Explaining income inequality in Europe using longitudinal data

Andra-Ecaterina Boca, Junfeng Yan, Thien An Pham

December 16, 2019

Abstract

Income inequality has long been the domain of macroeconomics and standard labor economics. Literature postulates that GDP growth, the openness of an economy, and its labor market structure as potential drivers of country-level inequality. We test this using a Mixed Effects approach in 36 select European countries with data collected by the World Bank. We find that although the strongest predictors of income inequality are solely a country's GDP per capita and employment in industry, a model accounting for all the previously-mentioned factors can be argued to be preferable over a simpler one.

Introduction

Literature review

With more countries experiencing unprecedented rates of growth, wealth inequality has become a more and more pressing issue. Wealth disparities can appear due to market distortions such as corruption or non-productive economic activity like rent-seeking, which we describe dedicating resources above what is necessary for keeping an asset alive (for example, extension of copyright law). Income inequality on the other hand tends to be a more insidious topic of interest. Within income inequality, we distinguish between earnings from assets such as inherited wealth and wage compensation. Some findings suggest that wealth inequality is disproportionately influenced by market-driven factors, through the relative gains made by the top of the income distribution rather than by diverging levels of compensation – as such, jobs such as superstar athletes or Wall Street employees make significantly more money due to market forces that set their fees or performance at historically unprecedented prices (Kaplan 2013).

On the other hand, there still remains a large proportion of variability in earnings that is not necessarily explained by the rise of 'superstar' professionals. In the United States, the minimum wage has been declining in real terms for decades (Autor 2016) and the rising competition in the developing world and advanced technology have created an incentive for developed countries to outsource and automate lower-skilled labor. To that end, focusing on the incomes of the top 1% relative to the rest of the population clouds our understanding of emerging compensating differentials within that '99%' demographic.

The European Union and other related European countries represents an important area to study in the domain of wage inequality. It is first and foremost a diverse space in terms of population demographics, labor laws, and even culture around work. On the other hand, it mostly acts as a unitary space of population and capital movement that correlate any changes we might see in wage inequality across time and across countries. On average, wage inequality is expectedly lower in Europe relative to the US given better labor laws coverage and more widespread collective bargaining (DiPrete 2008). Using a microeconomic household-level lens, Rodríguez-Pose (2009) showed that European income per capita relationship with income inequality is positive while educational outcomes are not significantly correlated with inequality in the long-run. In terms of collective bargaining, Dell'Aringa (2005) found that multi-bargaining systems (more centralized worker bargaining) might have an effect on wage dispersion in European context.

At a macroeconomic level, Budría (2005) found that returns to tertiary education play a significant role in transitioning from the bottom quintiles of the income distribution to the upper quintiles in 8 select European countries. The rising inequality does not come from lower employment in lower-skilled occupations: in fact, a study for the United Kingdom shows employment has increased in top-paid jobs (professional and management) as well as in the lowest-paid jobs (such as cleaning and personal care) according to Goos (2007). To that end, it is worth empirically examining the structure of employment in a country as a contributing factor to income inequality.

Theoretical framework

In order to study income inequality at an aggregated level such as country data, we need to take two theoretical views on the problem. Firstly, we need to account for the variation in wage dispersion as suggested by this strand of labor economics and secondly, we need to be able to model inequalities arising from changing macroeconomic fundamentals, such as productivity and trade at a country level, which are associated with economic development.

Traditionally, what has been used to model wealth inequality in developing countries is the Kuznets curve formalized by Simon Kuznets in the 1950's. This formula posits that countries tend to become more unequal as they experience increased productivity and growth (that is, as they experience development), and then see a decrease in inequality. This is theorized to appear due to the level of opening of an economy (pictured by exports, imports and foreign direct investment) that indicate that firms in developed economies respond to import competition from low-wage countries by moving non-skill-intensive activities abroad. The Kuznets curve is expected to showcase a U-shape where inequality rises as the transition from agricultural to industrial jobs and from rural to urban happens, and decreases back to its original starting point as income per capita increases to the levels of what are considered developed countries.

From a labor-focused theoretical standpoint, wage inequality is to be expected due to different compensation for differing levels of skill. Since the 1990's, a significant strand of literature has tried to explain changing compensation levels by the skill-based technological change (SBTC) hypothesis – that is, the tendency of developed economies to shift demand towards more skilled and educated workers (for a comprehensive literature survey see Katz (1999)). This explanation has been in competition with previous supply-side considerations: factors such as collective bargaining, unionization, wage differentials across industries that were mainly shifted by workers (the labor supply), rather than firms and consumers.

Based on previous literature, it remains crucial for the study of income inequality to model the differences across countries in their populations' labor earning potential. This paper takes a Mixed Effects approach to analyse the drivers of inequality in 36 European countries. We use World Bank longitudinal data on select macroeconomic variables between the years 2005 and 2016. Our outcome variable of interest is the Gini country-level coefficient tracked by Eurostat for the European Statistics on Income and Living Condition (EU-SILC) survey which was then merged to relevant World Bank indicators.

Specifically, this analysis can contribute to existing literature by providing a unique country-by-country lens to income inequality. While it might not allow us to draw any broad conclusions about the dynamics of this economic indicator, the analysis using a Mixed Effects approach can give us an insight into across-country and across-time variation in the Gini inequality coefficient. Based on previous literature that hypothesizes that GDP growth, the openness of an economy, and its labor market structure as potential drivers of country-level inequality, we test the null hypothesis that these variables have no effect on the Gini inequality index.

We discuss data and statistical methodology in **Section II**, results in **Section III** and give a discussion about conclusions and limitations of our model in **Section IV**. We test our hypothesis using a Mixed Effects approach in 36 select European countries with data collected by the World Bank. We find that although the strongest predictor of income inequality is employment in industry, a model accounting for all the previously-mentioned factors is statistically preferred over a simpler one.

Methods

Data description

The macroeconomic indicator data for the analysis comes from the World Bank DataBank publicly available data. It tracks macroeconomic variables for 36 selected countries across 11 years (2005-2016). The countries we are interested in are both part of the European Union (EU) and outside at different points in time (for example, Bulgaria and Romania accede to the EU in 2007). These countries are: Albania, Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom.

