. Homoscedasticity, where the residuals at each level of x maintain a constant variance, otherwise the model is heteroscedasticity. Lastly, normality, ensuring the model's residuals are normally distributed. (Altman and Krzywinski 2016) The following section with apply regression diagnostic tests to the model to identify the presence of multicollinearity, heteroscedasticity and autocorrelation, if present, they will be resolved.

5.1.1 Multicollinearity

Multicollinearity occurs when multiple explanatory variables are highly correlated. (Zach 2019) It is generally accepted a correlation below 0.7 is acceptable, while a correlation of 0.9 is the threshold beyond which problems will arise. For example, if education attainment and education expenditure had a strong positive linear relationship of 0.8. The MLR may assign a high value coefficient to the education attainment variable and low value coefficient to the education expenditure variable. This would result in difficulty understanding the effect each explanatory variable has on the dependent variable. For example, residuals within the education expenditure variable may influence the education attainment coefficient. As discovered from the correlation matrix and initial model's R-squared values, multicollinearity does not appear to be present. This can be further tested by applying the variance inflation factor (VIF) function within R to each model producing the following outputs:

```
> car::vif(Gimodel1)
  gdpGrowth    soprotEXP    healthEXP       eduEXP    eduATTAIN unemployment  log(FDIinc) log(netTrade)
   1.257581     1.793922     1.695395     1.411141     1.273819     1.134020     1.022923      1.081906
> car::vif(s80s20model1)
  gdpGrowth    soprotEXP    healthEXP       eduEXP    eduATTAIN unemployment  log(FDIinc) log(netTrade)
   1.257581     1.793922     1.695395     1.411141     1.273819     1.134020     1.022923      1.081906
```

Figure 6 - VIF Output for Gimodel1 and s80s20model1 respectively

The VIF result for each variable in both models ranges from 1.022923 and 1.793922. A VIF value of 1 typically suggests there is no multicollinearity present. Generally, a threshold of 5 is where multicollinearity begins to cause issues. (Craney and Surles 2002) These results are acceptable, confirming will not negatively influence either model's results.

5.1.2 Heteroscedasticity

Heteroscedasticity occurs when error variances in a linear regression model are non-constant. (Yobero 2016) The assumption of classical linear regression models is that errors appearing in the population regression function are homoscedastic, that they have the same variance. The presence of heteroscedasticity can be tested using the Bresuche-Pagan LM test. The bptest function in R produces the following output for each model:

```
> bptest(Gimodel1)

        studentized Breusch-Pagan test

data:  Gimodel1
BP = 56.494, df = 8, p-value = 2.26e-09
```

```
> bptest(s80s20model1)

        studentized Breusch-Pagan test

data:  s80s20model1
BP = 67.618, df = 8, p-value = 1.462e-11
```

Figure 7 - bpTest Output for Gimodel1 and S80s20model1 respectively

To formally test if heteroscedasticity is present in the model, an If statement is created in R comparing the BP test statistic to the critical value. If the BP score is greater than the critical value (qchisq(.95, df=3)), then reject the null hypothesis and confirm heteroscedasticity is present. First state the hypotheses:

$$h_0: BP > CV - \text{ Heteroscedasticity } \textit{is present}$$

$$h_1: BP \leq CV - \text{Heteroscedasticity } \textit{is not present}$$

Executing the If statement in R produces the following outputs:

```
> if(bptest(Gimodel1)$statistic > qchisq(.95, df=8)){
+   print("Reject the null hypothesis i.e. heteroscedascity is present")
+ } else {
+   print("Fail to reject the null hypothesis i.e. no heteroscedascity detected")
+ }
[1] "Reject the null hypothesis i.e. heteroscedascity is present"
```

```
> if(bptest(s80s20model1)$statistic > qchisq(.95, df=8)){
+   print("Reject the null hypothesis i.e. heteroscedascity is present")
+ } else {
+   print("Fail to reject the null hypothesis i.e. no heteroscedascity detected")
+ }
[1] "Reject the null hypothesis i.e. heteroscedascity is present"
```

Figure 8 - Formally testing presence of heteroscedasticity in Gimodel1 and S80s20model1 respectively

