Assignment 1: Data and Descriptive

Anine Therese Karlsen & Mona Lisa Jones

This assignment aims to acquire, process, and analyze sub-national GDP and population data for a subset of European countries. Calculate GDP per capita and explore regional inequity using various descriptive statistics and visualizations.

1 Introduction

While national GDP and GDP per capita are vital indicators of a country's aggregate economic health, they do not shed light on how wealth or income is distributed among its residents. A high national GDP can, paradoxically, coexist with pockets of regional deprivation (**lessmann2017?**).

The truth of this statement becomes more evident when taking a closer look in to into sub-national data. Regional wealth disparities are of prime concern, especially when crafting policies for equitable growth (**lessmann2017?**). A country's macro-level prosperity does not automatically guarantee that all its regions partake equally in this wealth. By studying smaller regions within a country, it is possible to get a more nuanced narrative about the state of regional economic disparities (**lessmann2017?**).

This assignment is anchored in this very premise. It seeks to acquire, process, and analyse sub-national GDP and population data for a selected subset of European countries, namely France, Denmark, Hungary, Portugal, and Slovakia, spanning the period from 2000 to 2020. The overarching aim is to calculate GDP per capita for these regions and delve into the intricacies of regional inequity. This exploration will be underpinned by various descriptive statistics, visualizations, and analytical techniques to paint a comprehensive picture of regional economic landscapes.

Using time series data observing the GDP per capita and GINI trends across regions and over time, we can discern whether certain areas are ahead or behind in economic performance.

2 Literature review

"In recent decades, the regional distribution of incomes within countries has attracted considerable interest among academics and policy makers." (**lessmann2017?**)

Slovakia, experienced significant economic growth. Their GDP per capita grew faster than the older EU member states, albeit from a lower starting point, leading to a convergence (**fullart?**).

The global financial crisis and subsequent European debt crisis had significant impacts on European economies. GDP per capita declined in several countries, particularly those hit hardest by the crisis, such as Greece, Spain, and Portugal. Some regions took many years to recover to pre-crisis levels, while others, especially in Northern Europe, recovered more quickly (**lewis?**).

Major metropolitan areas, like Paris in France, London in the UK, and Berlin in Germany, often grew faster than other regions in their respective countries, leading to increasing spatial disparities (**kiuru2019?**).

Denmark maintained relatively high levels of GDP per capita throughout this period, benefiting from a diversified economy, strong institutions, and a high degree of economic openness. It did face challenges during the 2008 economic crisis but recovered relatively quickly (**dynamics?**).

Slovenia, which joined the EU in 2004, showed growth in GDP per capita, though it too was affected by the global economic downturn in 2008. However, it managed to recover in the subsequent years (**fullart?**).

Regional disparities persisted, with the Île-de-France region (which includes Paris) significantly outperforming other regions in terms of GDP per capita. The gap between urban and rural areas also remained a topic of discussion (**regional?**).

Towards the end of this period, in early 2020, the COVID-19 pandemic posed a new set of challenges for European economies, causing significant contractions in GDP per capita across many regions (**regional?**).

3 Part A: Sub-national GDP and GDP per Capita

3.1 Data Acquisition and datasets

The first thing we did in this assignment was to acquire our data. Trough Eurostat, we could download the datasets nama_10r_3gdp and demo_r_pjanggr3 as csv files, as well as filter the data by our preferences before downloading it. As per the assignment, we filtered the dataset by choosing the years 2000 to 2020. Furthermore, we selected the NUTS 3 region for the nations we were given, which were Portugal, France, Hungary, Slovakia and Denmark. Finally, we specified the data to be in million Euro.

3.1.1 GDP (nama_10r_3gdp)

The nama_10r_3gdp dataset from Eurostat provides insights into GDP at regional level using the NUTS classification system. It furnishes GDP values in both current prices and adjusted for inflation, with some figures given in purchasing power standards (PPS) to account for price level differences between countries. The data is often structured by year and region Eurostat (2023a).

The GDP at market prices represents the final result of production activities of resident producer units within a region or nation. It is calculated as the sum of the gross value added across various institutional sectors or industries, augmented by taxes and reduced by subsidies on products (which are not allocated to specific sectors or industries). This also balances out in the total economy production account. In terms of methodology, while national accounts compile GDP from the expenditure side, regional accounts don't adopt this approach due to the complexities of accurately mapping inter-regional flows of goods and services.

The different measures for the regional GDP are absolute figures in € and Purchasing Power Standards (PPS), figures per inhabitant and relative data compared to the EU Member States average Eurostat (2023c).

3.1.2 Population (demo_r_pjanggr3)

Using the NUTS categorization once more, Eurostat's demo_r_pjanggr3 records annual population changes at the regional level. This dataset includes information on births, deaths, net migration, and may also include demographic information on age and gender. It's also often displayed in a year-by-region format, and the data usually spans in yearly intervals Eurostat (2023b).

Eurostat's primary source for yearly demographic data at the regional level stems from the Unified Demography (Unidemo) project. The project covers 37 countries and is the central repository for demographic and migration-related data. Specific metrics gathered under UNIDEMO encompass population counts at the close of the calendar year and events such as births and deaths occurring within that year. Additionally, data on marriages, divorces, and migration flows are recorded.

For the purpose of this research, the demographic data references the NUTS 2016 classification, which provides a detailed breakdown of the European Union's territory Eurostat (2021).

3.1.3 NUTS classification

The Nomenclature of Territorial Units for Statistics (NUTS) offers a stratified system to segment the economic territory of the EU (including the UK) to facilitate the consistent collection and harmonization of regional statistics across Europe. The NUTS regions range from NUTS 0 Country level to NUTS 3 small units such as municipalities level.

3.2 GDP per Capita Calculation

The formula for calculating GDP per Capita is as follows:

$y_i = GDP_i/population_i$

After calculating the GDP per capita for all NUTS 3 regions in our assigned countries, we can see that there is a large spread between the figures for the various regions. In this assignement we want to look at regional inequity; in order to do this in a valuable way we have to divide between the different countries. By doing this, we can gain important insights on regional differences that we can utilize, for instance, to discuss national policy on equity and sustainable economic development in regions.

3.2.1 Descriptive statistics

In this part we will report and interpret different types of essential descriptive statistics. Measuring regional income inequality is challenging due to heterogeneity of regions. The number of regions in our data set varies largely in size and population. Since the focus of this paper is purely growth and changes in inequities over time, the variations of size and population density becomes a minor issue because the country-level territorial heterogeneity is fixed.

"Interest in income inequality has led to the development of several ways of measuring it. Two types of measures are of interest in this paper—-static and dynamic. *Static measures* provide a snapshot of these inequalities at a point of time whereas the dynamic *measures capture historical trends.*" (læreboken)

In this part, we'll look at GDP per capita for our assigned countries on a NUTS 3 level. In addition, we'll use different kinds of descriptive statistics in order to further 0,37, is analyse this data. In this analysis, we'll use Wooldridge (2020) for help.

By using figures, we can visualize the GDP per capita, and look at how it varies among the different regions. In these figures, a line represent one NUTS 3 region.

Mean

Calculate the mean to provide a representative value for a dataset, facilitating understanding of its central tendency and serving as a benchmark against which deviations and anomalies can be assessed, in later steps when building and interpreting regression models.

MMR

A comparison of the GRDP (gross regional domestic product) per capita of the region with the highest income to the region with the lowest income (minimum per capita GRDP) provides a measure of the range of these disparities. If this measure is small (close to 1), then it would mean that the different regions have relatively equal incomes. If this measure is large, then the interpretation is more problematic, as it does not tell us if the high ratio is due to substantial variation in the distribution of per capita GDRPs or the presence of outliers. Nevertheless, maximum to minimum ratio (MMR) provides a quick, easy to comprehend, and politically powerful measure of regional income inequality.

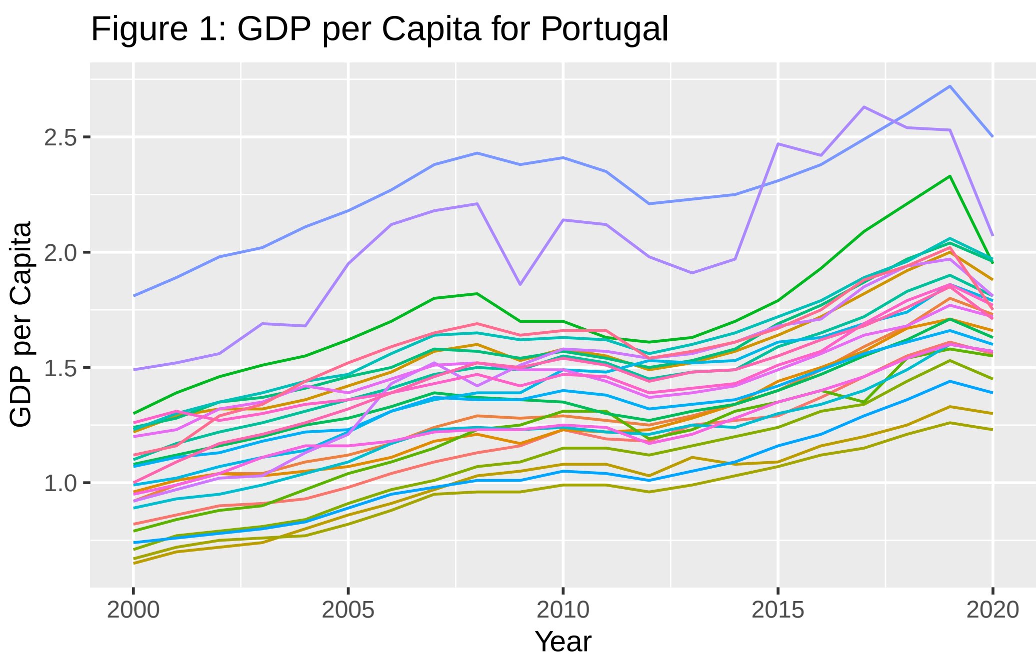
Standard diviation (SD)

Calculating SD to quantify the dispersion or variability of a data set around its mean. Helping us assess the degree of uncertainty, variability, or risk associated with an economic variable or parameter, which is crucial for understanding the reliability of estimations and predictions (Wooldridge, 2020).

Median

The median serves as a robust measure of central tendency, especially when a dataset may have outliers or is skewed. Unlike the mean, the median is not influenced by extreme values and, thus, can provide a clearer picture of the "typical" value in situations where the data distribution is not symmetrical (Wooldridge, 2020).

