Assignment 4:

Regional GDP Inequality in 4 Selected European Economies - Synthesised Results.

Kristoffer Tufta and Harald Vika

Wednesday 3 Dec, 2025

1 Assignment 4 Consolidation of Key findings

1.1 Abstract

1.2 Introduction

1.2.1 Background

1.2.2 Objetives

1.2.3 Sinificance

1.3 Literature Review - Kristoffer

1.3.1 Previous Work

A central reference point for our work is Lessmann & Seidel (2017), who shows that the link between economic development and regional inequality is not linear. Their results suggest something closer to an inverted-U (and in some cases N-shaped) pattern, where inequality tends to rise in the middle stages of development before falling again in high-income settings. This gives a natural expectation that fast-growing regions may show lower inequality if they are already relatively advanced, while poorer or developing regions might move in the opposite direction. Their work is also relevant because they rely on both traditional data and satellite-based luminosity data, which shows that inequality can be meaningfully captured even when official data is patchy — a recurring issue even in Eurostat datasets, as we experienced directly.

Eurostat's own methodological documentation is also part of the relevant background. Their notes on missing values help explain why regional GDP and population numbers are incomplete in certain years. These gaps arise from reclassification of regions, confidentiality rules, or delayed reporting. Understanding why these "quirks" happen is important because it directly affects our results.

For our segmentation work, we also build on Eurostat's definitions of population-density classes used for identifying rural areas, urban clusters, and high-density urban centres. While Eurostat's approach operates at the 1 km^2 continuous grid-cell level, their breakpoints (300 and 1500 people per square kilometer) give a practical structure for grouping NUTS2 regions by density. This matters because urban density is often correlated with both productivity and access to public-services, and therefore a likely influence when it comes to regional GDP inequality.

1.3.2 Research Gap

While there is existing literature on national GDP inequality trends and on broader development patterns, there is less focus on how short-term growth shocks play out inside individual NUTS2 regions, and even less on whether the effect of growth differs across structural categories like urban density, unemployment rates, or workforce size. Our assignments aim to examine exactly that: to see whether any of these structural categories influence GDP inequality in NUTS2-regions, and if so, to what degree.

1.4 Data and Methodology

1.4.1 Data Sources - Harald

Primary data refers to information collected firsthand by the researcher directly from original source. It is raw and unprocessed, typically gathered through methods such as surveys, interviews and observations. The data provided by Eurostat can be considered primary data because it is originally collected through official surveys and processed following established methodological standards for gathering information on specific topic within defined time periods.

In the first assignment, we used Eurostat to extract datasets on gross domestic product (GDP) at current market prices by NUTS 3 regions. From these datasets, we collected data for Germany, Ireland, Croatia annd Switzerland covering the years 2000 to 2023.

In the second assignment, we retrieved data from three additional datasets that could potentially influence regional inequality in 2017. These datasets included labour force by NUTS 2 region, unemployment rate by Nuts 2 region, and population density by NUTS 2 region. Each country collects its own data and reports it to Eurostat in accordance with ESA 2010, a global framework for national accounting. This ensures that the resulting statistics are consistent, reliable, and comparable across countries.

However, the data also comes with certain limitations. For example, Switzerland is not en EU member but an EFTA country (European Free Trade Association), which promotes free trade and economic cooperation. Because Switzerland did not have a formal datasharing agreement with Eurostate before 2008, some data for earlier years or specific topics may be missing (*Information on Data - National Accounts - Eurostat*, n.d.).

Another limitation relates to confidentiality. Ireland, for instance, lacks GDP data for 2015-2017 in the Mid-West and South-West regions due to confidentiality restrictions imposed by Ireland's Central Statistics Office. These restrictions are linked to changes in the national accounts in 2015 that significantly affected the measurement of Ireland's productive capacity (*County Incomes and Regional GDP 2015 - CSO - Central Statistics Office*, 2018).

Some counters also choose to delay the publication of certain statistics to ensure accuracy. As a result, data for the most recent years of 2022 and 2023, may still be incomplete due to its unavailability.

A common limitation across several datasets concerns changes in regional boundaries. Regions may merge to form larger areas or split into smaller units, affecting how data is collected, aggregated and updated. This issue is particularly relevant in Germany, where regional reforms occur more frequntly due to regional policy. Coratia and Ireland face similar challenges, as changes to their NUTS regions have affected population data availability at the NUTS 3 level.