As expected from the informal Bptest p-value, for each model, the BP score is greater than the critical value meaning the null hypothesis can be rejected confirming heteroscedasticity is present. This voids the assumptions of MLR where the spread of residuals should be constant. Therefore, the results from these models are less precise, increasing the likelihood that coefficient estimates are further from the correct value. As the aim of these models is to identify the significant determinants of income inequality, heteroscedasticity must be resolved, otherwise it may lead to conclusions of variables appearing statistically significant when they are not. (Altman and Krzywinski 2016)

As noticed in the initial results of Gimodel1, each of the government expenditure variables were found to be significant. It is suspected these variables are a potential cause of heteroscedasticity within the model. Additionally, the dependent variable Gini coefficient will be logged. As tertiary education attainment was found not to be statistically significant, it is removed from the model. The following model is created:

Gimodel2:

$$\ln(gini_i) = b_0 + b_1(gdpGrowth_i) + b_2(govEXP_i) + b_3(unemployment_i) + b_4(\ln(FDIinc_i)) + b_5(\ln(netTrade)) + u_t$$

```
Call:
lm(formula = log(gini) ~ gdpGrowth + govEXP + unemployment +
    log(FDIinc) + log(netTrade), data = Portfoliodata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.33179 -0.05420  0.00324  0.07970  0.27437

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.917039   1.360269   2.880 0.004244 **
gdpGrowth     -0.002557   0.001785  -1.432 0.153010
govEXP        -0.010646   0.001135  -9.380  < 2e-16 ***
unemployment   0.012600   0.001561   8.070 1.35e-14 ***
log(FDIinc)    0.048063   0.047724   1.007 0.314634
log(netTrade) -0.064711   0.018274  -3.541 0.000456 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.109 on 327 degrees of freedom
Multiple R-squared:  0.3559,	Adjusted R-squared:  0.346
F-statistic: 36.13 on 5 and 327 DF,  p-value: < 2.2e-16
```

Figure 9 - Gimodel2 Summary

As expected, the government expenditure variable remains statistically significant at the 100% level, as well as unemployment and net trade. This model was then tested for the presence of heteroscedasticity:

```
> if(bptest(Gimodel2)$statistic > qchisq(.95, df=5)){
+   print("Reject the null hypothesis i.e. heteroscedascity is present")
+ } else {
+   print("Fail to reject the null hypothesis i.e. no heteroscedascity detected")
+ }
[1] "Fail to reject the null hypothesis i.e. no heteroscedascity detected"
```

Figure 10 - Gimodel2 test for presence of heteroscedasticity

The BP score is less than the critical value therefore, fail to reject the null hypothesis concluding the updated model is homoscedastic. Next, heteroscedasticity is resolved in the s80s20 model taking similar steps as shown in Gimodel2 by creating the following model:

s80s20model2:

$$\ln(s80s20_i) = b_0 + b_1(gdpGrowth_i) + b_2(eduATTAINchange_i) + b_3(unemployment_i) + b_4(\log(FDIinc_i)) + b_5(\ln(netTrade_i) + u_t$$

```
Call:
lm(formula = log(s80s20) ~ gdpGrowth + eduATTAINchange + unemployment +
    log(FDIinc) + log(netTrade), data = Portfoliodata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.40286 -0.14408 -0.02427  0.14066  0.61179

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.018716   2.641612   1.521  0.12915
gdpGrowth        0.004737   0.003254   1.456  0.14645
eduATTAINchange -0.057739   0.216844  -0.266  0.79020
unemployment     0.023996   0.002993   8.019 1.92e-14 ***
log(FDIinc)     -0.009115   0.091664  -0.099  0.92085
log(netTrade)   -0.094087   0.035058  -2.684  0.00765 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2091 on 327 degrees of freedom
Multiple R-squared:  0.1902,    Adjusted R-squared:  0.1778
F-statistic: 15.36 on 5 and 327 DF,  p-value: 1.429e-13
```

Figure 11 - S80s20model2 Summary

As expected, unemployment remains statistically significant at the 100% level, while net trade is significant at the 99.9% level. A model using S80s20 as the dependent variable with government expenditure as an explanatory variable continuously produced evidence of heteroscedasticity. As government expenditure has been demonstrated to be significant in each model before, it was removed to create a homoscedastic s80s20 model. Additionally, tertiary education attainment was changed to the rate of change annually on tertiary education attainment. This model was tested for the presence of heteroscedasticity:

```
> if(bptest(s80s20model2)$statistic > qchisq(.95, df=5)){
+   print("Reject the null hypothesis i.e. heteroscedascity is present")
+ } else {
+   print("Fail to reject the null hypothesis i.e. no heteroscedascity detected")
+ }
[1] "Fail to reject the null hypothesis i.e. no heteroscedascity detected"
```

Figure 12 - S80s20model2 test for presence of heteroscedasticity

Heteroscedasticity has now been resolved for each model. Each have been formally tested and verified as homoscedastic, aligning with MLR assumptions. Therefore, regression diagnostic testing can progress to the final stage, testing for autocorrelation.