3.2.2 Portugal

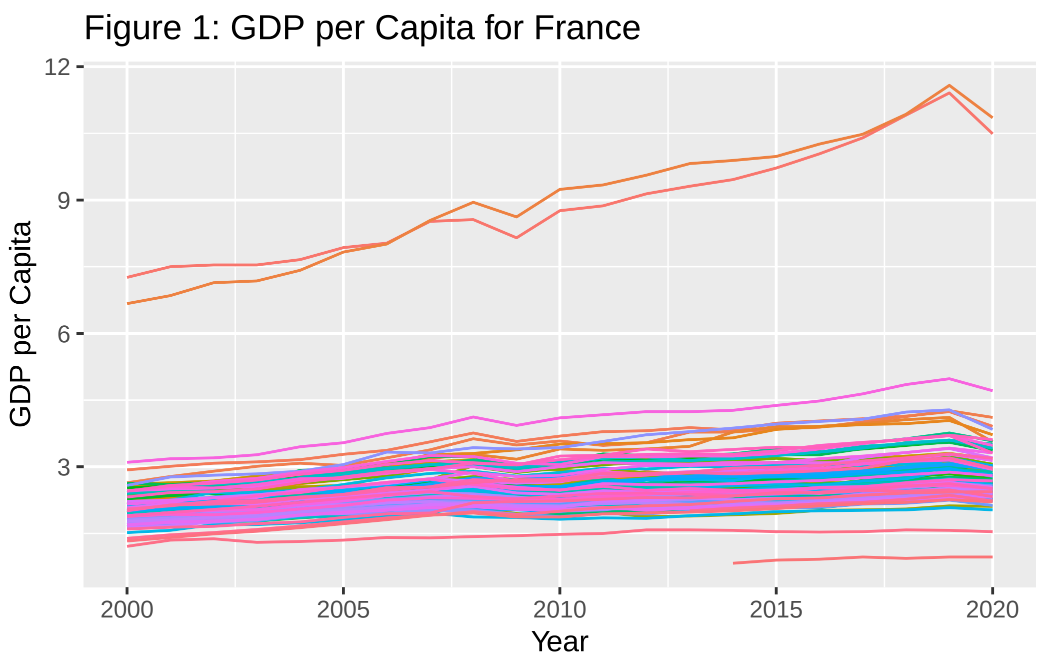



```
        GDP_per_capita
mean         1.4185524
median       1.3900000
std_dev      0.3702905
minimum      0.6500000
maximum      2.7200000
```

By looking at figure for Portugal, we can see that the GDP per capita in Portugal's regions appears to be fairly consistent. There is however some regional variability. We can see that the regions around the big cities like Lisbon have a higher GDP per capita compared to some more rural areas. Since Lisbon is the capital of Portugal, there is probably a higher concentration of industries, making it a economic center (which again makes the GDP per capita higher).

To continue, we can see that the mean is a little higher that the median, something that might indicate that regions like Lisbon are pulling up the average. If we compare the standard derivation for Portugal with the other countries, we'll see that is fairly low in comparison. This might mean that there is not a lot of variability between the GDP per capita across different regions in Portugal. The gap between minimum and maximum is also low compared to other countries, something that'll also show us that the economic disparity in Portugal might not be as high as it is in other countries.

3.2.3 France

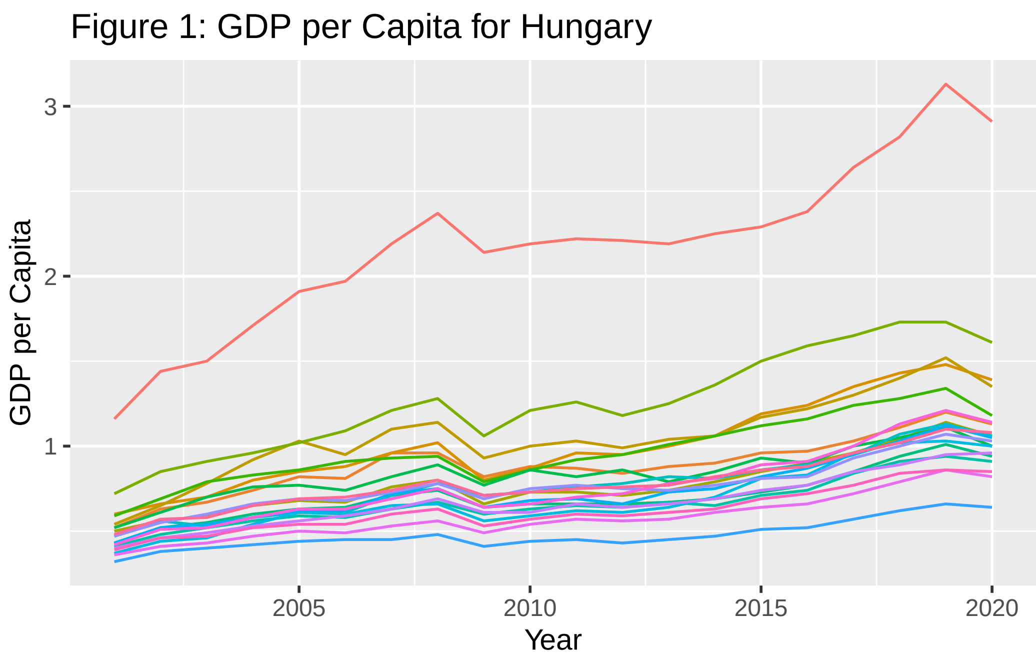



```
        GDP_per_capita
mean          2.630444
median        2.440000
std_dev       1.053767
minimum       0.830000
maximum      11.580000
```

When looking at the figure for France, we can right away see that there are some regions that have a much higher GDP per capita compared to the other regions. The region that has the highest GDP per capita for all years is the île-de-France region, one that also includes Paris. This significant difference between the regions with the highest GDP per capita and the lowest, shows us that there is a high concentration of economic activity and wealth in a few urban regions. Similar to Portugal, we can also again see that there is a difference between urban and rural regions.

Just as in Portugal, there is also a higher mean in France as well. Something that is different from the data in France compared to Portugal, is that the standard derivation is higher, and the difference between minimum and maximum is large. This strengthens what we have look at earlier in the figure, with some regions having a high concentration of wealth.

3.2.4 Hungary

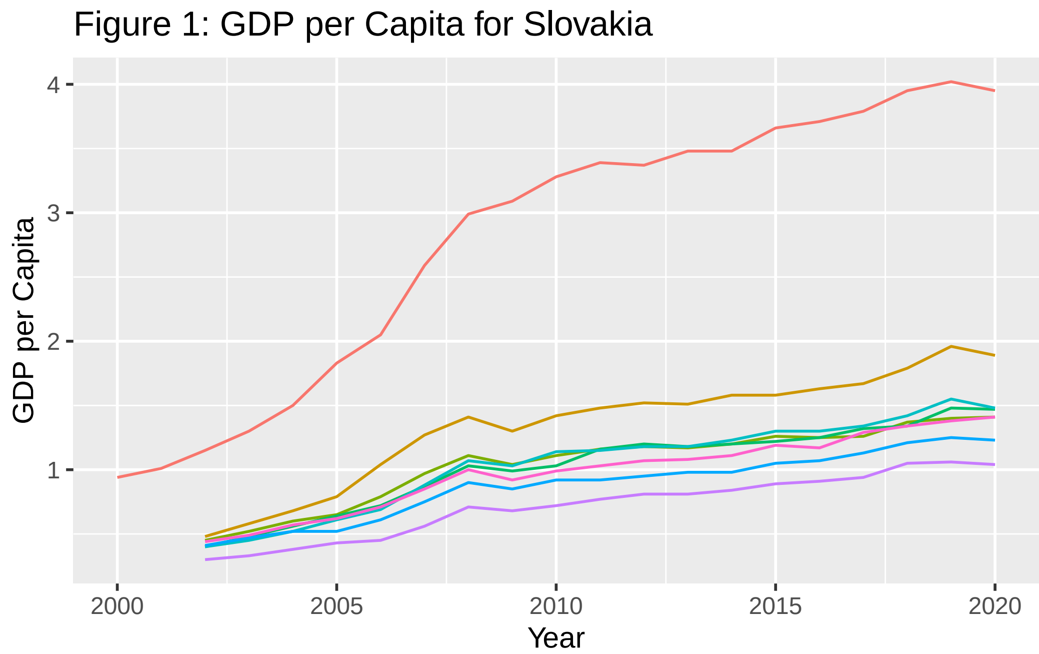



```
        GDP_per_capita
mean         0.8598000
median       0.7650000
std_dev      0.4059723
minimum      0.3200000
maximum      3.1300000
```

In Hungary, most of the regions have similar GDP per Capita. One region that sticks out by having a higher value, is the region of Budapest, Hungary's biggest city.

We have here as well an mean that is larger than the median, high standard derivation, and a large gap between minimum and maximum.

3.2.5 Slovakia



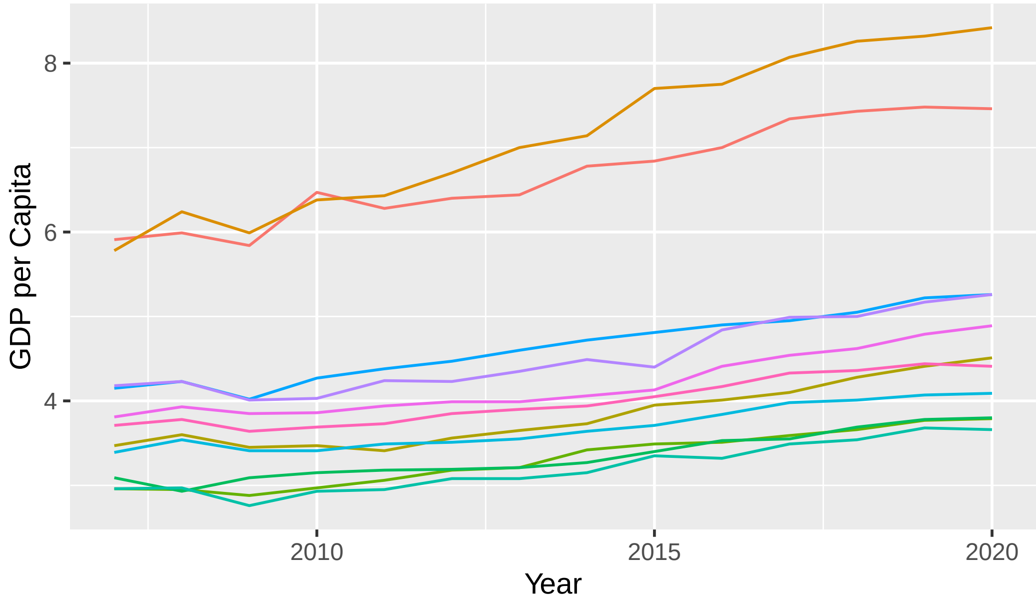
	GDP_per_capita
mean	1.2501948
median	1.0950000
std_dev	0.8018259
minimum	0.3000000
maximum	4.0200000

In Slovakia as well, we have one region that has a much higher GDP per capita than the rest of the regions. This region is Bratislava, which is the biggest city and capital, something that might point to this city being the economic capital of Slovakia as well.