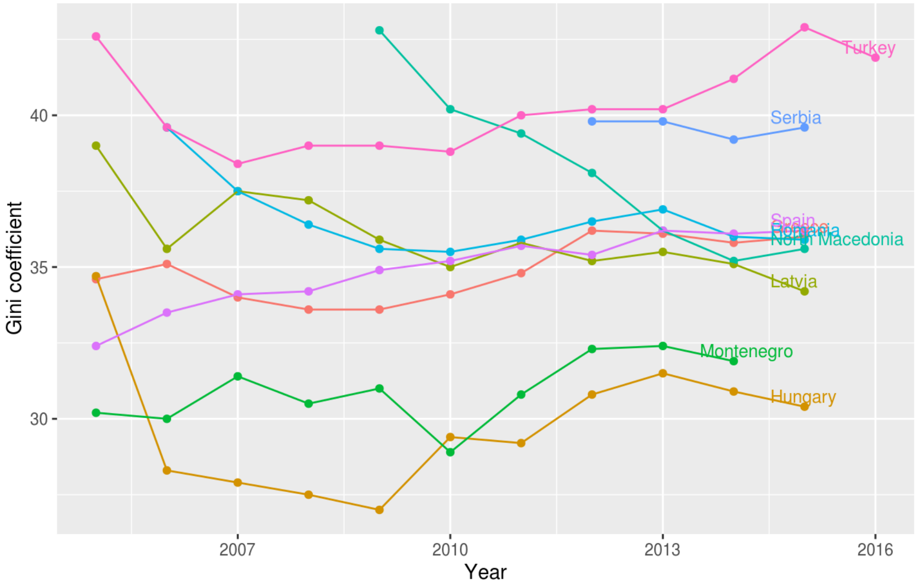
The outcome variable is the **Gini inequality coefficient** tracked by Eurostat. The Gini coefficient is a metric of statistical dispersion for wealth. In percentage terms, a Gini coefficient of 0 represents a perfectly equal country showcasing perfect homogeneity in income across a distribution, while a Gini coefficient of 100 represents complete inequality. Eurostat considers income everything encompassing wages, self-employment earnings, private income from investment and property, transfers between households (such as remittances or gifts) and social transfers (pensions, benefits and others). For a more in-depth description of all the other variables, please consult the Appendix.

Descriptive statistics

To be able to better understand the data, we include some visual descriptions of the development across time of select variables and their relationships with our outcome variable (the Gini coefficient).

Figure 1: Change in inequality throughout time

(selected countries)



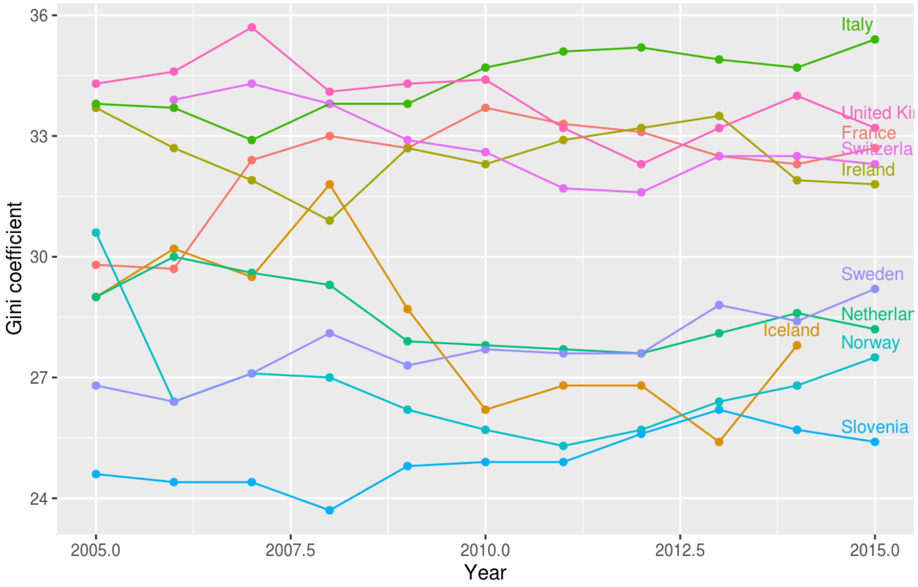


Figure 2: Change in inequality throughout time
(selected countries)

In **Figure 1** and **Figure 2**, we plot the change in inequality measured by our outcome variable of interest, the Gini coefficient. It is notable that countries that are in transition from developing to developed (primarily Eastern European countries) tend to showcase larger inequality and variability, as opposed to more developed Western European countries in **Figure 2**. This is in line with the Kuznets economic hypothesis outlined in the **Section 1**. Typically, we see increased inequality for countries in transition.

In terms of the explanatory variables we consider based on the theory, we note that most countries have similar trajectories in GDP per capita growth, unemployment or employment in Industry for example. Most of the other macroeconomic health variables in the dataset follow the same parallel pattern of stable growth or decrease. An example for GDP per capita in **Figure 3** showcases such trajectory.

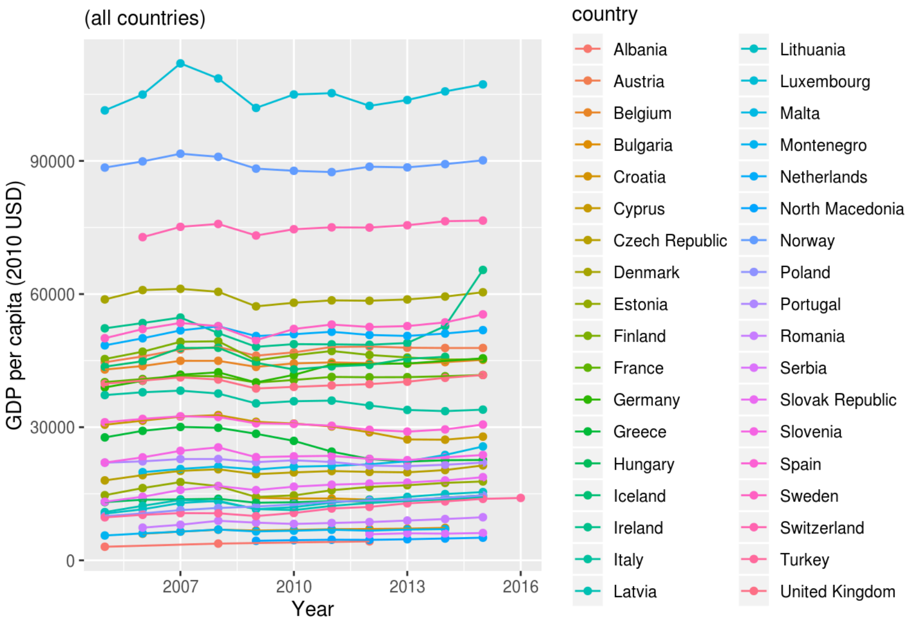


Figure 3: Change in GDP per capita throughout time
(all countries)