5.1.3 Autocorrelation

Autocorrelation is the representation of the degree in similarity between a given time series and lagged version over consecutive time intervals. (Altman and Krzywinski 2016) It is especially important to be identified and resolved as these models use a panel dataset spanning multiple years. If autocorrelation is present, the standard errors of coefficients in the model are likely to be underestimated resulting in explanatory variables incorrectly deemed statistically significant. Using the newly created homoscedastic models, the Durbin-Watson test is performed. An If statement was created to determine whether the autocorrelation is positive, negative, within a zone of indecision, else autocorrelation is not present. The If statement uses the DW test statistic from the model, comparing this with dL and Du variables in the Durbin-Watson Significance Tables. The If statements were executed for each model producing the following outputs:

```
> if(testStat > 0 && testStat < dl){
+   print("Reject the null hypothesis i.e. positive autocorrelation")
+ } else if(testStat > dl && testStat < du) {
+   print("Zone of indecision")
+ } else if(testStat > 4-dl && testStat < 4) {
+   print("Reject the null hypothesis i.e. negative autocorrelation")
+ } else if(testStat > 4-du && testStat < 4-dl) {
+   print("Zone of indecision")
+ } else{
+   print("Do not reject null hypothesis")
+ }
[1] "Reject the null hypothesis i.e. positive autocorrelation"
```

```
> if(testStat > 0 && testStat < dl){
+   print("Reject the null hypothesis i.e. positive autocorrelation")
+ } else if(testStat > dl && testStat < du) {
+   print("Zone of indecision")
+ } else if(testStat > 4-dl && testStat < 4) {
+   print("Reject the null hypothesis i.e. negative autocorrelation")
+ } else if(testStat > 4-du && testStat < 4-dl) {
+   print("Zone of indecision")
+ } else{
+   print("Do not reject null hypothesis")
+ }
[1] "Reject the null hypothesis i.e. positive autocorrelation"
```

Figure 13 - Formal test of autocorrelation for Gimodel2 and S80s20model2 respectively

Positive autocorrelation was found in both models. It must be resolved otherwise residual variances may be underestimated, the R-squared is overestimated or the coefficient variances are underestimated. (Altman and Krzywinski 2016) The Cochrane-Orcutt estimation can be applied to each model, adjusting for serial correlation within the error term. The Cochrane-Orcutt function was applied to each model within R creating the following models:

cochGimodel2:

$$\ln(gini_i) = b_0 + b_1(gdpGrowth_i) + b_2(govEXP_i) + b_3(unemployment_i) + b_4(\ln(FDIinc_i)) + b_5(\ln(netTrade)) + u_t$$

```
Call:
lm(formula = log(gini) ~ gdpGrowth + govEXP + unemployment +
    log(FDIinc) + log(netTrade), data = Portfoliodata)

                 Estimate  Std. Error t value  Pr(>|t|)
(Intercept)     5.11117043  0.76314881   6.697 9.301e-11 ***
gdpGrowth      -0.00191613  0.00087901  -2.180 0.0299819 *
govEXP         -0.01060407  0.00141469  -7.496 6.276e-13 ***
unemployment    0.00918150  0.00187081   4.908 1.458e-06 ***
log(FDIinc)    -0.00886322  0.02251742  -0.394 0.6941217
log(netTrade)  -0.05014497  0.01509661  -3.322 0.0009966 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0583 on 326 degrees of freedom
Multiple R-squared:  0.201 ,  Adjusted R-squared:  0.1887
F-statistic: 16.4 on 5 and 326 DF,  p-value: < 1.943e-14

Durbin-Watson statistic
(original):    0.32077 , p-value: 1.086e-54
(transformed): 2.15165 , p-value: 9.177e-01
```