Slovakia has also a mean higher that the median, and a large gap between minimum and maximum. In addition, the standard derivation is pretty high, meaning that there is some regions (or one region in this case) that is far away from the rest of the regions when it comes to economic development.

3.2.6 Denmark

Figure 1: GDP per Capita for Denmark




```
        GDP_per_capita
mean          4.419221
median        4.000000
std_dev       1.343933
minimum       2.760000
maximum       8.420000
```

Lastly, we have Denmark. We can see similar pattern here as well, with the capital Copenhagen being one of the regions with the highest GDP per capita.

We can also see the same as the previous countries, with the mean being higher than the median, which shows us that regions like Copenhagen might drag the mean up by being much larger than the rest of the regions.

4 Part B: Regional Inequity

4.1 Gini Coefficient Calculation

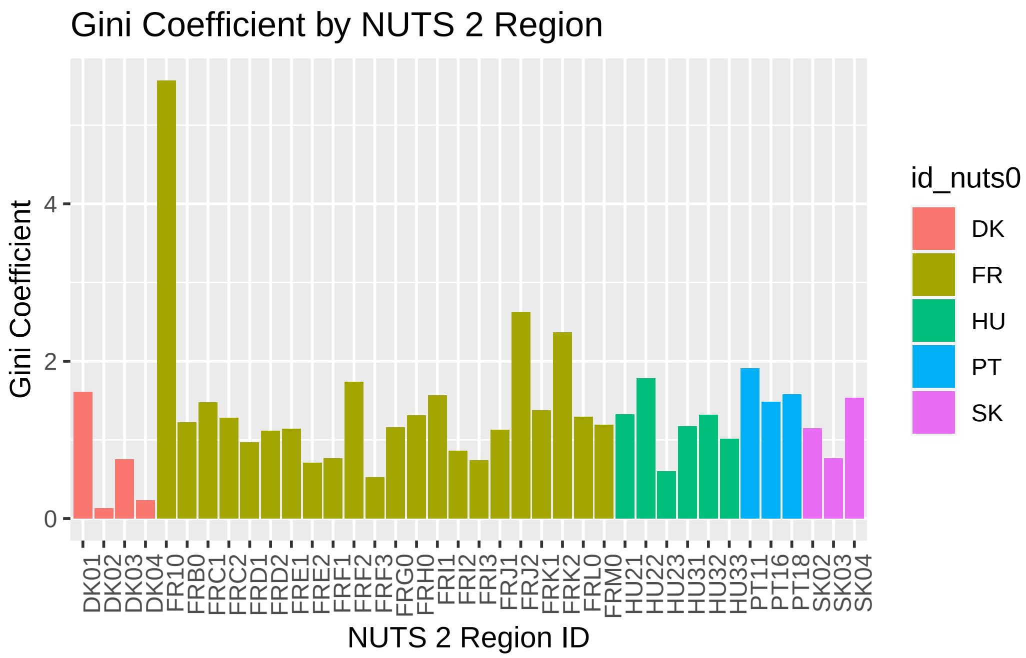
With the use of the NUTS3 GDP per capita data and this formula:

$GINW_j = \frac{1}{2\bar{y}_j} \sum_i^{n_j} \sum_l^{n_j} \frac{p_i}{P_j} \frac{p_l}{P_j} |y_i - y_l|$

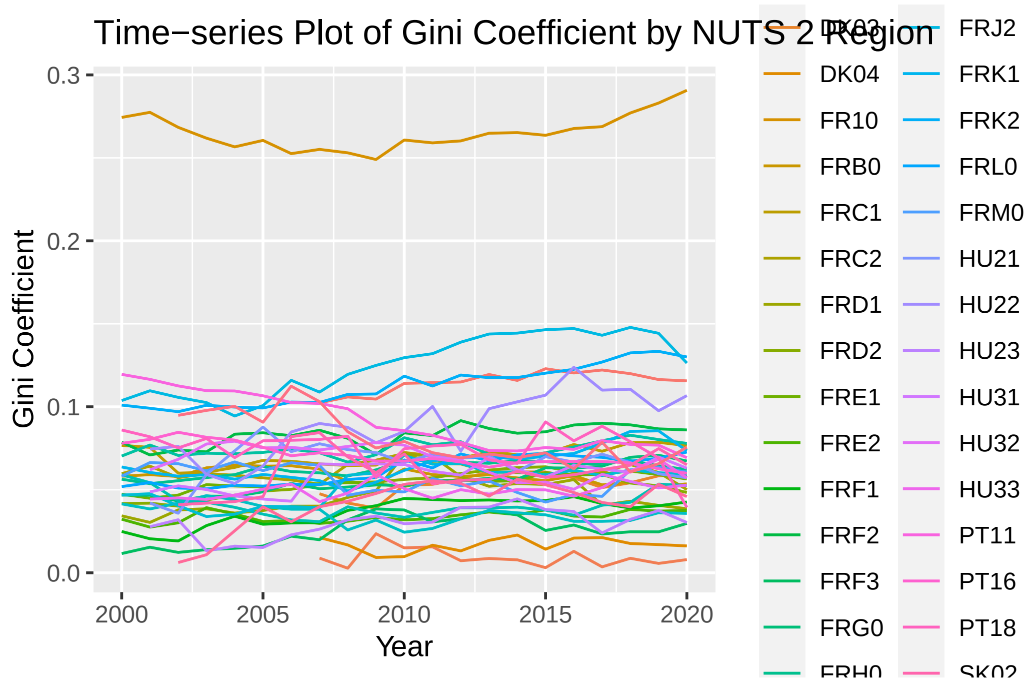
we will compute the population-weighted GDP Gini coefficient for each European NUTS2 region in our assigned countries.

The gini coefficient can help us measure inequality in a distribution, as is therefore a useful tool for us to use when we look at regional inequity. The closer the gini coefficient is to 1, the bigger the inequality is; a number closer to 0 equals equality. When looking at the gini coefficient for NUTS 2 regions, we also get a better overview over differences in income between different regions, and it also makes it easier to find the reasons as to why there is a difference between the regions Hasell & Roser (2023).

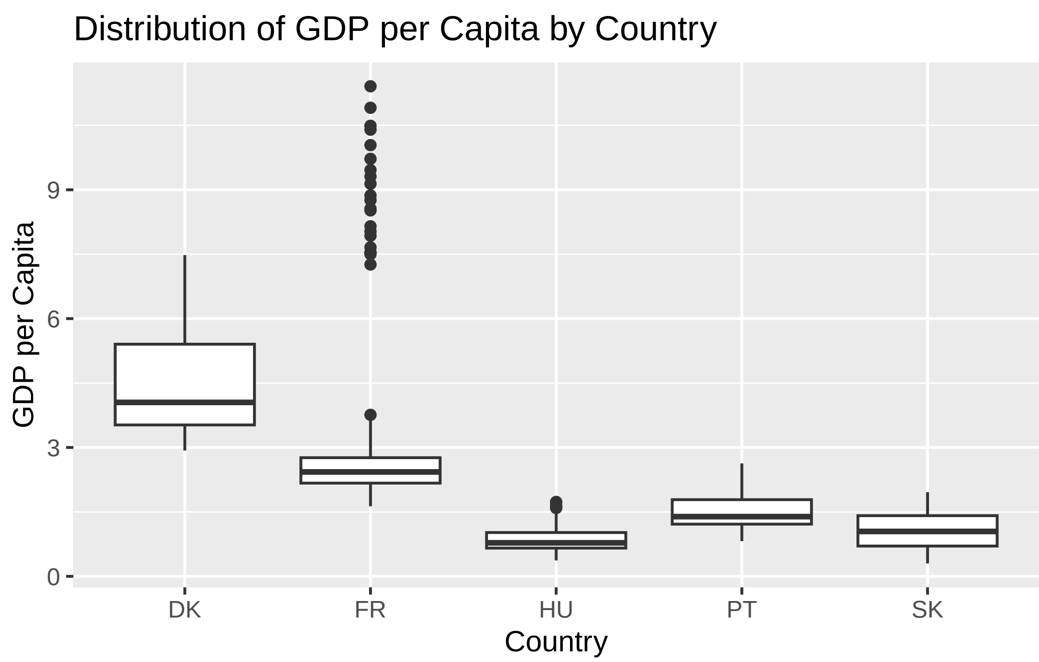
After calculating the gini coefficients, we can see that there are some similarities to the data we got from GDP per capita for NUTS 3 regions. In order to see these similarities better, as well as look for other important aspects that can be provided trough the calculations, we will visualize the data in three different ways.



...


Time−series Plot of Gini Coefficient by NUTS 2 Region
Gini Coefficient
0.3
0.2
0.1
0.0
2000
2005
2010
2015
2020
Year
DK03
DK04
FR10
FRB0
FRC1
FRC2
FRD1
FRD2
FRE1
FRE2
FRF1
FRF2
FRF3
FRG0
FRH0
FRJ2
FRK1
FRK2
FRL0
FRM0
HU21
HU22
HU23
HU31
HU32
HU33
PT11
PT16
PT18
SK02

...


Distribution of GDP per Capita by Country
GDP per Capita
9
6
3
0
DK
FR
HU
PT
SK
Country

... p. 116

4.2 Discussion

5 Assignment 2: Cross-sectional Estimates

5.1 2A: Growth and Inequity

5.1.1 1. Model Estimation

Scale-Location Plot is another way to check for homoscedasticity. The spread of the residuals should be roughly the same across all levels of the fitted values. If the spread increases or decreases with the fitted values, it suggests heteroscedasticity.

Residuals vs. Leverage Plot to identify influential observations that might unduly influence the regression. Points with high leverage or large residuals (outliers) can be identified. Cook's distance lines may be added to help identify points that have a large influence on the model.

$GINI = \beta_0 + \beta_1 \cdot \text{GDP per capita} + \epsilon$

Gini is the dependent variable (y), while GDP per capita is the independent variable (x).