Finally, there were limitations related to the time coverage of the datasets. Althought the assignments aimed to examine data from 2000 to 2023, the population density dataset for NUTS 2 regions only covers 2012-2023, and the unemployment rate dataset covers 2013-2023. Consequently, when including these variables, we were restricted to using data from 2013 to 2023 to maintain consistency across datasets.

1.4.2 Methodological approach - Kristoffer

1.4.2.1 Cross-Sectional Estimation

We first combined the Eurostat datasets covering GDP (in million euros) and population, , then calculated GDP per capita and Gini coefficients using a formula weighted by population, like Lessmann & Seidel (2017). We then visualized the calculated Gini coefficients for the different regions using different plots from ggplot2. In assignment 2, we made a cross-sectional analysis of the year 2017 using a regression model to test the effect of change in regional GDP per capita on regional inequality. We specified the following linear regression model:

$$\mathrm{Gini}_{i,2017} = \alpha + \beta\,\Delta\mathrm{GDPpc}_{i,2016\to2017} + u_i$$

Where:

- $\mathrm{Gini}_{i,2017}$ — Calculated Gini coefficient of region i in 2017 (regional GDP per capita inequality)
- $\Delta\mathrm{GDPpc}_{i,2016\to2017}$ — percent change in GDP per capita from 2016 to 2017 in region i
- α — intercept term (baseline inequality when GDP-per-capita change is zero)

- β — slope coefficient showing how inequality changes with a one-percentage-point increase in GDP-per-capita growth

- u_i — error term capturing unobserved regional factors

We then experimented with other determinants of inequality by making a new OLS model with a set of new variables: population density (persons per square kilometer), unemployment rate and total workforce, also fetched from Eurostat. We specified the MLR as follows:

$$\text{Gini}_{i,2017} = \alpha + \beta_1 \,\text{Workforce}_i + \beta_2 \,\text{PopDensity}_i + \beta_3 \,\text{Unemployment}_i + u_i$$

where α is the intercept, β_1, β_2, and β_3 are the slope coefficients for each explanatory variable, and u_i is the error term capturing unobserved factors affecting inequality.

1.4.2.2 Segmentation

We also segmented regions into categories to examine if the development–inequality relationship differed across different structural environments. Population density was split into "Rural", "Medium Density", and "High Density" using Eurostat-inspired breakpoints (0–300, 300–1500, 1500+). For workforce size, we used quantile-based grouping (Low, Medium, High). Unemployment rate was also split into Low-High segments, but here we set the breakpoints manually at 0%-5%, 5%-10% and 10+%. These subsets were used to run separate regressions for each group.

1.4.2.3 Panel Estimation

Finally, we estimated fixed-effects models using the full panel, applying plm() with region, country, year, and two-way effects. Before estimation, we converted the dataset into a pdata.frame and cleaned remaining NA/Inf values to avoid dropped regions. The panel models allowed us to control for all time-invariant regional characteristics, giving a "less noisy" picture of how regional GDP inequality changes inside regions over time.

1.4.2.4 Handling of NA-values and Heteroskedasticity

After joining the datasets onto eachother, including the new variables used for segmenting, we ended up with a few obvious gaps in our data and some NA/Inf rows. These were removed to keep the panel consistent, but a side effect of this is that we only have panel data from 2013 onwards.

We also ran Breusch–Pagan tests to check for heteroskedasticity across all regressions. Where alternative functional forms were required, we introduced quadratic and logarithmic transformations of the key variables and evaluated their residual behaviour the same way.