Figure 14 - cochGimodel2 Summary

cochs80s20model2:

$$\ln(s80s20_i) = b_0 + b_1(gdpGrowth_i) + b_2(eduATTAINchange_i) + b_3(unemployment_i) + b_4(\log(FDIinc_i)) + b_5(\ln(netTrade_i) + u_t$$

```
Call:
lm(formula = log(s80s20) ~ gdpGrowth + eduATTAINchange + unemployment +
    log(FDIinc) + log(netTrade), data = Portfoliodata)

                  Estimate   Std. Error t value Pr(>|t|)
(Intercept)      3.95888745  1.42151704   2.785 0.005666 **
gdpGrowth        0.00049876  0.00151219   0.330 0.741744
eduATTAINchange  0.01221924  0.09008451   0.136 0.892188
unemployment     0.00985762  0.00343510   2.870 0.004377 **
log(FDIinc)     -0.01254064  0.04151049  -0.302 0.762763
log(netTrade)   -0.08586797  0.02804684  -3.062 0.002385 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1082 on 326 degrees of freedom
Multiple R-squared:  0.053 ,  Adjusted R-squared:  0.0385
F-statistic: 3.6 on 5 and 326 DF,  p-value: < 3.167e-03

Durbin-Watson statistic
(original):    0.30709 , p-value: 2.65e-55
(transformed): 2.08237 , p-value: 7.869e-01
```

Figure 15- cochS80s20model2 Summary

5.2 Interpreting Results

After completing the regression diagnostic testing stage, the results for each model created through this paper are demonstrated in the following tables:

Gimodel1 –

$$gini_i = b_0 + b_1(gdpGrowth_i) + b_2(soprotEXP_i) + b_3(healthEXP_i) + b_4(eduEXP_i) + b_5(eduATTAIN_i) + b_6(unemployment_i) + b_7(\log(FDIinc_i)) + b_8(\ln(netTrade_i) + u_t$$

Gimodel2 –

$$\ln(gini_i) = b_0 + b_1(gdpGrowth_i) + b_2(govEXP_i) + b_3(unemployment_i) + b_4(\ln(FDIinc_i)) + b_5(\ln(netTrade)) + u_t$$

cochGimodel3 –

Cochrane Orcutt Transformation: $\ln(gini_i) = b_0 + b_1(gdpGrowth_i) + b_2(govEXP_i) + b_3(unemployment_i) + b_4(\ln(FDIinc_i)) + b_5(\ln(netTrade)) + u_t$

	Gimodel1	Gimodel2	cochGimodel2	
Intercept	80.4224*	3.917039**	5.11117***	
gdpGowth	-0.10627*	-0.002557	-0.001916*	
soprotEXP	-0.11823*			
healthEXP	-0.77828***			
eduEXP	-0.99203***			
govEXP		-0.010646***	-0.01060***	
EduATTAIN	0.02364			
eduATTAINchange				
Unemployment	0.31624***	0.012600***	0.00918***	
log(FDIinc)	0.86256	0.048063	-0.00886	
Log(netTrade)	-2.58246***	-0.064711***	-0.05014***	
RSE	3.072 on 324 df	0.109 on 327 df	0.0583 on 326 df	
R-squared (adjust R-squared)	0.4193 (0.405)	0.3559 (0.346)	0.201 (0.1887)	
F-stat	29.25 on 8 and 34DF, p-value < 2.2e-16	36.13 on 5 and 327 DF, p-value < 2.2e-16	16.4 on 5 and 326 DF, p-value: < 1.943e-14	
DW-Stat		0.32077, p-value 1.086e-54	2.15165, p-value 9.177e-01	
Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 7 – Gini Empirical Results

Gimodel1 included eight explanatory variables, however this model was found to be heteroscedastic. Therefore, the three government expenditure variables were added to create a single government expenditure value. Additionally, in Gimodel2 and cochGimodel2, the log value of the Gini coefficient is used. Therefore, when interpreting the coefficients, the exp() function is applied to each with R to identify the effect each explanatory variable unit increase has on the Gini coefficient. For example, government expenditure in cochGimode2 has a negative coefficient of -0.01060 at the 100% significance level on the Gini coefficient. Therefore, for each unit increase in government expenditure in healthcare, social protection and education as rate of GDP, Gini coefficient is reduced by exp(-0.001916) or 0.989456. Additionally, net trade maintained a significance at the 100% level across all three models. In cochmodel3, logging is used for both net trade and the Gini coefficient. Therefore, each 1% increase in net trade reduces the Gini coefficient by 0.0583%. GDP Growth was found to be significant at the 99% level in Gimodel1 and cochGimodel2 in reducing the Gini coefficient. In cochmodel2, each unit increase in GDP growth reduced the Gini coefficient by exp(-0.001916) or 0.9980858. The R-squared value fell from 0.4193 in Gimodel1 to 0.201 in cochGimodel2. While this in itself it not preferable, it was necessary to ensure heteroskedasticity and autocorrelation were resolved, otherwise the results would be less precise. Although as mentioned, the purpose is to identify the significance of relationships rather than predict therefore the R-squared is less important.