	All
(Intercept)	0.055 ***
	(0.010)
log_GDP_per_capita	0.019
	(0.011)
r.squared	0.080
adj.r.squared	0.055
statistic	3.142
p.value	0.085

*** p < 0.001; ** p < 0.01; * p < 0.05.

Regression statistics of all countries for the year 2010

Beta 0 = 0.0551 (intercept)

Beta 1 = 0.0193 (slope coefficient)

Regression statistics of all countries seperately for the year 2010

	SK	DK	HU	PT	FR
(Intercept)	0.069	-0.127	0.075 **	0.079	-0.028
	(0.010)	(0.085)	(0.010)	(0.015)	(0.026)
log_GDP_per_capita	-0.018	0.124	0.049	-0.008	0.107 ***
	(0.034)	(0.059)	(0.029)	(0.030)	(0.026)
r.squared	0.221	0.689	0.410	0.067	0.457
adj.r.squared	-0.558	0.534	0.262	-0.866	0.430
statistic	0.284	4.432	2.778	0.072	16.820
p.value	0.689	0.170	0.171	0.833	0.001

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

5.1.2 2. Model Diagnostics

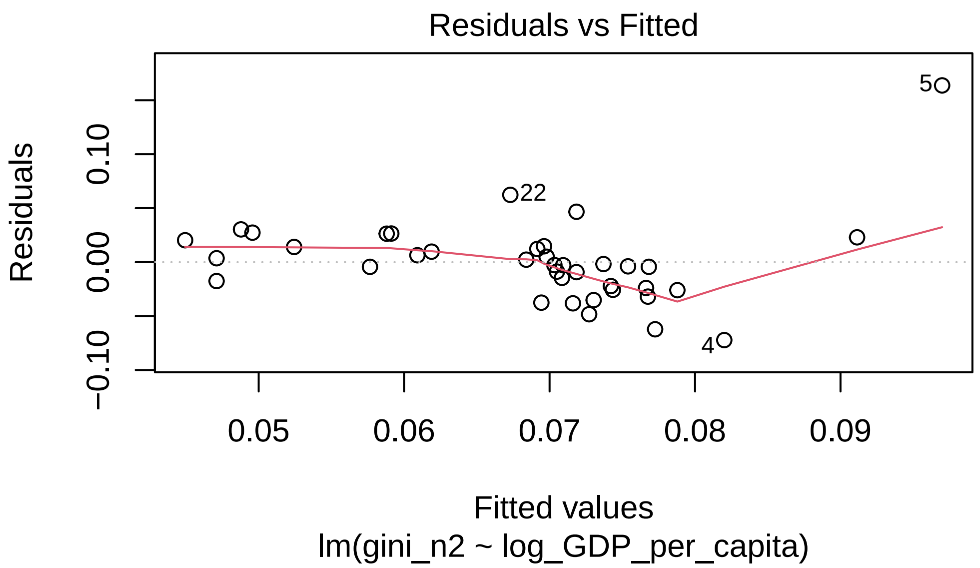
We'll now look at some of the numbers we got from the linear regression model (for all countries combined):

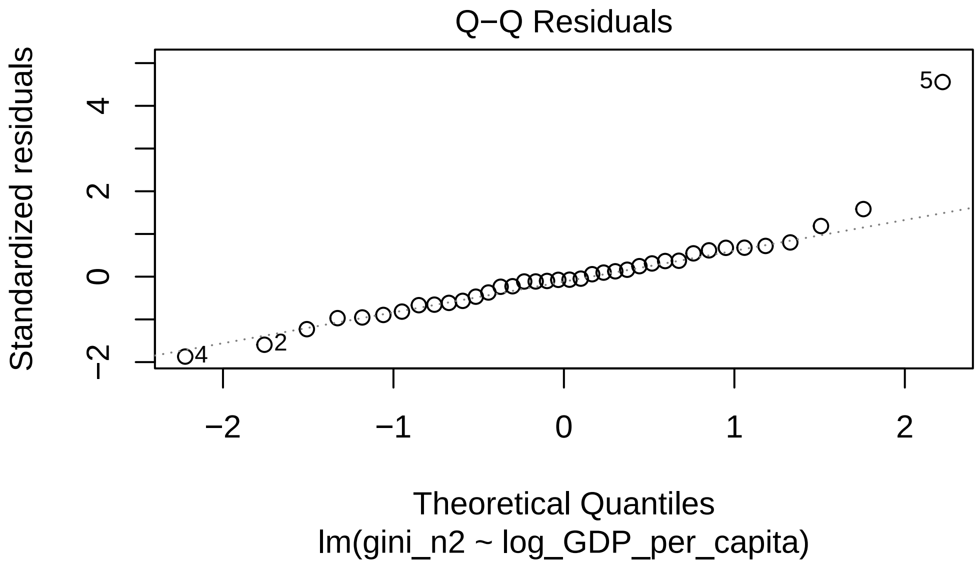
- Coefficients:
  - Intercept is 0.0551 (expected value of gini when GDP per capita is 0). Statistically significant (indicated by p-value).
  - Estimated coefficient for GDP per capita is 0.0193, if the natural logarithm of GDP per capita increase by one, then the gini coefficient will increase by 0.0193. The p-value associated with the coefficent is however not statistically significant at 5% level.
- Goodnes of fit:
  - Multiple R-squared (0.08028) indicate that around 8% of the variability in the GINI coefficient is explained by the model. This is low, which can suggest that the model dosen't explain the variation in gini.
  - Adjusted R-squared is even lower (0.05474), is therefore expected that the model dosen't really explain the variance in the dependent variable (gini).
- Model significance:
  - The F-statistic (3.142) indicate the significance of the regression model. We can see here, that with the p-value of 0.08474, the model isn't significant at a 5% level.

By looking at these numbers, we can see that the model may not reliably predict the gini coefficient. It also suggest that GDP per capita may not be a valid predictor of the gini coefficient.

We also examined our selected countries seperately in order to see how the reliability and validity of the model might vary between countries. However, since there are too few observations for most of the countries, it makes it hard to make a conclusion of the reliability. What we can see from this examination, is that the models for Denmark, Portugal and Slovakia are not statistically significant. France seem however to have significant coefficiants, and Hungary have a moderate R-squared (but lacks significance in the slope).

5.1.3 3. Visualization





We've made two plots that can help us understand the relationship between GDP per capita and the Gini coefficient, and that together with the regression statistics can help with discussing if the classical OLS assumptions hold for the model. The first one is a plot that shows us residuals vs fitted values, which can help us check the homoscedasticity assumption of a linear regression model. The residuals should be randomly scattered around the horisontal 0 line, something that can inducate that the variances of the error terms are constant. In our plot, the residuals are in some extent randomly distributed, and there is also no clear pattern; this suggest that there is likely no significant issues with heteroscedasticity or non-linearity.

The other plot - normal Q-Q plot - is used to assess if the residuals of the linear model are normally distributed. In our plot, most of the points follow the line closely, suggesting that the residuals are normally distributed. There are however some outliers in the tails, that suggest some variation from normality.

…

In both of these plots, we can see that there are some outliers, these can affect the fit of the model. These outliers might come from regions that have a high GDP per capita compared to rest of the regions, like Paris in France that is a financial hub.

```
`geom_smooth()` using formula = 'y ~ x'
```

In this plot we are visualizing the relationship between GDP per capita and gini by using a mix of the two previous plots. We can here as well, see extreme outliers.

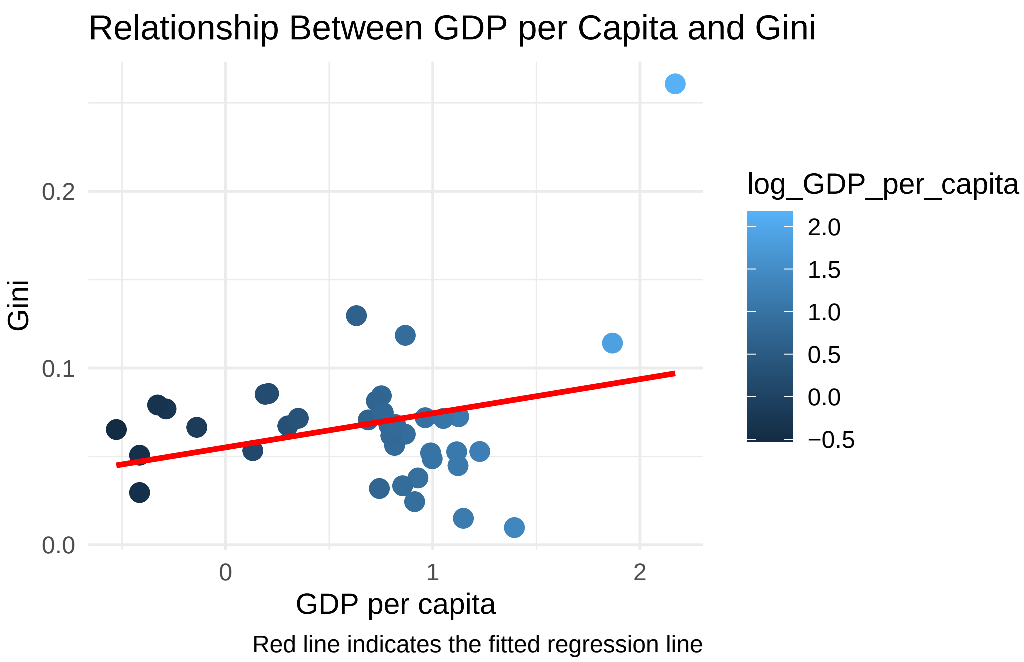


Figure 1: Relationship Between GDP per Capita and Gini

5.2 2B: Exploring Other Determinants of Inequity

5.2.1 1. Data Acquistion

In order to conduct a Multiple Linear Regression model, we need to have some independent variables to use in the model and compare them with the dependent variable. The first variable, education, can explain income inequality, since it can influence income distribution in a region. If access to education is unequal, then higher education levels might incease income disparities (Rodriguez-Pose & Tselios, 2008). The population density, our second variable, can explain inequality since regions with a higher population density might have different economic behaviours. Our last variable, rail network (infrastructure), can influence economic development and accessibility, which also can affect income inequality in a region (Chatterjee & Turnovsky, 2012).

5.2.2 2. Multiple Linear Regression Model

We will in this part do a Multiple Linear Regression model by using the variables education (in percentage of pupils and students in education, % of total population), population density and rail network in km. This model will tell us if these variables can help explain change in the gini coefficient.

In both our simple linear regression model, and now in our multiple linear regression model, we use the logarithmic function which makes it easier to linearize the relationship between the variables. By using it for GDP per capita, we can reflect changes more effectively. For rail network and population density, the logarithm function ensure that the model capture proportional changes and deals better with the wide range of values.