1.5 Empirical Findings - Harald

1.5.1 Cross- sectional Estimates

In assignment 2, we began with a cross-sectional analysis examining the relationship between regional inequality and short-term ecnonmic development, measured as the change in GDP per capita from 2016 to 2017. The dependent variable in our model was regional inequality, measured by the Gini coefficent, and the independent variable was the change in GDP per capita. This produced the following estimates:

	Estimate	Standard Error	t value	Pr(>|t|)
(Intercept)	0.134	0.016	8.514	0.0000 ***
chg_gdpc	-0.002	0.003	-0.787	0.4350

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05*

Residual standard error: 0.08068 on 50 degrees of freedom

Multiple R-squared: 0.01224, Adjusted R-squared: -0.007518

F-statistic: 0.6194 on 50 and 1 DF, p-value: 0.4350

The results shows an estimated intercept of ($\hat{\alpha} = 0.1336$), representing the average level of inequality in regions with zero GDP-per-capita growth. The coefficient for GDP-per-capita growth is negative ($\hat{\beta} = -0.0024$), which indicates a slight negative relationship between economic growth and inequality. However, the model's explanatory power is extremely low, with ($R^2 = 0.012$), meaning that changes in GDP-per-capita account for only 1.2% of the variation in inequality. Thus, although the coefficient suggests a negative relationship, the effect is statistically insignificant and not substantively meaningful. We can conclude that the short-term GDP-per-capita growth does has a small to none explanatory power for regional inequality in this analyse.

We then went on to collect three more variables to determine whether they serve as strong predictor of regional inequality then GDP-per-capita growth. These variables were labour force, unemployment rate, and population density. After cleaning and tiding the data, we estimated a multiple regression model to analyse their cross-sectional relationship with inequality:

$$\text{Gini}_{i,2017} = \alpha + \beta_1 \, \text{Workforce}_i + \beta_2 \, \text{PopDensity}_i + \beta_3 \, \text{Unemployment}_i + u_i$$

Where:

- $\text{Gini}_{i,2017}$ — Calculated Gini coefficient of region i in 2017 (regional GDP per capita inequality)
- $\Delta\text{Workforce}_i$ — Total labour force in region i, including both employed and unemployed individuals

- $\Delta \text{PopDensity}_i$ — Population density of region i, indicating how densely populated the region is
- $\Delta \text{Unemployment}_i$ — Unemployment rate in region i, representing presons withouth work who are available and is seeking work.
- α — intercept term (baseline inequality when the variables change is zero)
- $\beta_1, \beta_2, \beta_3$ — Coefficient showing how inequality changes with a one-percentage-point change in each variable
- u_i — error term capturing unobserved regional factors

Running the regression model gave us the following estimates:

	Estimate	Standard Error	t value	Pr(>|t|)	
(Intercept)	0.150168	0.034	4.465	0.0001	***
Workforce	0.000043	0.000	2.380	0.0216	*
Pop_km2	-0.000041	0.000	-2.763	0.0083	**
Unemp_prct	-0.011422	0.006	-1.919	0.0613	.

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05*

Residual standard error: 0.06949 on 45 degrees of freedom

Multiple R-squared: 0.2826, Adjusted R-squared: 0.2347

F-statistic: 5.908 on 45 and 3 DF, p-value: 0.0017

This model gives us a R^2 of 0,283 and an adjusted R^2 of 0,235 meaning that the these three variables together has a 28% explanatory power of the variation in regional inequality, which is higher then the change in GDP-per-capita. The *workforce* variable is statistically significant at the 5% level (p = 0.0216). It's coefficient, $\beta = 0.000043$, implies that regional inequality increases slightly as the labour force grows by one thousand workers. This may reflect that regions with larger or more active labour markets experience greater occupational and wage diversity, leading to somewhat higher inequality.

Population density is statistically significant at the 1% level (p = 0.0083) and has a negative coefficient ($\beta = -0.000041$). This suggests that more densely populated—typically more urbanized—regions tend to have lower inequality, possibly due to better access to education, more specialized employment opportunities, and a broader labour market.

The *unemployment* variable has a coefficient of $\beta = -0.0114$ and is marginally significant at the 10% level (p = 0.0613). The negative sign indicates that regions with higher unemployment rates appear to have lower inequality, although the weak significance prevents us from drawing a firm conclusion.

Overall assessment: Both labour force size and population density are stronger predictors of regional inequality than short-term GDP-per-capita growth. These structural variables appear to better capture regional economic differences than year-to-year changes in output.

1.5.2 Alternative functional forms and panel estimates

In this section, we explored alternative functional forms of the regression models using the variables from the second assignment workforce, population density, and unemployment rate to examine their effects on regional inequality, measured by the Gini coefficient. Specifically, we considered linear, logarithmic, and quadratic functional forms. While cubic specifications were also tested, they produced more estimates without providing meaningful improvements in explanatory power, so they were not pursued further.