Referring to the hypotheses established, it can be concluded GDP growth, Net FDI inflow and education attainment are not statistically significant determinants of income inequality. In Gimodel1, social protection expenditure was not found to be a statistically significant determinant of income inequality, while healthcare expenditure and education expenditure are statistically significant determinants of income inequality. For each model, unemployment rates and net trade were statistically significant.

S80s20model1 –

$$s80s20_i = b_0 + b_1(gdpGrowth_i) + b_2(soprotEXP_i) + b_3(healthEXP_i) + b_4(eduEXP_i) + b_5(eduATTAIN_i) + b_6(unemployment_i) + b_7(\log(FDIinc_i)) + b_8(\ln(netTrade_i) + u_t$$

S80s20model2 –

$$s80s20_i = b_0 + b_1(gdpGrowth_i) + b_2(eduATTAINchange_i) + b_3(unemployment_i) + b_4(\log(FDIinc_i)) + b_5(\ln(netTrade_i) + u_t$$

cochS80s20model3 –

Cochrane Orcutt Transformation: $s80s20_i = b_0 + b_1(gdpGrowth_i) + b_2(eduATTAINchange_i) + b_3(unemployment_i) + b_4(\log(FDIinc_i)) + b_5(\ln(netTrade_i) + u_t$

	S80s20model1	S80s20model2	S80s20model3
Intercept	5.5774519*	4.01716	0.005666**
gdpGowth	-0.00582.	0.004737	0.000498
soprotEXP	-0.00563.		
healthEXP	-0.04161***		
eduEXP	-0.06868***		
govEXP			
EduATTAIN	0.00069		
eduATTAINchange		-0.057739	0.01221924
Unemployment	0.01899***	0.023996***	0.0098576**
log(FDIinc)	-0.00622	-0.009115	-0.01254064
Log(netTrade)	-0.13159***	-0.094087**	-0.08586797**
RSE	0.1771 on 324 df	0.2091 on 327 DF	0.1082 on 326DF
R-squared (adjust R-squared)	0.4243(0.4101)	0.1902 (0.1778)	0.053 (0.0385)
F-stat	29.85 on 8 and 324 DF, p-value: < 2.2e-16	15.36 on 5 and 327 DF, p-value: < 1.429e-13	3.6 on 5 and 326 DF, p-value: < 3.167e-03
DW-Stat		0.3070873, p-value: 2.65e-55	2.08237, p-value: 7.869e-01
Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Table 8 - S80s20 Empirical Results

S80s20model1 included eight explanatory variables, however this model was found to be heteroscedastic. Therefore, the government expenditure variables were removed as it was demonstrated as statistically significant across all Gini models and S80s20model1. Additionally, in S80s20model2 and cochS80s20model2, the log value of the s80s20 coefficient is used. Therefore, when interpreting the coefficients, the exp() function is applied to each with R to identify the effect each explanatory variable unit increase has on the S80s20 decile ratio. For example, unemployment in cochGimode2 has a positive coefficient of 0.0098576 at the 99.9% significance level on the S80s20 decile ratio. Therefore, for each unit increase in unemployment, S80s20 decile ratio is increased by exp(0.0098576) or 1.009906. Net Trade maintained a significance of at least 99.9% across all three s80s20 models. For each 1% increase in net trade, the s80s20 quintile ratio fell by 0.08586%. No other variables were statistically significant within this model. The R-squared value fell from 0.4243 in S80s20model1 to 0.053 in cochS80s20model2. This suggests there are a number of omitted variables correlated with the S80s20 decile ratio. This is understood to be a result of removing the government expenditure variables