	Model	Model 2	Model 3
(Intercept)	0.050	0.008	-0.206
	(0.091)	(0.078)	(0.114)
log_GDP_per_capita	0.020	0.011	0.017
	(0.014)	(0.012)	(0.012)
students_percentage	0.000	-0.006	-0.001
	(0.004)	(0.004)	(0.006)
log(pop_density)		0.041 **	0.032 *
		(0.012)	(0.014)
log(rail_km)			0.020
			(0.018)
r.squared	0.098	0.370	0.467
adj.r.squared	0.033	0.300	0.370
statistic	1.515	5.284	4.812
p.value	0.237	0.005	0.006

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Multiple Linear Regression Model

5.2.3 3. Model Interpretation

Our first model in the Multiple Linear Regression model examine how education in addition to GDP per capita can help explain the gini coefficient. The second model look at both education, GDP per capita, and population density, while the third model examine them all and also add rail network.

For model 1, the adjusted R-squared is 0.033, indicating that the model explains around 3% of the variability in the gini coefficient. Since this is relatively low, it suggests that model 1 is not suitable in explaining the variance in gini.

6 Assignment 3: Alternative Functional Forms and Panel Estimates

6.1 Part A: Testing Development Effects Across Subsets

6.1.1 Subset Analysis:

To begin the third assignment, we will divide the data into different subsets. We have chosen to divide the countries based on size and population; we'll therefore have one subset with France, and one subset with the rest of the countries. We have several reasons for the dividing the countries like this:

1. France is one of the largest countries in Europe, with a great economic and political influence. The larger size and population of France may lead to different economic dyamics compared to the other smaller countries.
2. Larger countries might have a greater diversity of economic activities, wich affects the distribution of wealth and income differentely than in smaller countries with less diverse economies. We can see this by looking at how some of the NUTS 2 regions in France differ far more than in other countries.
3. The size and population can influence public policy, like social welfare, which impacts income distribution and inequality (source).
4. Larger countries often have larger internal markets, which can affect economic development and income distribution differently compared to smaller economies, where ecternal trade might be more crucial.

Population (as of 2023):

France: 67,75

Denmark: 5,85

Slovakia: 5,4

Hungary: 9,71

Portugal: 10,33

Size:

France: 551 695

Denmark: 42 952

Slovakia: 49 000

Hungary: 93 000

Portugal: 92 000

```r
# Dataframe with just france
gini_nuts_2_FR <- gini_nuts_2 |>
  filter(id_nuts0 == "FR", gini_n2 > 0)

# Dataframe with the other countries
gini_nuts_2_C <- gini_nuts_2 |>
  filter(id_nuts0 != "FR", gini_n2 > 0)
```

Descriptive analysis:

```r
summary(gini_nuts_2_FR)
```

```
  id_nuts3                year        population            GDP
 Length:462         Min.   :2000   Min.   : 120753   Min.   :  2285
 Class :character   1st Qu.:2005   1st Qu.: 311745   1st Qu.:  6783
 Mode  :character   Median :2010   Median : 524817   Median : 11606
                    Mean   :2010   Mean   : 665173   Mean   : 23585
                    3rd Qu.:2015   3rd Qu.: 682556   3rd Qu.: 17982
                    Max.   :2020   Max.   :2606234   Max.   :246937
 GDP_per_capita     id_nuts2           id_nuts0            gini_n2
 Min.   : 1.630   Length:462         Length:462         Min.   :0.01157
 1st Qu.: 2.170   Class :character   Class :character   1st Qu.:0.04343
 Median : 2.430   Mode  :character   Mode  :character   Median :0.05903
 Mean   : 2.750                                         Mean   :0.06960
 3rd Qu.: 2.763                                         3rd Qu.:0.07157
 Max.   :11.410                                         Max.   :0.29069
```

```r
summary(gini_nuts_2_C)
```

```
  id_nuts3                year        population           GDP
 Length:295         Min.   :2000   Min.   : 93259   Min.   : 1480
 Class :character   1st Qu.:2007   1st Qu.:360397   1st Qu.: 3290
 Mode  :character   Median :2011   Median :457344   Median : 4718
                    Mean   :2011   Mean   :473560   Mean   : 8189
                    3rd Qu.:2016   3rd Qu.:561764   3rd Qu.: 7851
                    Max.   :2020   Max.   :826244   Max.   :59251
 GDP_per_capita    id_nuts2           id_nuts0            gini_n2
 Min.   :0.300   Length:295         Length:295         Min.   :0.002747
 1st Qu.:0.770   Class :character   Class :character   1st Qu.:0.046143
 Median :1.180   Mode  :character   Mode  :character   Median :0.063624
 Mean   :1.743                                         Mean   :0.062308
 3rd Qu.:1.825                                         3rd Qu.:0.077469
 Max.   :7.480                                         Max.   :0.123817
```

Regression analysis:

```r
# France
lmFR <- lm(gini_n2 ~ log(GDP_per_capita), data = gini_nuts_2_FR)
# Other
lmC <- lm(gini_n2 ~ log(GDP_per_capita), data = gini_nuts_2_C)

huxreg(France = lmFR,
       Others = lmC,
       statistics = c("r.squared","adj.r.squared", "statistic", "p.value"))
```

	France	Others
(Intercept)	-0.034 ***	0.062 ***
	(0.005)	(0.002)
log(GDP_per_capita)	0.110 ***	0.001
	(0.005)	(0.002)
r.squared	0.490	0.000
adj.r.squared	0.489	-0.003
statistic	442.257	0.113
p.value	0.000	0.737

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Comparative analysis:

```r
# Welch two sample t-test
t.test <- t.test(gini_nuts_2_FR$gini_n2, gini_nuts_2_C$gini_n2)
print(t.test)
```

```
   Welch Two Sample t-test

data:  gini_nuts_2_FR$gini_n2 and gini_nuts_2_C$gini_n2
t = 2.6138, df = 741.78, p-value = 0.009136
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.001815167 0.012769002
sample estimates:
 mean of x  mean of y
0.06959963 0.06230754
```

6.1.2 Subset Analysis Discussion:

To conclude your analysis, you'll need to interpret and discuss your results with a focus on how the effects of development (as measured by GDP per capita) on inequality (measured by the Gini coefficient) differ between your two subsets: France, and the group of Hungary, Denmark, Portugal, and Slovakia. Here's a structured approach to crafting your conclusion:

1. ***Summary of Findings****: Start by summarizing the key findings from your analysis. This includes the main trends you observed in each subset and the results of your regression and comparative analyses. Highlight whether the relationship between GDP per capita and inequality was stronger, weaker, or different in nature in France compared to the other group of countries.*
2. ***Interpretation of Results****: Delve into what these findings mean. For instance, if you found that economic development has a more significant impact on reducing inequality in France than in the other countries, discuss possible reasons. These could include differences in social welfare policies, economic structures, or the scale of economic activities due to the size and population of France.*
3. ***Contextualization with Existing Literature****: Relate your findings to existing research and theories. How do your results align with or differ from previous studies on the relationship between economic development and inequality? Discuss any theories or models that help explain your results.*
4. ***Explanation of Variations****: Address the variations observed between the subsets. This could involve discussing how factors like economic diversity, public policy differences, market size, or international trade might influence the relationship between GDP per capita and inequality differently in larger versus smaller countries.*
5. ***Limitations****: Acknowledge any limitations in your study. This might include the limited number of countries in your analysis, the potential impact of external factors not considered in your study, or limitations in the data.*
6. ***Implications****: Discuss the broader implications of your findings. What do they suggest about economic policy in European countries? How might they inform future research or policy-making aimed at reducing inequality?*
7. ***Concluding Thoughts****: End with a concluding statement that encapsulates the essence of your findings and their significance in understanding economic development and inequality in Europe.*

Remember, a good conclusion not only summarizes findings but also provides a thoughtful interpretation, places the research in a broader context, and opens up avenues for future research or policy considerations.

6.2 Part B: Exploring Alternative Functional Forms

6.2.1 Functional Form Exploration:

```r
# Logarithmic model
log_model <- lm(gini_n2 ~ log(GDP_per_capita), data = gini_filtered)

# Quadratic model
quad_model <- lm(gini_n2 ~ GDP_per_capita + I(GDP_per_capita^2), data = gini_filtered)

# Cubic model
cubic_model <- lm(gini_n2 ~ GDP_per_capita + I(GDP_per_capita^2) + I(GDP_per_capita^3),

# Check the summary of each model
summary(log_model)
```

```
Call:
lm(formula = gini_n2 ~ log(GDP_per_capita), data = gini_filtered)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.075857 -0.022756 -0.005897  0.014948  0.187821 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         0.052001   0.002200  23.631   <2e-16 ***
log(GDP_per_capita) 0.021644   0.002414   8.965   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04018 on 755 degrees of freedom
Multiple R-squared:  0.09621,	Adjusted R-squared:  0.09502 
F-statistic: 80.38 on 1 and 755 DF,  p-value: < 2.2e-16
```

```r
summary(quad_model)
```

```
Call:
lm(formula = gini_n2 ~ GDP_per_capita + I(GDP_per_capita^2), 
    data = gini_filtered)

Residuals:
```

```
      Min        1Q    Median        3Q       Max
-0.093684 -0.017704 -0.000659  0.011892  0.124592

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.080381   0.003099   25.93   <2e-16 ***
GDP_per_capita      -0.019151   0.001839  -10.41   <2e-16 ***
I(GDP_per_capita^2)  0.003955   0.000197   20.07   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02821 on 754 degrees of freedom
Multiple R-squared:  0.555, Adjusted R-squared:  0.5538
F-statistic: 470.2 on 2 and 754 DF,  p-value: < 2.2e-16
```

```r
summary(cubic_model)
```

```
Call:
lm(formula = gini_n2 ~ GDP_per_capita + I(GDP_per_capita^2) +
    I(GDP_per_capita^3), data = gini_filtered)

Residuals:
      Min        1Q    Median        3Q       Max
-0.083703 -0.017575 -0.000724  0.012043  0.121584

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          8.267e-02  4.170e-03  19.828  < 2e-16 ***
GDP_per_capita      -2.191e-02  3.823e-03  -5.730 1.45e-08 ***
I(GDP_per_capita^2)  4.778e-03  1.021e-03   4.681 3.38e-06 ***
I(GDP_per_capita^3) -5.925e-05  7.208e-05  -0.822    0.411
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02822 on 753 degrees of freedom
Multiple R-squared:  0.5554,    Adjusted R-squared:  0.5537
F-statistic: 313.6 on 3 and 753 DF,  p-value: < 2.2e-16
```

Based on the output of the three models (logarithmic, quadratic, and cubic) for your data, let's analyze and compare them to determine which ones may provide a better fit for the relationship between GDP per capita and the Gini coefficient.