We constructed separate regression tables for each functional form linear, logarithmic, and quadraticand compared their performance in terms of coefficient significance, explanatory power, and overall model fit. Below, we summarize the key findings and interpret the effects of each independent variable.

Workforce

Linear model:

- When all independent variables are zero, the regional inequality coefficient is estimated at $\beta = 0.1773$.
- Workforce has a coefficient of $\beta = 0.000036$ (p = 0.0665), which is marginally significant at the 10% level.
- Change in GDP per capita has a coefficient of $\beta = -0.002122$ (p = 0.4931), indicating it is not statistically significant.
- Overall, the model's explanatory power is low: $R^2 = 0.074$, adjusted $R^2 = 0.035$, and the F-test (F = 1.877, p = 0.163) suggests the model is not statistically significant.

Logarithmic model:

- Transforming change in GDP per capita using a log transformation improves model performance.

- When all independent variables are zero, the regional inequality coefficient is estimated at $\beta = 0.0.2292$.
- The coefficient for GDP per capita becomes $\beta = -0.059611$ and is highly significant (p = 0.0008), showing a negative relationship with the Gini coefficient.
- Workforce remains largely unchanged at $\beta = 0.000033$ (p = 0.0604, marginally significant).
- Explanatory power improves: $R^2 = 0.288$, adjusted $R^2 = 0.0251$, and the F-test is highly significant (F = 8.078, p = 0.0011).

Quadratic model:

- The quadratic specification produces less significant results than the logarithmic model.

- When all independent variables are zero, the regional inequality coefficient is estimated at $\beta = 0.1379$.
- GDP per capita: linear term $\beta = 0.004327$ (p = 0.4066), quadratic term $\beta = -0.000468$ (p = 0.1302), both insignificant.
- Workforce: $\beta = 0.000029$ (p = 0.1480), effect is weak.
- $R^2 = 0.1194$, adjusted $R^2 = 0.06199$, and the model is not statistically significant overall (p = 0.116).

Summary: The logarithmic model best captures the relationship between workforce, GDP growth, and inequality, mainly due to the strong significance of log-transformed GDP per capita.

2. Population Density

Linear model:

- When change in GDP per capita and population density are zero, the regional inequality coefficient is estimated at $\beta = 0.1506$.

- Population density has a coefficient of $\beta = -0.000037$ with a p-value of 0.0228, indicating statistical significance at the 5% level.
- Change in GDP per capita has a coefficient of $\beta = 0.000054$ (p = 0.6166), which is not significant.
- Model performance is limited: $R^2 = 0.111$, adjusted $R^2 = 0.072$, and the F-test is 7.36 with a p-value of 0.00668, suggesting marginal overall significance.

Logarithmic model:

- Applying a log transformation to change in GDP per capita improves model fit and significance.

- When change in GDP per capita and population density are zero, the regional inequality coefficient is estimated at $\beta = 0.229228$.
- GDP per capita becomes $\beta = -0.05793$, p = 0.0007, showing a strong negative relationship with inequality.
- Population density remains similar: $\beta = -0.000039$, p = 0.0047, still significant.

- Overall explanatory power increases substantially: $R^2 = 0.364$, adjusted $R^2 = 0.331$, and the F-test p-value is 0.0001, indicating a highly significant model.

Quadratic model:

- Including a quadratic term for GDP per capita allows for a non-linear relationship with inequality.

- When change in GDP per capita and population density are zero, the regional inequality coefficient is estimated at $\beta = 0.137873$.
- GDP per capita: linear term $\beta = 0.0077$ (p = 0.1136, not significant), quadratic term $\beta = -0.000682$ (p = 0.0198, significant), indicating a slight non-linear effect.
- Population density: $\beta = -0.000041$, p = 0.0091, remains significant.
- Model fit is moderate: $R^2 = 0.213$, adjusted $R^2 = 0.161$, with overall model p = 0.0123, showing statistical significance.

Summary: The logarithmic model provides the best explanatory power for population density, with both GDP growth and population density showing significant impacts on regional inequality. The quadratic model captures a slight non-linear effect but has lower overall explanatory power.

3. Unemployment Rate

Linear model:

- The regional inequality coefficient is $\beta = 0.1758$ when GDP per capita growth and unemployment are zero.