Referring to the hypotheses established, it can be concluded GDP growth, Net FDI inflow and education attainment are not statistically significant determinants of income inequality. In Gimodel1, social protection expenditure was not found to be a statistically significant determinant of income inequality, while healthcare expenditure and education expenditure are statistically significant determinants of income inequality. For each model across both income inequality metrics, unemployment rates and net trade were statistically significant. Additionally, while government expenditure was removed from S80s20model2 and cochS80s20model2, it was statistically significant at the 100% level on each other model, although this may be because of the 100% significance for education expenditure and healthcare expenditure influencing the overall significance, as social protection was social protection expenditure was found statistically significant at the 99% and 95% levels in comparison. Therefore, the model is unable to conclude government expenditure into social protection is a statistically significant determinant of income inequality.

6. Conclusion

Overall, government expenditure into education and healthcare as rate of total GDP are each significant determinant in reducing income inequality. The results for government expenditure into social protection are inconclusive. Net trade was found to be statistically significant in reducing income inequality. Unemployment was found to be the strongest positive significant determinant of income inequality. The remaining explanatory variables were found to be insignificant determinants of income inequality.

The results finding net trade is a negative determinant of income inequality supports Asteriou *et al.* (2014) findings. This can be attributed to the EU single market, allowing developed economies to benefit from their infrastructure and highly skilled workforce, increasing total exports value. However, in America, high export tariffs reduce profitability potentially leading to losses. This promotes increasing imports. Government expenditure into healthcare and education were found to be significant negative determinants of income inequality aligning with Marrero and Rodríguez (2012) findings. This may be attributed to the extremely high cost of non-government assisted healthcare plans. As demonstrated in the Nordic countries, increasing investment into public services reduces income inequality levels. Government expenditure into education reduces inequality due to the higher levels of education reducing the potential of long-term unemployment. Finding net FDI inflow insignificant is contradictory to Alderson and Nielsen (1999) and Mihaylova (2015) findings. However, this study focused on net FDI rather than total FDI inflow as rate of GDP. This suggests developed economies in the EU are strongly investing in other EU countries, ensuring the skilled income rates rise in accordance across EU countries, rather than focusing on single countries who receive a disproportionate rate of FDI inflow to FDI outflow.

The insignificance of tertiary education attainment align with Sánchez and Pérez-Corral (2018) finding education expenditure not significant on reducing income inequality. This may be because the tertiary education attainment data consisted of 30–34-year-olds, who are less likely to become unemployed due to having sufficient experience. Additionally, this may be a result of stagnating skilled income rates due to the increasing levels of individuals with tertiary education. (Checchi and García-Peñalosa 2009) Alternatively, this may simply be a result of the lagged effects of education on economies.

This area should be revisited when more data becomes available incorporating the short-term and long-term effects of Covid-19 and associated restrictions. Particularly studying the effects of closing businesses or furloughing employees for extended periods of time and how this has impacted income inequality in the short term along with the potential long-term social and economic impacts.

7. Bibliography

Alderson, A.S., Nielsen, F. (1999) 'Income Inequality, Development, and Dependence: A Reconsideration', *American Sociological Review*, 64(4), 606–631.

Altman, N., Krzywinski, M. (2016) 'Regression diagnostics', *Nature Methods*, 13(5), 385–386.

Asteriou, D., Dimelis, S., Moudatsou, A. (2014) 'Globalization and income inequality: A panel data econometric approach for the EU27 countries', *Economic Modelling*, 36, 592–599.

Balcilar, M., Gupta, R., Ma, W., Makena, P. (2021) 'Income inequality and economic growth: A re-examination of theory and evidence', *Review of Development Economics*, n/a(n/a), available: https://onlinelibrary.wiley.com/doi/abs/10.1111/rode.12754 [accessed 6 May 2021].

Barro, R.J. (1979) 'On the Determination of the Public Debt', *Journal of Political Economy*, 87(5), 940–971.

Barro, R.J. (2000) 'Inequality and Growth in a Panel of Countries', *Journal of Economic Growth*, 5(1), 5–32.

Bohn, H. (1990) 'Tax Smoothing with Financial Instruments', *The American Economic Review*, 80(5), 1217–1230.

Brueckner, M., Lederman, D. (2015) *Effects of Income Inequality on Aggregate Output* [online], Policy Research Working Papers, The World Bank, available: http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-7317 [accessed 6 May 2021].