Two particular variables require a more careful analysis. In **Figure 5** and **Figure 6**, we can see the trajectories of foreign direct investment (FDI) and trade across time might be skewed by outliers if specified as fixed in our model. In the case of both FDI and trade, Luxembourg is quite starkly different to all the other countries. Luxembourg is a small country with a small production level on its own, and as such the large amount of trade and FDI to the country represent outliers as a ratio of a small GDP. Malta and Cyprus are similar cases in terms of FDI and trade respectively. For our analysis, this means the relatively larger starting points of these countries show a correlation to their growth trajectory.

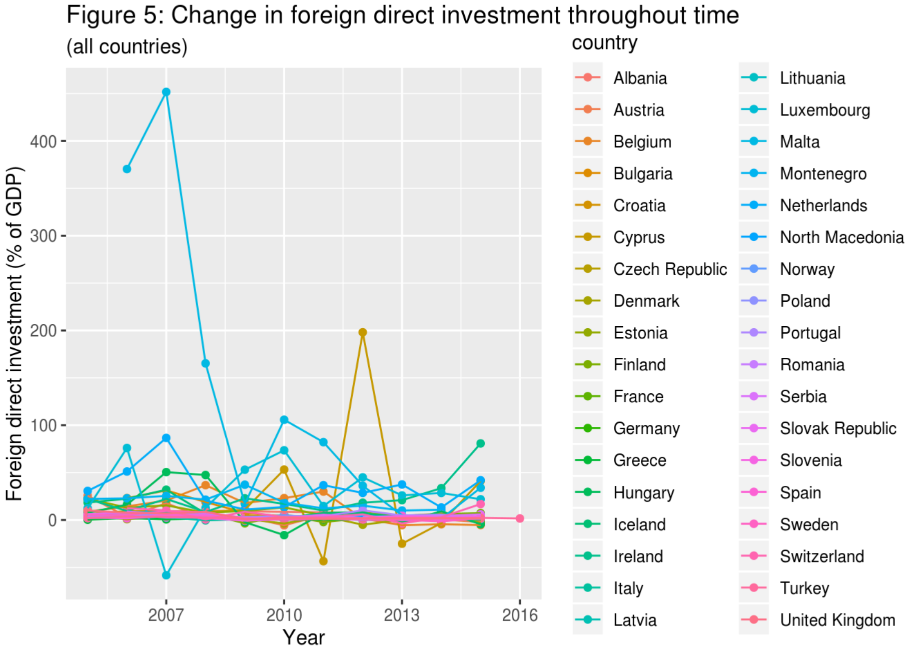


Figure 5: Change in foreign direct investment throughout time
(all countries)
Foreign direct investment (% of GDP)
400
300
200
100
0
2007
2010
2013
2016
Year
country
Albania
Austria
Belgium
Bulgaria
Croatia
Cyprus
Czech Republic
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Iceland
Ireland
Italy
Latvia
Lithuania
Luxembourg
Malta
Montenegro
Netherlands
North Macedonia
Norway
Poland
Portugal
Romania
Serbia
Slovak Republic
Slovenia
Spain
Sweden
Switzerland
Turkey
United Kingdom

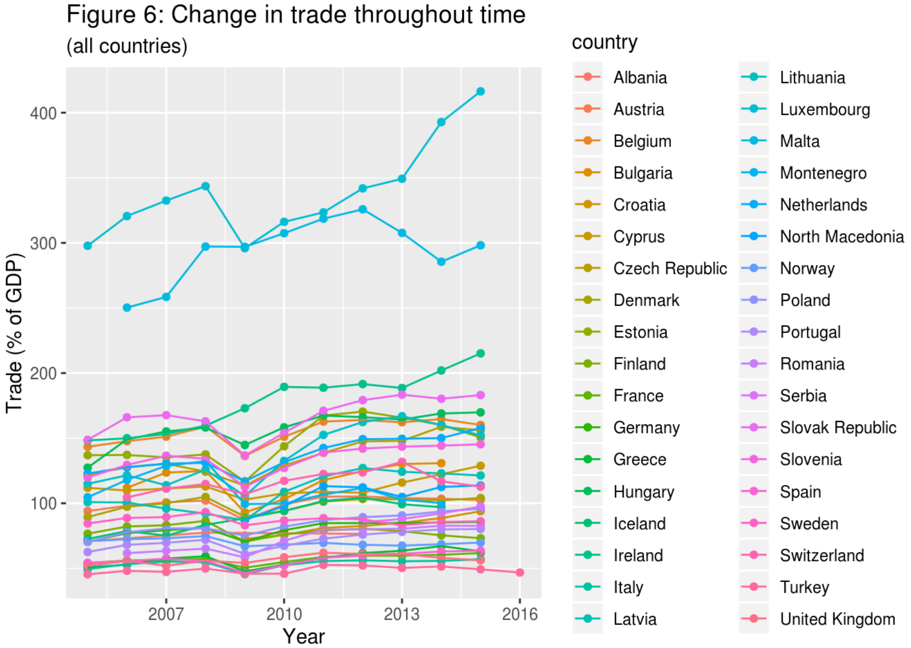


Figure 6: Change in trade throughout time
(all countries)
Trade (% of GDP)
400
300
200
100
2007
2010
2013
2016
Year
country
Albania
Austria
Belgium
Bulgaria
Croatia
Cyprus
Czech Republic
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Iceland
Ireland
Italy
Latvia
Lithuania
Luxembourg
Malta
Montenegro
Netherlands
North Macedonia
Norway
Poland
Portugal
Romania
Serbia
Slovak Republic
Slovenia
Spain
Sweden
Switzerland
Turkey
United Kingdom