- Unemployment rate: $\beta = -0.001039$, p = 0.03, statistically significant, suggesting higher unemployment is associated with slightly lower inequality.
- Change in GDP per capita: $\beta = -0.0000248$, p = 0.936, not significant.
- Model fit is weak: $R^2 = 0.10$, adjusted $R^2 = 0.062$, and the F-test p-value is 0.083, indicating borderline overall significance.

Logarithmic model:

- Transforming GDP per capita using a log improves model performance.

- The regional inequality coefficient is $\beta = 0.233016$ when GDP per capita growth and unemployment are zero.

- GDP per capita: $\beta = -0.051$, p = 0.005, now highly significant with a negative effect on inequality.
- Unemployment rate: $\beta = -0.00707$, p = 0.1096, not significant, though the effect remains negative.
- Explanatory power increases: $R^2 = 0.27$, adjusted $R^2 = 0.2336$, F-test p = 0.0018, highly significant overall.

Quadratic model:

- Including a quadratic term for GDP per capita allows for a potential non-linear effect.
- The regional inequality coefficient is $\beta = 0.160671$ when GDP per capita growth and unemployment are zero.
- GDP per capita: linear term $\beta = 0.0062$ (p = 0.2114, not significant), quadratic term $\beta = -0.000488$ (p = 0.0102, borderline significant).
- Unemployment rate: $\beta = -0.009216$, p = 0.0516, borderline significant.
- Model fit is lower than the log model: $R^2 = 0.1517$, adjusted $R^2 = 0.09633$, with F-test p = 0.0539, borderline significant overall.

Summary: The logarithmic transformation of GDP per capita produces the strongest model for unemployment, improving both explanatory power and overall significance. Unemployment shows a weak but negative relationship with inequality, while quadratic effects of GDP per capita are borderline significant.

Overall Conclusion for the alternative functional forms:

Across all variables, logarithmic models generally provide the best explanatory power and statistical significance, particularly through the transformation of GDP per capita. Quadratic models capture some non-linear effects but often at the cost of reduced overall significance. Linear models are generally the weakest in explaining regional inequality. The key takeaway is that GDP per capita growth consistently shows a negative relationship with inequality when appropriately transformed, while workforce, population density, and unemployment rate exhibit more nuanced, context-dependent effects.

The panel regression results reveal a consistently positive association between change in GDP per capita and the dependent variable across all specifications. The magnitude of the effect is small (0.00017–0.00024), but precision improves substantially when controlling for regional heterogeneity using NUTS2 fixed effects, as reflected in markedly lower standard errors and RMSE. Models with country only or year only fixed effects explain very little variation ($R^2 = 0.005$–0.009), whereas NUTS2 and NUTS2 & Year fixed effects substantially improve model fit ($R^2 = 0.102$–0.109) and are preferred according to AIC and BIC. These findings highlight the importance of accounting for subnational regional characteristics, which capture significant variation in the outcome variable. Overall, the results suggest that regional-level factors are key drivers of the dependent variable, while the positive effect of change in GDP per capita remains robust across specifications.

1.6 Discussion - Begynner kvar for oss, samskriver etterpå

- Key insight:

The logarithmic models gives the best results for the alternative functions.

1.6.1 Key insights

1.6.2 Policy Implications

1.7 Limitations and Future Research - Harald

- lack of data
- What variable to use
- lack of information on what variables that may have a bigger impact on regional inequalities.
- Different countries, different ways of collecting data
- Gain more insight in each country and gain a more understanding on collecting data.

1.7.1 Reasearch limitations

1.8 Conclusion

1.8.1 Summary

1.8.2 Final Reflection

1.9 References

Appendix

County incomes and regional GDP 2015 - CSO - central statistics office. (2018). CSO. https://www.cso.ie/en/releasesandpublications/er/cirgdp/countyincomesandregionalgdp2015/

Information on data - national accounts - eurostat. (n.d.). https://ec.europa.eu/eurostat/web/national-accounts/information-data

Lessmann, C., & Seidel, A. (2017). Regional inequality, convergence, and its determinants – a view from outer space. *European Economic Review*, *92*, 110–132. https://doi.org/10.1016/j.euroecorev.2016.11.009