6.2.2 *Logarithmic Model*

- ***Adjusted R-squared**: 0.09502, indicating that approximately 9.5% of the variation in the Gini coefficient is explained by the model.*
- ***Significance**: Both the intercept and the log(GDP_per_capita) are highly significant ($p < 0.001$).*
- ***Model Fit**: The relatively low R-squared value suggests that while the model is statistically significant, it explains a small portion of the variance in the Gini coefficient.*

6.2.3 *Quadratic Model*

- ***Adjusted R-squared**: 0.5538, a substantial improvement over the logarithmic model, explaining about 55.38% of the variation in the Gini coefficient.*
- ***Significance**: All coefficients, including GDP_per_capita and its square, are highly significant.*
- ***Model Fit**: This model shows a much better fit than the logarithmic model, indicating that a quadratic relationship might be more appropriate for describing the data.*

6.2.4 *Cubic Model*

- ***Adjusted R-squared**: 0.5537, almost identical to the quadratic model, indicating that adding the cubic term doesn't substantially improve the model.*
- ***Significance**: The linear and quadratic terms are significant, but the cubic term (GDP_per_capita^3) is not ($p = 0.411$).*
- ***Model Fit**: The lack of significance of the cubic term and the similar R-squared value to the quadratic model suggest that the cubic model doesn't provide additional explanatory power over the quadratic model.*

6.2.5 *Conclusion and Justification*

*Based on these results, the **quadratic model** appears to be the best fit for your data. It significantly improves upon the logarithmic model and captures more of the variance in the Gini coefficient. The improvement in fit from the logarithmic to the quadratic model suggests that the relationship between GDP per capita and income inequality is not linear but follows a curved trajectory, which could align with economic theories like the Kuznets curve.*

*The **cubic model**, while adding complexity, does not provide a significantly better fit than the quadratic model. The non-significance of the cubic term suggests that this additional complexity is not justified.*

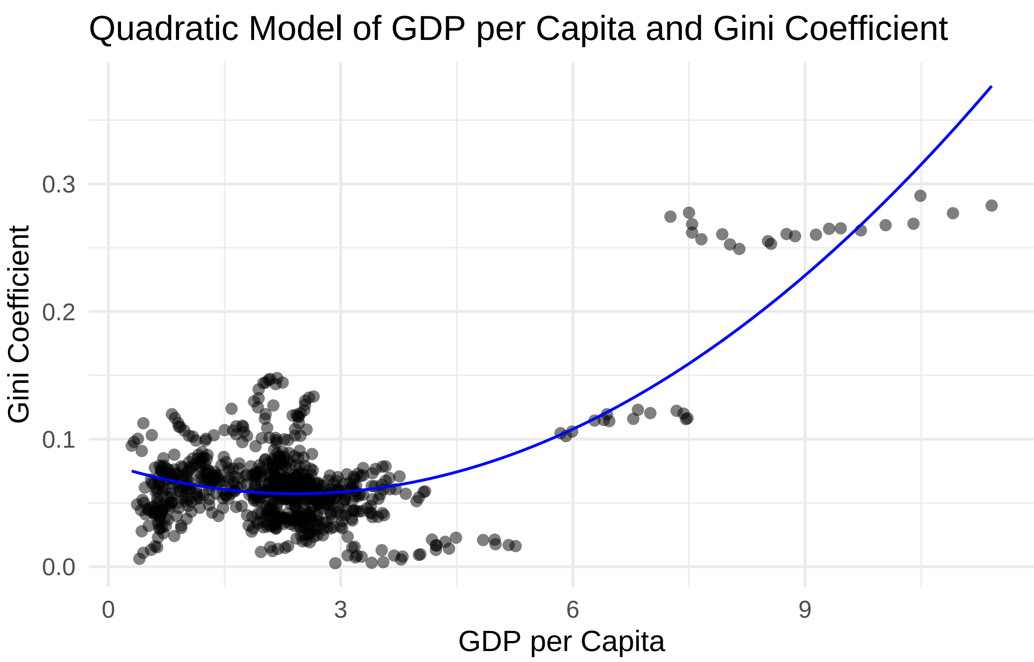
Therefore, I would recommend the quadratic model as the primary model for your analysis, with the logarithmic model as a secondary option for comparison. The quadratic model's *better fit and theoretical alignment with expectations of a non-linear relationship between economic development and inequality make it a strong choice for your analysis.*

6.2.6 Estimation and Visualization:

```
# Quadratic model

# Create a new data frame for plotting predicted values
plot_data <- data.frame(GDP_per_capita = seq(min(gini_filtered$GDP_per_capita), max(gini
plot_data$gini_pred <- predict(quad_model, newdata = plot_data)

# Plotting
ggplot(gini_filtered, aes(x = GDP_per_capita, y = gini_n2)) +
  geom_point(alpha = 0.5) +  # Plot the actual data points
  geom_line(data = plot_data, aes(x = GDP_per_capita, y = gini_pred), color = 'blue') +
  labs(title = "Quadratic Model of GDP per Capita and Gini Coefficient",
       x = "GDP per Capita",
       y = "Gini Coefficient") +
  theme_minimal()
```

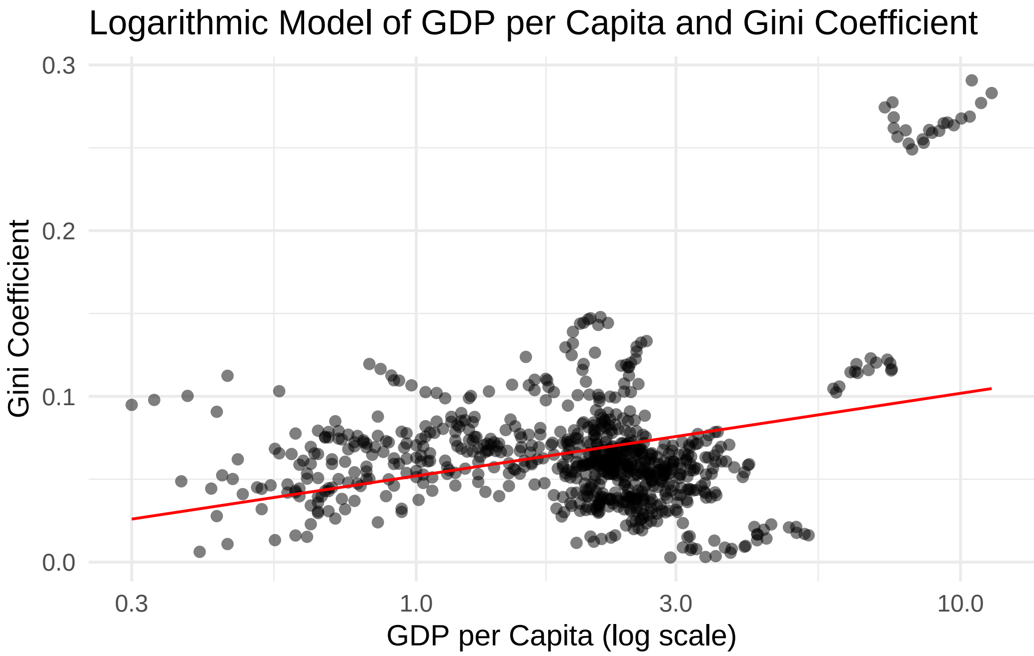



```
# Logarithmic model

# Create a new data frame for plotting predicted values
```

```r
plot_data <- data.frame(GDP_per_capita = seq(min(gini_filtered$GDP_per_capita), max(gin
plot_data$gini_pred <- predict(log_model, newdata = plot_data)

# Plotting
ggplot(gini_filtered, aes(x = GDP_per_capita, y = gini_n2)) +
  geom_point(alpha = 0.5) +  # Plot the actual data points
  geom_line(data = plot_data, aes(x = GDP_per_capita, y = gini_pred), color = 'red') +
  scale_x_log10() +  # Logarithmic scale for the x-axis
  labs(title = "Logarithmic Model of GDP per Capita and Gini Coefficient",
       x = "GDP per Capita (log scale)",
       y = "Gini Coefficient") +
  theme_minimal()
```

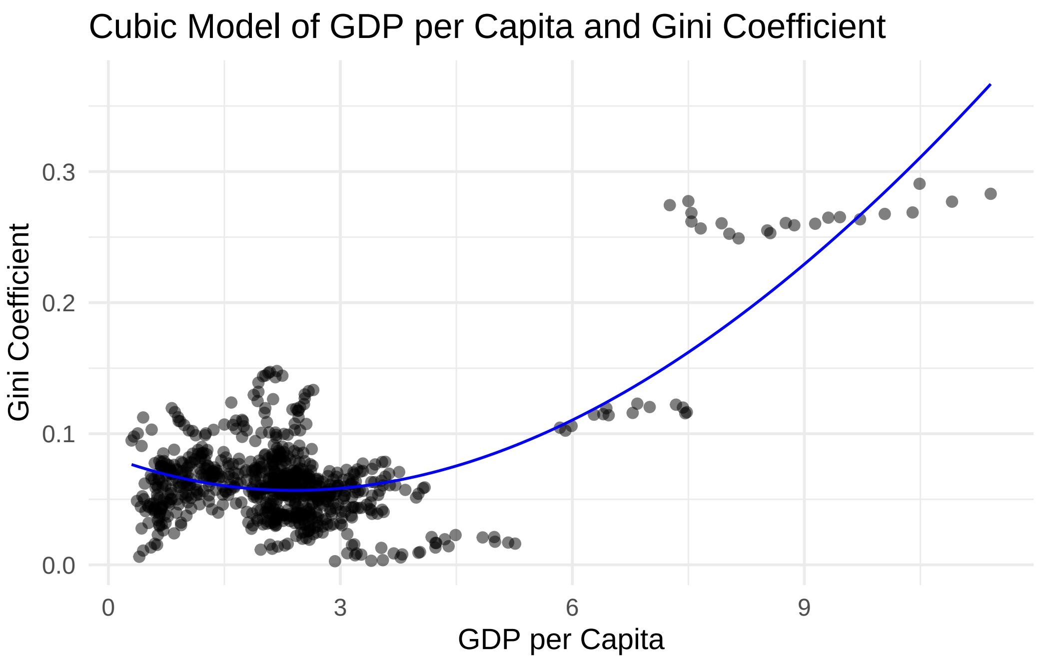



```r
# Assuming your cubic model is named 'cubic_model'
# and your dataframe is 'gini_filtered'

# Create a new data frame for plotting predicted values
plot_data <- data.frame(GDP_per_capita = seq(min(gini_filtered$GDP_per_capita), max(gin
plot_data$gini_pred <- predict(cubic_model, newdata = plot_data)

# Plotting
ggplot(gini_filtered, aes(x = GDP_per_capita, y = gini_n2)) +
  geom_point(alpha = 0.5) +  # Plot the actual data points
  geom_line(data = plot_data, aes(x = GDP_per_capita, y = gini_pred), color = 'blue') +
```