Figure 7: Inequality explained by GDP per capita

(all countries)

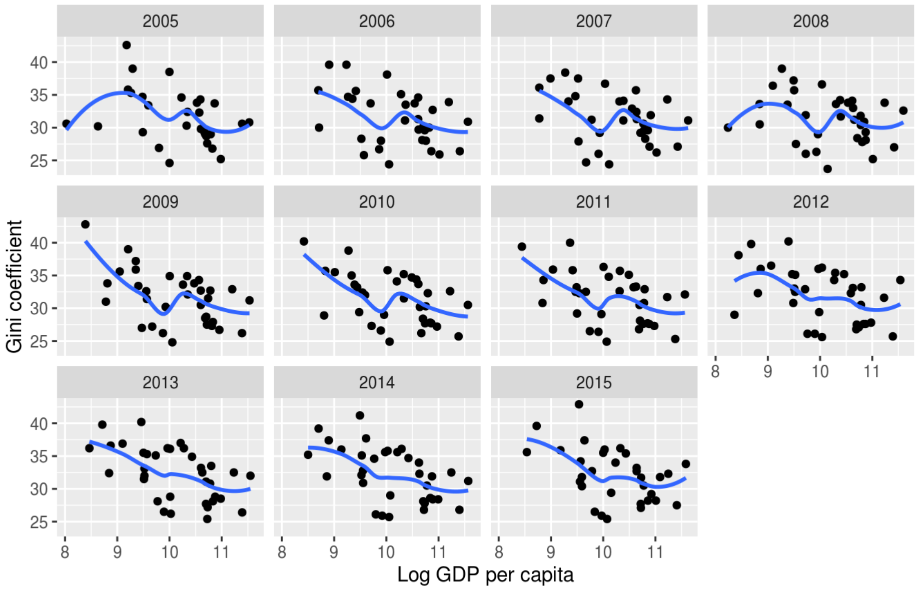
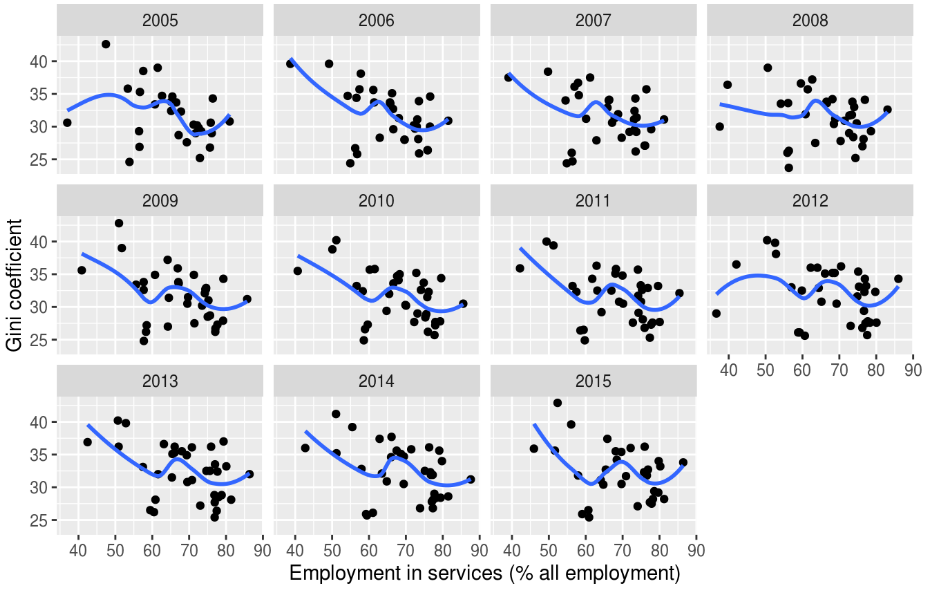


Figure 7 showcases the relationship between different GDP per capita and inequality value pairs for each country. We note a distinctly decreasing pattern. As GDP increases, the Gini coefficient decreases or, in other words, inequality decreases. We expect to see poorer countries have a larger inequality index.

Figure 8: Inequality explained by employment in services

(all countries)



Similarly, we expect that countries with a higher level of population involved in economic activities away from agriculture and industry will perform better in terms of productivity. We note the same decreasing pattern between engagement in services and inequality as in **Figure 8**.

In selecting what variables would best represent a country's employment and labor structure, we find that net investment as a percentage of GDP, tax revenue, education expenditure as a percentage of total government expenditure do not vary significantly with inequality. These plots are available in **Section 2** of the Appendix. We then assume they are artifacts of each country and also stay quite constant over time. In terms of demographics, we find that net migration has a weak relationship with inequality, but we choose not to focus on this variable in our final model.

Statistical methodology

As per our literature review and theoretical framework, GDP growth, the openness of an economy, and its labor market structure are the three most frequently identified drivers of wage inequality in an economy. As such, we use the log of GDP per capita to represent economic growth, trade value and FDI to represent openness of economy , percentage of employment in industry, percentage of employment in service and in industry, and percentage of population in urban areas to account for different demographic structure among the observed countries. Additionally, we created a dummy variable for year 2008 to account for the great recession. Our baseline model of wage inequality,represented using the gini coefficient, is written as such:

$$gini = \beta_0 + \beta_1 * lggdppercapita + \beta_2 * trade + \beta_3 FDI + \beta_4 industryemployment + \beta_5 employmentservice + \beta_6 urbanpopulation + \beta_7 *$$

In other words, this is a naive OLS model. However,given the longitudinal nature of our data,which has regular repeated observations of different countries, this method will clearly give biased results. We then turn to marginal models with generalized estimating equations (GEE) (Liang and Zeger 1986) and mixed-effect models (Laird, Ware, and others 1982) to tackle this regression questions. In short, GEE estimate the sample population average coefficient for a given set of parameters and produces robust standard errors for each parameter, due to the fact that it specifies a working correlation structure for each observed individual. On the other hand, mixed effect model allows for heterogeneity within the estimated intercept as well as some regression coefficients for the sample population (random effect),in addition to estimating the population average for a given set of parameters (fixed effect).