Bubbico, R.L., Freytag, L. (2018) 'Inequality in Europe', *European Investment Bank*, 50.

Castells-Quintana, D., Ramos, R., Royuela, V. (2015) 'Income inequality in European Regions: Recent trends and determinants', *Review of Regional Research*, 35(2), 123–146.

Checchi, D., García-Peñalosa, C. (2009) 'Labour Market Institutions and the Personal Distribution of Income in the OECD', *Economica*, available: http://doi.wiley.com/10.1111/j.1468-0335.2009.00776.x [accessed 7 May 2021].

Cobham, A., Sumner, A. (2014) 'Is inequality all about the tails?: The Palma measure of income inequality', *Significance*, 11(1), 10–13.

Craney, T.A., Surles, J.G. (2002) 'Model-Dependent Variance Inflation Factor Cutoff Values', *Quality Engineering*, 14(3), 391–403.

Dabla-Norris, E., EDabla-Norris@imf.org, Kochhar, K., KKochhar@imf.org, Suphaphiphat, N., NSuphaphiphat@imf.org, Ricka, F., FRicka@imf.org, Tsounta, E., ETsounta@imf.org (2015) 'Causes and Consequences of Income Inequality: A Global Perspective', *Staff Discussion Notes*, 15(13), 1.

De Battisti, F., Porro, F., Vernizzi, A. (2019) *The Gini Coefficient and the Case of Negative Values*, SSRN Scholarly Paper ID 3610829, Social Science Research Network, Rochester, NY, available: https://papers.ssrn.com/abstract=3610829 [accessed 1 May 2021].

Dorn, F., Fuest, C., Potrafke, N. (2018) *Globalization and Income Inequality Revisited*, SSRN Scholarly Paper ID 3143398, Social Science Research Network, Rochester, NY, available: https://papers.ssrn.com/abstract=3143398 [accessed 6 May 2021].

Dreger, C., Reimers, H.-E. (2013) 'Does euro area membership affect the relation between GDP growth and public debt?', *Journal of Macroeconomics*, 38, 481–486.

Economic Inequality - Overview, How to Measure, the Gini Coefficient [online] (2021) *Corporate Finance Institute*, available: https://corporatefinanceinstitute.com/resources/knowledge/economics/economic-inequality/ [accessed 1 May 2021].

Education Statistics - All Indicators | DataBank [online] (2020) *World Bank DataBank*, available: https://databank.worldbank.org/source/education-statistics-%5e-all-indicators [accessed 28 Apr 2021].

Eurostat (2020) Income Quintile Share Ratio (S80/S20) by Sex - Products Datasets - Eurostat [online], *Eurostat*, available: https://ec.europa.eu/eurostat/web/products-datasets/-/tessi180 [accessed 28 Apr 2021].

Eurostat (2021) Gini Coefficient of Equivalised Disposable Income - EU-SILC Survey - Products Datasets - Eurostat [online], *Eurostat*, available: https://ec.europa.eu/eurostat/web/products-datasets/-/ilc_di12 [accessed 28 Apr 2021].

Filauro, S., Fischer, G. (2021) 'Income inequality in the EU: General trends and policy implications', *VoxEU.org*, available: https://voxeu.org/article/income-inequality-eu-trends-and-policy-implications [accessed 29 Apr 2021].

Fischer, R.D. (2001) 'The evolution of inequality after trade liberalization', *Journal of Development Economics*, 12th INTERAMERICAN SEMINAR ON ECONOMICS, 66(2), 555–579.

Frost, J. (2017) 'How To Interpret R-squared in Regression Analysis', *Statistics By Jim*, available: http://statisticsbyjim.com/regression/interpret-r-squared-regression/ [accessed 15 Mar 2021].

Gastwirth, J.L. (1971) 'A General Definition of the Lorenz Curve', *Econometrica*, 39(6), 1037–39.

In 't Veld, J. (2019) 'The Economic Benefits of the EU Single Market in Goods and Services', *Journal of Policy Modeling*, 41.

Jakobsson, A., Bourdet, Y., Gullstrand, J. (2006) 'Trade Openness and Income Inequality', 51.

Kiselakova, D., Sofrankova, B., Onuferova, E., Cabinova, V. (2020) 'Assessing the effect of innovation determinants on macroeconomic development within the EU (28) countries', *Problems and Perspectives in Management*, 18(2), 277–289.