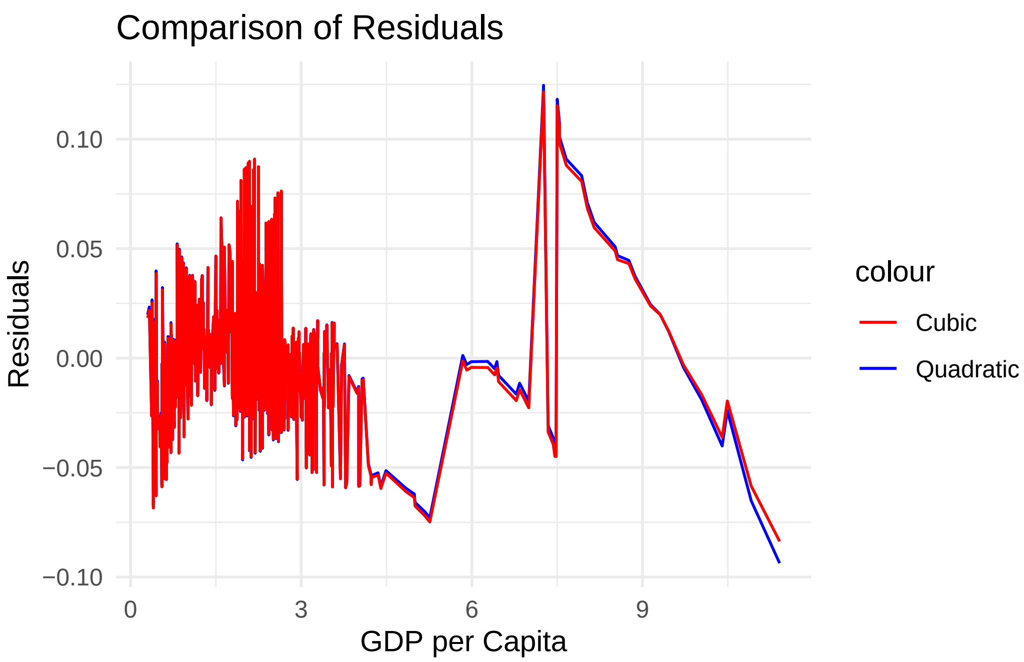
```r
  labs(title = "Cubic Model of GDP per Capita and Gini Coefficient",
       x = "GDP per Capita",
       y = "Gini Coefficient") +
  theme_minimal()
```




```r
# Calculate residuals
residuals_quad <- gini_filtered$gini_n2 - predict(quad_model, gini_filtered)
residuals_cubic <- gini_filtered$gini_n2 - predict(cubic_model, gini_filtered)
```

```r
# Create a dataframe for plotting residuals
residuals_data <- data.frame(GDP_per_capita = gini_filtered$GDP_per_capita,
                             Residuals_Quad = residuals_quad,
                             Residuals_Cubic = residuals_cubic)

# Plotting residuals
ggplot(residuals_data) +
  geom_line(aes(x = GDP_per_capita, y = Residuals_Quad, color = "Quadratic")) +
  geom_line(aes(x = GDP_per_capita, y = Residuals_Cubic, color = "Cubic")) +
  labs(title = "Comparison of Residuals",
       x = "GDP per Capita",
       y = "Residuals") +
  theme_minimal() +
  scale_color_manual(values = c("Quadratic" = "blue", "Cubic" = "red"))
```




```
sd(residuals_quad)
```

```
[1] 0.02817368
```

```
sd(residuals_cubic)
```

```
[1] 0.02816105
```

```
mean(residuals_quad^2)
```

```
[1] 0.0007927076
```

```
mean(residuals_cubic^2)
```

```
[1] 0.000791997
```

6.2.7 Results Interpretation:

Based on the numbers and plots from the logarithmic and quadratic models, let's discuss and compare these models in terms of their coefficients, implications, and overall fit compared to a basic linear model.

6.2.8 *Logarithmic Model*

6.2.8.1 *Coefficients*

- ***Intercept***: *Statistically significant, indicating that when GDP per capita is at its minimum (log-transformed), there is still a base level of inequality.*
- ***log(GDP_per_capita)***: *Also significant, implying that as GDP per capita increases, the Gini coefficient (a measure of inequality) changes in a way that's captured by the logarithmic transformation of GDP per capita.*

6.2.8.2 *Implications*

- *The logarithmic model suggests that the relationship between GDP per capita and income inequality is not linear but has a diminishing effect. As the economy grows, each additional unit of GDP per capita has a smaller impact on changing the level of inequality.*
- *This could imply that in lower-income ranges, increases in GDP per capita are more effective in reducing or increasing inequality than at higher income levels.*

6.2.8.3 *Fit*

- *The adjusted R-squared value is relatively low, indicating that while the model has statistical significance, it explains only a small portion of the variance in the Gini coefficient.*
- *Compared to a linear model, this suggests that a simple linear relationship does not adequately capture the dynamics between GDP and inequality.*

6.2.9 *Quadratic Model*

6.2.9.1 *Coefficients*

- ***Intercept***: *Significant, similar to the logarithmic model.*
- ***GDP_per_capita***: *The linear term is negative and significant.*
- ***GDP_per_capita^2***: *The quadratic term is positive and significant.*

6.2.9.2 *Implications*

- *The quadratic model suggests a more complex relationship that aligns with the Kuznets curve hypothesis: as an economy grows (to a point), inequality may initially decrease, but beyond a certain level of GDP pe-r capita, further growth could lead to increasing inequality.*
- *This model captures the possibility of a 'turning point' in economic development where the benefits of growth might start having less impact on reducing inequality or even begin to exacerbate it.*

6.2.9.3 *Fit*

- *The quadratic model has a significantly higher adjusted R-squared value than the logarithmic model, indicating a much better fit.*
- *This improved fit suggests that the quadratic model captures the nuances in the relationship between GDP per capita and income inequality more effectively than a simple linear or logarithmic model.*

6.2.10 *Comparison to the Original Linear Model*

- *Both the logarithmic and quadratic models provide more insights into the relationship between GDP per capita and income inequality compared to a basic linear model. The linear model would likely fail to capture the diminishing returns (as suggested by the logarithmic model) or the potential turning point (as suggested by the quadratic model).*
- *The quadratic model, with its higher explanatory power (adjusted R-squared), seems to be a more appropriate choice for capturing the complex, non-linear dynamics in the data.*
- *The choice between logarithmic and quadratic models should also consider the theoretical underpinnings of the economic relationships being studied, where the quadratic model aligns well with established economic theories like the Kuznets curve.*

In conclusion, while the logarithmic model offers a better fit than a simple linear model by introducing the concept of diminishing returns, the quadratic model provides an even more nuanced understanding of the relationship between GDP per capita and income inequality, highlighting the potential for a turning point in how economic growth affects inequality.

6.3 Part C: Heteroskedasticity Testing and Causality Discussion

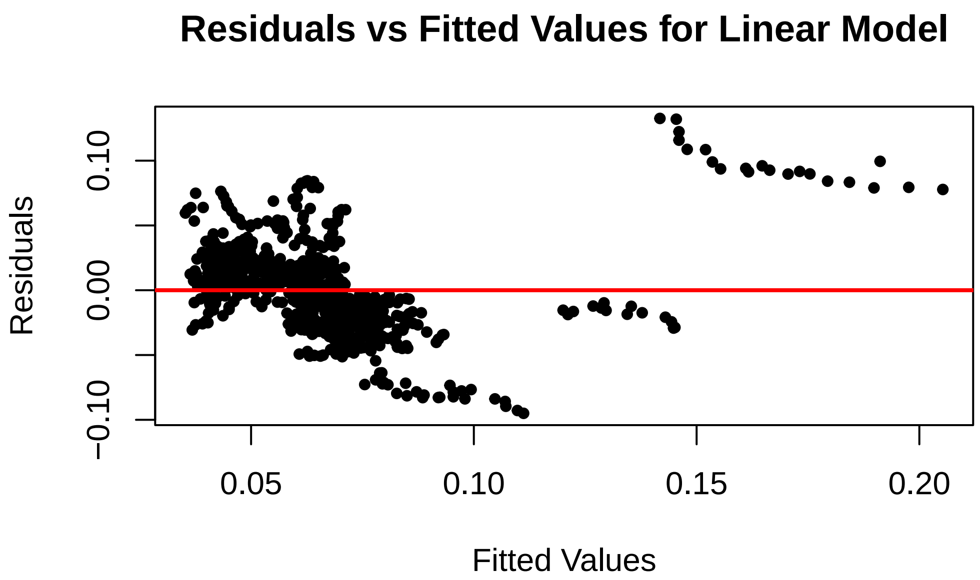
6.3.1 Heteroskedasticity Testing:

```
lm_01 <- lm(gini_n2 ~ GDP_per_capita, data = gini_filtered)
```

```
# Assuming you have a linear regression model called linear_model_ols
# Here's how to plot the residuals vs. fitted values for the linear model

# Obtain fitted values and residuals from the model
fitted_values <- fitted(lm_01)
residuals <- resid(lm_01)

# Create the plot
plot(fitted_values, residuals,
     xlab = "Fitted Values",
     ylab = "Residuals",
     main = "Residuals vs Fitted Values for Linear Model",
     pch = 20)
abline(h = 0, col = "red", lwd = 2) # Add a horizontal line at 0
```



Yes, the plot you've provided is a typical "Residuals vs Fitted Values" plot for a linear regression model. It's used to assess the assumption of homoscedasticity (constant variance *of residuals across levels of the predictor variable). Here's what you can infer from the plot:*

- ***Horizontal Red Line****: This line represents where residuals would be if they were perfectly predicted without any error. Residuals above the line were underpredicted by the model, and those below the line were overpredicted.*
- ***Pattern in Residuals****: Ideally, residuals should be randomly dispersed around the red line, with no clear pattern. In your plot, there seems to be a funnel shape, with the residuals spreading out as the fitted values increase. This pattern is a classic indication of heteroskedasticity, suggesting that the variance of the residuals is not constant across all levels of fitted values.*
- ***Presence of Outliers****: Towards the right end of the plot, the points are widely spread out from the center line, indicating potential outliers or extreme values that the model does not predict well.*

If this is your plot, and these are the correct interpretations based on what you see, it indicates that your linear model may be suffering from heteroskedasticity and possibly from outliers or high-leverage points as well. You might need to consider model diagnostics and remedial measures, such as transforming variables, adding missing variables, or using robust standard errors.

```r
# Assuming you have a linear_model, log_model, and quad_model already estimated

# Breusch-Pagan test for the original linear model
bptest_linear <- bptest(lm_01)

# Breusch-Pagan test for the logarithmic model
bptest_log <- bptest(log_model)

# Breusch-Pagan test for the quadratic model
bptest_quad <- bptest(quad_model)

# Output the test results
print(bptest_linear)
```

```
	studentized Breusch-Pagan test

data:  lm_01
BP = 231.48, df = 1, p-value < 2.2e-16
```

```r
print(bptest_log)
```

```
	studentized Breusch-Pagan test

data:  log_model
BP = 142.82, df = 1, p-value < 2.2e-16
```

```r
print(bptest_quad)
```

```
	studentized Breusch-Pagan test

data:  quad_model
BP = 80.325, df = 2, p-value < 2.2e-16
```

The Breusch-Pagan test is used to detect heteroskedasticity in a regression model. The test works by checking if the variances of the errors from the regression are dependent on the values of the independent variables, which would violate one of the key assumptions of ordinary least squares (OLS) regression.