For our specific research question, the mixed effect model is preferred. As informed by **Figure 1**, the observed countries clearly started at different level of Gini coefficient at the start of coefficient period. Moreover, it is apparent in **Figure 5** that different countries exhibit different trade volume and FDI trajectory within our observation period. It is therefore necessary to introduce country-specific random intercept, denoted as b_i, as well as country-variable-specific random coefficients,denoted as β_{ij} to account inter-country differences in terms of starting level and variable coefficients. Both random effects are assumed to follow a probability distribution, such as $N\left(0, \sigma_i\right)$. In addition, for a sufficiently complex phenonemon such as inter-coutnry income inequality, it is to some degree inevitable that there are unobserved factors or unobserved differences between country that are significant to income inequality. Country-level random intercept and coefficient can to some degree accounted for this potential omitted variable bias. Our tentative models are therefore:

$$model1: gini = \beta_0 + \beta_1 * lggdppercapita + \beta_2 * trade + \beta_3 FDI + \beta_4 industryemployment +$$

$$+\beta_5 employmentservice + \beta_6 urbanpopulation + \beta_7 * year2008 + b_i$$

and

$$model2: gini = \beta_0 + \beta_1 * lggdppercapita + \beta_2 * trade + \beta_3 FDI + \beta_4 industryemployment + \beta_5 employmentservice +$$

$$+\beta_6 urbanpopulation + \beta_7 * year2008 + b_i + b_{i,trade}$$

After using maximum likelihood to estimate our model, we however find that the random effect and random intercept for model 2 has a correlation of 0.92, meaning they are perfectly negatively correlated with each other. Although we are not sure of the exact cause behind this estimates, it is discretionary to avoid estimates that are close to the boundary of possible values and we therefore discarded model 2. Out of discretion, we also fit the model with random slopes for every other variable. It appears that every random effect that we estimated are met with similar problems.

With only model 1 left, we now concern ourselves with this problem: is the full specification necessary or is a simpler version of the model preferable. To shrink down the number of specified variables, we first conduct a single variable hypothesis test with null hypothesis $H0: \beta = 0$. We can calculate a z-statistic, $z = \frac{\hat{SE}}{\hat{\beta}}$. This z-statistic is simply given by our model output. We assume that significant variables will have a z-score that is larger than 1.96, which translates to a p-value smaller than 0.05 if we were to assume the sample distribution to be normal. We however, also retain trade volume since its theoretical importance and its z-statistic are close to our threshold value. With this test, the strongest three predictors are the log of gdp per capital, employment in industry and trade volume.The simpler model is therefore:

$$model3: gini = \beta_0 + \beta_1 * lggdppercapita + \beta_2 trade + \beta_4 industryemployment + b_i$$

Running the model again with the simpler specification also produces lower AIC and BIC than the full specification. It is important to note that now trade has a significant z-statistic.

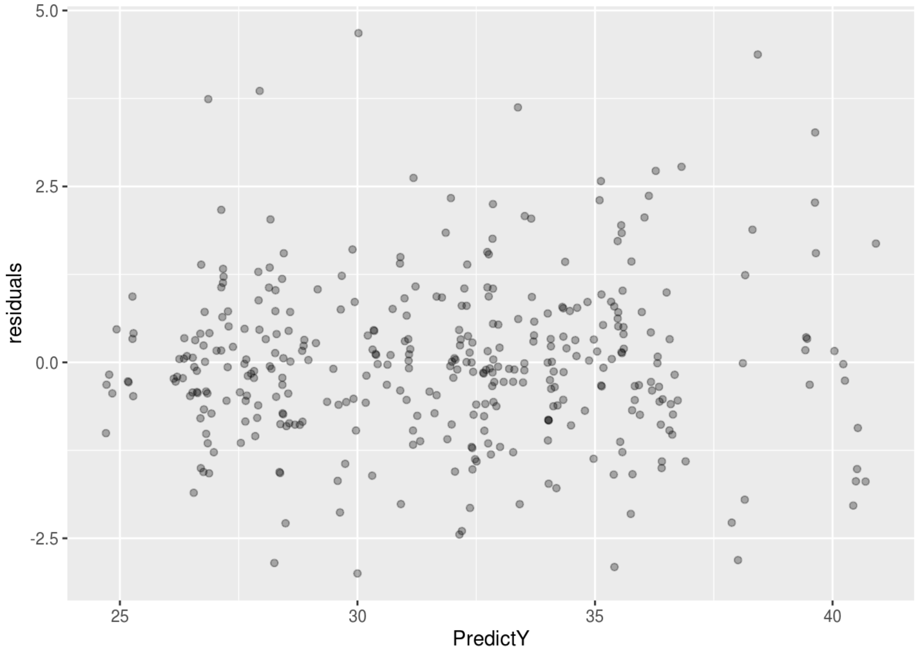
However, we did not simply conclude that the simpler specification is preferable over the full specification. It is important to stress again the focal point of this paper: we are attempting to explain main driver of income inequality instead of building a predictive model of income inequality. As such, it is undesirable to discard variables that theory suggests to be significant simply due to insufficiently large z-statistics. To test a more involved hypothesis, $H0: \mathbf{L}\beta = 0$, that whether multiple slopes are 0, we calculate a W-statistic as follows:

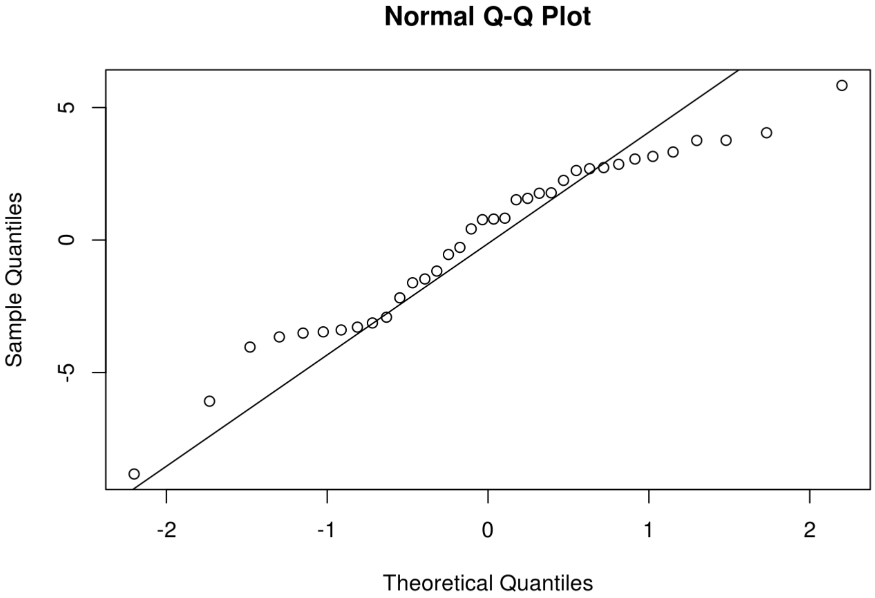
$$\mathbf{W}^2 = (\mathbf{L}\hat{\beta})^T (\mathbf{L}\hat{Cov}(\hat{\beta})\mathbf{L}^\mathbf{T})^{-1} (\mathbf{L}\hat{\beta})$$