Kolev, G., Niehues, J. (2016) 'The Inequality-Growth Relationship - An Empirical Reassessment', *IW-Report*.

Kuznets, S. (1955) 'Economic Growth and Income Inequality', *The American Economic Review*, 45(1), 1–28.

Larch, M. (2010) *Fiscal Performance and Income Inequality : Are Unequal Societies More Deficit-Prone? Some Cross-Country Evidence* [online], European Commission. Directorate-General for Economic and Financial Affairs, European Commission: BE, available: https://data.europa.eu/doi/10.2765/42402 [accessed 6 May 2021].

Leow, K.W., Tan, E.C. (2019) 'Determinants of poverty: A dynamic panel data analysis with controls for income level and inequality', *Malaysian Journal of Economic Studies*, available: https://search.informit.org/doi/abs/10.3316/informit.815682654149888 [accessed 29 Apr 2021].

Maestri, V., Roventini, A. (2012) *Inequality and Macroeconomic Factors: A Time-Series Analysis for a Set of OECD Countries*, 2012/21, LEM Papers Series, LEM Papers Series, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, available: https://ideas.repec.org/p/ssa/lemwps/2012-21.html [accessed 7 May 2021].

Marrero, G.A., Rodríguez, J.G. (2012) 'Inequality of Opportunity in Europe', *Review of Income and Wealth*, 58(4), 597–621.

Martinez-Vazquez, J., Moreno-Dodson, B., Vulovic, V. (2012) *The Impact of Tax and Expenditure Policies on Income Distribution: Evidence from a Large Panel of Countries*, SSRN Scholarly Paper ID 2188608, Social Science Research Network, Rochester, NY, available: https://papers.ssrn.com/abstract=2188608 [accessed 6 May 2021].

Michálek, A., Výbošťok, J. (2019) 'Economic Growth, Inequality and Poverty in the EU', *Social Indicators Research*, 141(2), 611–630.
Mihaylova, S. (2015) 'Foreign direct investment and income inequality in Central and Eastern Europe', *Theoretical and Applied Economics*, XXII (2015), 23–42.
Mussini, M. (2017) 'Decomposing Changes in Inequality and Welfare Between EU Regions: The Roles of Population Change, Re-Ranking and Income Growth', *Social Indicators Research*, 130(2), 455–478.
Sánchez, Á., Pérez-Corral, A.L. (2018) 'Government Social Expenditure and Income Inequalities in the European Union', *Revista Hacienda Pública Española*, 227(4), 133–156.
Solt, F. (2020) 'Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database', *Social Science Quarterly*, 101(3), 1183–1199.
Sylwester, K. (2000) 'Income inequality, education expenditures, and growth', *Journal of Development Economics*, 63(2), 379–398.
Szmigiera, M. (2021) Pay Gap between CEOs and Workers by Country 2018 [online], *Statista*, available: https://www.statista.com/statistics/424159/pay-gap-between-ceos-and-average-workers-in-world-by-country/ [accessed 29 Apr 2021].
Tertiary Educational Attainment by Sex, Age Group 30-34 - Products Datasets - Eurostat [online] (2021) *Eurostat*, available: https://ec.europa.eu/eurostat/web/products-datasets/-/t2020_41 [accessed 29 Apr 2021].
Trapeznikova, I. (2019) 'Measuring income inequality', *IZA World of Labor*, available: https://wol.iza.org/articles/measuring-income-inequality/long [accessed 1 May 2021].
United Nations (2021) World Economic Situation and Prospects 2021 | Department of Economic and Social Affairs [online], *United Nations: Department of Economic and Social Affairs Economic Analysis*, available: https://www.un.org/development/desa/dpad/publication/world-economic-situation-and-prospects-2021/ [accessed 28 Apr 2021].
World Development Indicators | DataBank [online] (2021) available: https://databank.worldbank.org/source/world-development-indicators [accessed 28 Apr 2021].
Yobero, C. (2016) RPubs - Methods for Detecting and Resolving Heteroskedasticity: An R Tutorial [online], available: https://rpubs.com/cyobero/187387 [accessed 15 Mar 2021].
Zach (2019) 'A Guide to Multicollinearity & VIF in Regression', *Statology*, available: https://www.statology.org/multicollinearity-regression/ [accessed 15 Mar 2021].