Here are the results for each of the models you tested:

6.3.2 *Original Linear Model (lm_01)*

- ***BP Statistic**: 231.48*
- ***Degrees of Freedom**: 1*
- ***p-value**: < 2.2e-16*

6.3.3 *Logarithmic Model (log_model)*

- ***BP Statistic**: 142.82*
- ***Degrees of Freedom**: 1*
- ***p-value**: < 2.2e-16*

6.3.4 *Quadratic Model (quad_model)*

- ***BP Statistic**: 80.325*
- ***Degrees of Freedom**: 2*
- ***p-value**: < 2.2e-16*

6.3.5 *Analysis of the Results*

For all three models, the p-values are extremely small (less than 2.2e-16), which is effectively zero for practical purposes. This strongly suggests that we can reject the null hypothesis of homoscedasticity (constant variance of errors) for all models. In other words, there is a presence of heteroskedasticity.

The BP statistic, which is chi-square distributed, is substantially large for all models, further supporting the rejection of the null hypothesis.

6.3.6 *Degrees of Freedom*

The degrees of freedom correspond to the number of independent variables in the model that are being tested for their association with the variance of the residuals. The original linear model and the logarithmic model have 1 degree of freedom, suggesting that only one independent variable was included in each of these models. The quadratic model has 2 degrees of freedom, indicating that there are two terms being considered (GDP per capita and its square).

6.3.7 *Implications*

The presence of heteroskedasticity means that the standard errors of the coefficients estimated by OLS may be biased, which can lead to incorrect conclusions when conducting hypothesis tests and calculating confidence intervals. It's especially important in econometrics and other fields where inference is as important as prediction.

To address heteroskedasticity, you may consider using robust standard errors (also known as heteroskedasticity-consistent standard errors) or transforming the model, perhaps by using a different functional form or applying weighted least squares (WLS) regression. It's also worthwhile to investigate the source of heteroskedasticity which could be due to outliers, omitted variables, or a misspecified model.

To address heteroskedasticity in your models, you can use robust standard errors, which are designed to provide consistent standard errors even when the homoscedasticity assumption is violated. In R, you can calculate robust standard errors using the `vcovHC()` *function from the* `sandwich` *package and the* `coeftest()` *function from the* `lmtest` *package.*

```
# Assuming your model objects are named lm_01, log_model, and quad_model

# Calculate robust standard errors for the original linear model
coeftest_lm_01 <- coeftest(lm_01, vcov = vcovHC(lm_01, type = "HC"))

# Calculate robust standard errors for the logarithmic model
coeftest_log_model <- coeftest(log_model, vcov = vcovHC(log_model, type = "HC"))
```

```r
# Calculate robust standard errors for the quadratic model
coeftest_quad_model <- coeftest(quad_model, vcov = vcovHC(quad_model, type = "HC"))

# Print out the results
print(coeftest_lm_01)
```

```
t test of coefficients:

               Estimate Std. Error t value  Pr(>|t|)    
(Intercept)   0.0306764  0.0035457  8.6516 < 2.2e-16 ***
GDP_per_capita 0.0153020  0.0016849  9.0816 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
print(coeftest_log_model)
```

```
t test of coefficients:

                    Estimate Std. Error t value  Pr(>|t|)    
(Intercept)        0.0520005  0.0023696  21.945 < 2.2e-16 ***
log(GDP_per_capita) 0.0216443  0.0045083   4.801 1.904e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
print(coeftest_quad_model)
```

```
t test of coefficients:

                     Estimate  Std. Error t value  Pr(>|t|)    
(Intercept)         0.08038055  0.00348132 23.0891 < 2.2e-16 ***
GDP_per_capita     -0.01915147  0.00229651 -8.3394 3.537e-16 ***
I(GDP_per_capita^2)  0.00395477  0.00029606 13.3581 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This will provide you with the t-statistics and p-values that are robust to heteroskedasticity for the coefficients in each of your models. If you find that the significance of your coefficients *changes when using robust standard errors, it suggests that your original standard errors (and thus t-statistics and p-values) were indeed affected by heteroskedasticity.*

As for other measures to deal with heteroskedasticity:

- ***Weighted Least Squares (WLS)**: If you can model the variance of the residuals as a function of the independent variables, you can use this variance function to weight your observations inversely to their variance.*
- ***Transformation of Variables**: In some cases, applying a transformation to the dependent variable (such as a log transformation) can help stabilize the variance of the residuals.*
- ***Adding Missing Variables or Interaction Terms**: Sometimes heteroskedasticity is a symptom of model misspecification. If important variables or interactions are omitted, adding them can improve homoskedasticity.*
- ***Checking for Outliers**: Large outliers can cause heteroskedasticity. Investigate any data points that have large residuals and determine if they are valid observations or data errors.*

Remember, while robust standard errors can adjust for heteroskedasticity, they do not correct the underlying cause of it, so it's often worthwhile to explore these other options as well.

The robust standard errors have been recalculated for your models, and the t-test of coefficients has been performed using these robust errors. Let's interpret the results for each model:

6.3.8 *Original Linear Model*

- ***Intercept**: The coefficient for the intercept is significant ($p < 2.2e\text{-}16$), indicating that when GDP per capita is zero, the expected value of the Gini coefficient is approximately 0.0307.*
- ***GDP_per_capita**: The coefficient for GDP per capita is also significant ($p < 2.2e\text{-}16$), which suggests a positive relationship between GDP per capita and the Gini coefficient. For every one-unit increase in GDP per capita, the Gini coefficient is expected to increase by about 0.0153.*

6.3.9 *Logarithmic Model*

- ***Intercept**: The intercept is significantly different from zero ($p < 2.2e\text{-}16$), with an expected Gini coefficient value of 0.052 when the log of GDP per capita is zero (which corresponds to GDP per capita being 1, due to the log transformation).*
- ***log(GDP_per_capita)**: The coefficient for the logarithm of GDP per capita is significant ($p = 1.904e\text{-}06$), indicating that as the GDP per capita increases by a certain percentage, the Gini coefficient increases by about 0.0216.*

6.3.10 *Quadratic Model*

- ***Intercept***: *The intercept is significantly different from zero ($p < 2.2e\text{-}16$).*
- ***GDP__per__capita***: *The linear term of GDP per capita is significant and negative ($p = 3.537e\text{-}16$), suggesting that initially, as GDP per capita increases, the Gini coefficient decreases.*
- ***I(GDP__per__capita^2)***: *The quadratic term is significant and positive ($p < 2.2e\text{-}16$), indicating that at higher levels of GDP per capita, the effect reverses, and the Gini coefficient starts to increase, confirming the presence of a quadratic relationship.*

6.3.11 *Overall Interpretation*

All three models show significant coefficients with very low p-values, even after adjusting for heteroskedasticity using robust standard errors. This reaffirms the statistical significance of your findings.

- *The linear model suggests a direct, positive relationship between GDP per capita and the Gini coefficient.*
- *The logarithmic model indicates a diminishing effect, where percentage increases in GDP per capita have a consistent impact on the Gini coefficient.*
- *The quadratic model supports the idea of the Kuznets curve, with inequality first decreasing and then increasing as GDP per capita grows.*

The robust standard errors have provided more reliable estimates of the standard errors, ensuring that your significance tests are valid despite the presence of heteroskedasticity. This reinforces the credibility of your model coefficients and the conclusions you can draw from them.

6.3.12 Causality Discussion:

6.4 Part D: Panel Estimates

6.4.1 Panel Estimation Task:

6.4.2 Panel Estimation Analysis:

Panel Estimation Discussion:

Chatterjee, S., & Turnovsky, S. J. (2012). Infrastructure and inequality. *European Economic Review*, *56*(8), 1730–1745.

Eurostat. (2021). *Population change - Demographic balance and crude rates at regional level (NUTS 3) (demo_r_gind3)*. https://ec.europa.eu/eurostat/cache/metadata/en/demo_r_gind3_esms.ht

Eurostat. (2023a). *Gross domestic product (GDP) at current market prices by NUTS 3 regions.* https://ec.europa.eu/eurostat/databrowser/view/nama_10r_3gdp/default/table?lang=en.
Eurostat. (2023b). *Population on 1 January by broad age group, sex and NUTS 3 region.* https://ec.europa.eu/eurostat/databrowser/view/demo_r_pjanaggr3/default/table?lang=en.
Eurostat. (2023c). *Regional economic accounts (reg_eco10).* https://ec.europa.eu/eurostat/cache/metadata/
Hasell, J., & Roser, M. (2023). Measuring inequality: What is the Gini coefficient? *Our World in Data.*
Rodriguez-Pose, A., & Tselios, V. (2008). *EDUCATION AND INCOME INEQUALITY IN THE REGIONS OF THE EUROPEAN UNION.* https://click.endnote.com/viewer?doi=10.1111%2Fj.146
9787.2008.00602.x&token=WzM0MzQ0MzksIjEwLjExMTEvai4xNDY3LTk3ODcuMjAwOC4wMDYwMi
Avuk84iH8myPb7qaaaJIQk.
Wooldridge, J. M. (2020). *Introductory Econometrics - A Modern Approach* (7th ed.). Cengage.