We then used this test to see whether the slopes of variables that did not pass the individual hypothesis test, namely trade volume, foreign direct investment, service employment, urban population, and year 2008 are in fact 0. With this test, we obtain a p-value of 0.01754489, which means that we have significant evidence to reject the null hypothesis that these slopes are 0. With our theoretical consideration and this test result, we conclude our final model to be the first model that includes all the variables.

Model Diagnostics

Since mixed-effect model assume errors, denoted ϵ_i, as well as the random intercept to be normally distributed and to have a mean 0, we lastly graph a residual plots and QQ plots to see if this assumption is being violated.





Residuals plots shows that ϵ_1 have mostly no pattern and is balanced 0. The QQ plot appears a bit more concerning as the random intercepts appear to stray from the normal line at both ends.

Results

As discussed above, we found that the best model for our data is the random intercept and slope one where we determine wage inequality by measuring log of GDP per capita, a country's trade, a country's foreign direct investment, unit percentage of employment in a country's industries, unit percentage of employment in a country's industries, unit percentage of employment in a country's services, percentage of urban population in a country, and whether or not a country experienced an economic crisis in 2008. We can see that the variance is pretty high of 10.645, suggesting wage inequality varies a lot by countries.

	Gini coefficient		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	59.90	48.53 – 71.26	**<0.001**
Log GDP per capita	-1.60	-3.06 – -0.14	**0.031**
Percent Urban Population	-2.38	-11.49 – 6.72	0.608
Foreign Direct Investment	-0.00	-0.01 – 0.00	0.612
Employment Industry	-0.20	-0.34 – -0.05	**0.007**

Trade	-0.01	-0.02 – 0.00	0.060
Employment Services	-0.06	-0.18 – 0.06	0.335
Year 2008	0.08	-0.37 – 0.53	0.731
Random Effects			
σ^2	1.43		
$\tau_{00\ id}$	10.64		
ICC	0.88		
N_{id}	36		
Observations	365		
Marginal R^2 / Conditional R^2	0.212 / 0.906		

By looking at the p values, log of GDP per capita of a country and unit percentage of employment in a country’s industries are two statistically significant predictors that have p values of less than 0.05 - 0.031 and 0.07 respectively. The fixed effects also tell us that these are the only two predictors with more than two standard deviations away.

- **logGDPpercapita**: if a country’s log of GDP per capita increases by 1 unit, their wage inequality is expected to decrease by 1.60%.
- **trade**: if a country trades 1% more, their wage inequality is expected to decrease by 0.01%.
- **fdi**: if a country invests in another country 1% more, their wage inequality is expected to remain unchanged.
- **emplInd**: if a country has 1% higher employment in their industries, their wage inequality is expected to decrease by 0.20%.
- **emplServ**: if a country has 1% higher employment in their services, their wage inequality is expected to decrease by 0.06%.
- **percUrbanPop**: if a country’s population in urban areas increases by 1%, their wage inequality is expected to decrease by 2.38%.
- **year2008**: if the year of the observation is 2008 (`year2008` = 1), the country’s wage inequality is expected to increase by 0.08%.

The output also tells us the correlation of fixed effects, or the expected correlation of the regression coefficients. For example, if the coefficient for the log of GDP per capita increases by 1 unit, it is likely that the coefficient for trade decreases by 0.076% and vice versa. Or if the coefficient for trade increases by 1%, it is likely that the coefficient for percentage of employment in a country’s industries increases by 0.12%. This may seem to represent multicollinearity but not necessarily. It tells you that should you test the model again and the coefficient changes, other correlated coefficients may change as well in a positive or negative direction.

Conclusions

Using World Bank and Eurostat data between 2005 and 2011, we test whether a country’s GDP growth, trade openness, and its labor market structure have a significant effect on its economic inequality using a Mixed Effects approach. We find that although the strongest predictors of income inequality are a country’s GDP per capita and employment in industry, a model accounting for all the previously-mentioned factors can be argued to be preferable over a simpler one in referral to economic theory.

Limitations

First, it is important to keep in mind the limited scope of our project. We use a theoretical model that seeks to prove a very widely-reaching hypothesis. The Kuznets curve variables are hypothesized for the life cycle of a country from industrialization to its actual observed growth which can last decades – we are, on the other hand, looking at a mere 11 years of country-level data. While we do see some significant effects across countries, the development path of a single country in this limited time frame might not tell us so much. Similarly, there remains a lot of omitted variable bias that we should correct for, especially in terms of demographics. Controlling for an aging population, for example, might result in a different statistical significance for our variables of interest.

Secondly, as shown in the descriptive statistics in Section 3, there are outlier countries such as Luxembourg and Malta that might drive us to different analysis results when we add trade and FDI into our model. More data work with and without the outliers might lead to more robust results.

Thirdly, our model diagnostics, specifically the QQ plot, show that the random intercept at smaller or larger values may not follow a normal distribution. Although this might undermine the validity of our model, with more rigorous statistical method or tools we can account for such limitation.

Acknowledgements

We want to thank everyone that made this paper possible: first and foremost, our STAT 494 Correlated Data professor Brianna Heggeseth for her patient guidance and help every time we were stuck, as well as the class preceptors, Yuren (Rock) Pang, Raven McKnight and Kieran Liming, for their continuous support.

Appendix

Section 1: Variable description

We track the following variables in our compiled World Bank dataset:

- Population: population of country in absolute terms at given year Urban population: population of a country living in a city in absolute terms at given year
- Percent urban population: ratio of urban population to total population (%)
- Imports: all the goods imported from the rest of the world to specific country reported as percentage of GDP (%)
- Exports: all the goods exported to the rest of the world to specific country reported as percentage of GDP (%)
- Trade: the sum of exports and imports of goods and services as a share of gross domestic product; reported as a percentage of GDP (%)
- Foreign direct investment: net inflows of investment to a country, reported as a percentage of GDP (%)
- GDP (Gross Domestic Product): sum of value added by all the producers, sometimes referred to as productivity or growth, reported in constant 2010 USD
- GDP per capita (Gross Domestic Product per capita): gross domestic product divided by midyear population, reported in constant 2010 USD
- Net investment in non-financial assets: investment in fixed assets, inventories, valuables, and non-produced assets, reported as a percentage of GDP (%)
- Education expenditure: government expenditure on education, reported as a percentage of GDP (%)
- Unemployment: total population out of labor force that is currently unemployed, reported as a percentage of labor force (%)
- Tax revenue: compulsory transfers to the central government for public expenditure purposes, reported as a percentage of GDP (%)
- Employment in agriculture/industry/services: should sum up to 100%, reported as a percentage of total employment (%)
- Net migration: the number of immigrants minus the number of emigrants, in absolute terms.
- Wage inequality: the 80/20 wage inequality ratio calculated by Eurostat.

Section 2: Visual relationships of variables and Gini coefficient

Figure A1: Inequality explained by trade

(all countries excluding Luxembourg and Malta)

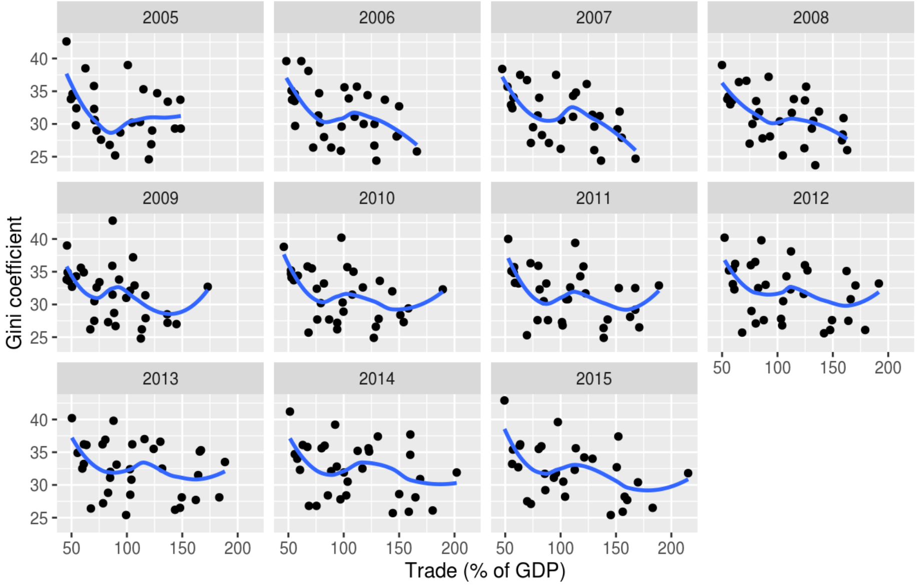


Figure A2: Inequality explained by net investment

(all countries

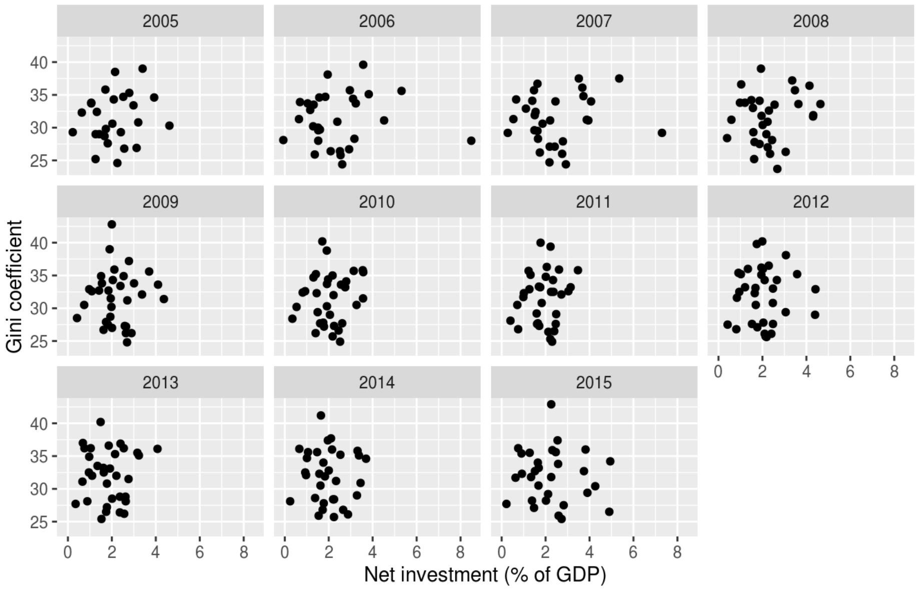


Figure A3: Inequality explained by education expenditure

(all countries)

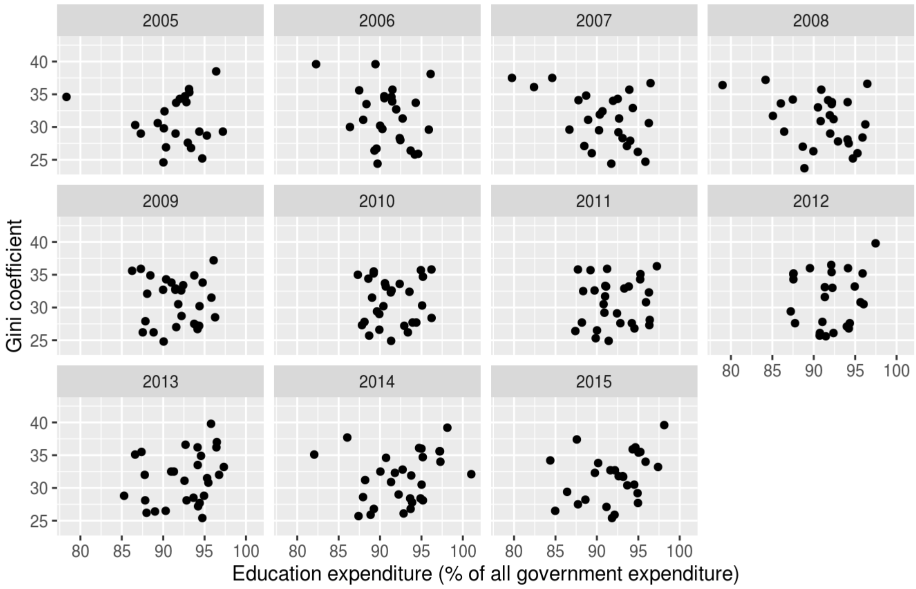


Figure A4: Inequality explained by tax revenue
(all countries)

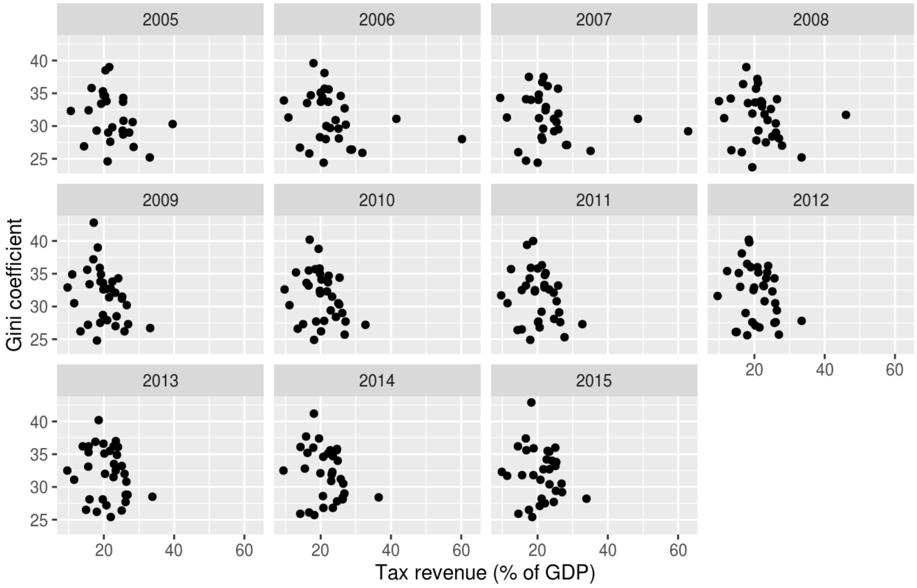
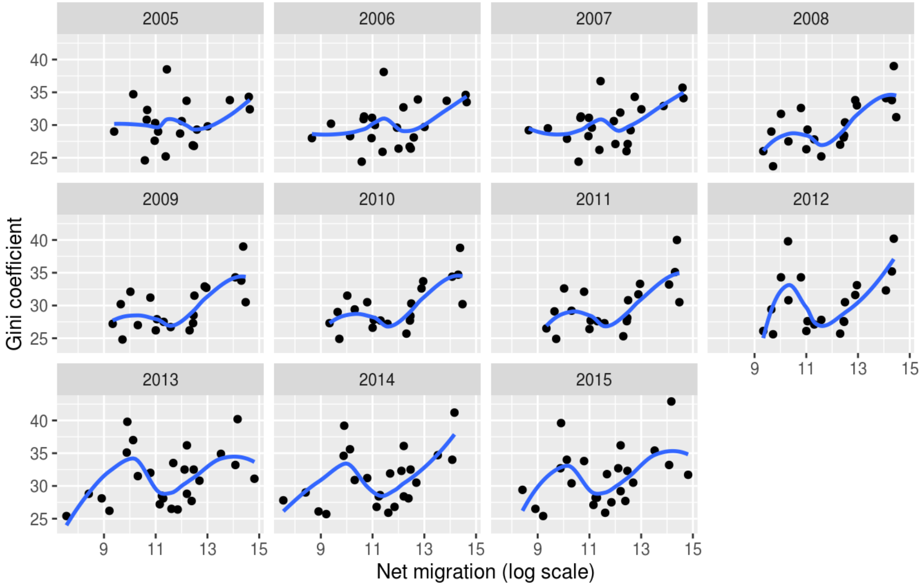


Figure A5: Inequality explained by net migration
(all countries)



References

Autor, Manning, D. 2016. “The Contribution of the Minimum Wage to U.s. Wage Inequality over Three Decades: A Reassessment.” *American Economic Journal: Applied Economics* 8 (1): 58–99.

Budría, & Pereira, S. 2005. “Educational Qualifications and Wage Inequality: Evidence for Europe.”

Dell’Aringa, & Pagani, C. 2005. “Regional Wage Differentials and Collective Bargaining in Italy.” *Rivista Internazionale Di Scienze Sociali*, 267–87.

DiPrete, T. A. 2008. “Labor Markets, Inequality, and Change: A European Perspective.” *Work and Occupations* 32 (2). National Bureau of Economic Research: 119–39.

Goos, A., M. & Manning. 2007. “Lousy and Lovely Jobs: The Rising Polarization of Work in Britain.” *Review of Economics and Statistics* 89 (1): 118–33.

Kaplan, & Rauh, S. N. 2013. “It’s the Market: The Broad-Based Rise in the Return to Top Talent.” *Journal of Economic Perspectives* 27 (3): 35–56.

Katz, L. F. 1999. “Changes in the Wage Structure and Earnings Inequality.” *Handbook of Labor Economics* 3: 1463–1555.

Laird, Nan M, James H Ware, and others. 1982. “Random-Effects Models for Longitudinal Data.” *Biometrics* 38 (4): 963–74.

Liang, Kung-Yee, and Scott L Zeger. 1986. “Longitudinal Data Analysis Using Generalized Linear Models.” *Biometrika* 73 (1). Oxford University Press: 13–22.

Rodríguez-Pose, & Tselios, A. 2009. “Education and Income Inequality in the Regions of the European Union.” *Journal of Regional Science* 49 (3): 411–37.