LIS
Working Paper Series

No. 573

Income Inequality and Redistribution in Europe

Simona Tóthová

November 2011

CROSS-NATIONAL
DATA CENTER
in Luxembourg

Luxembourg Income Study (LIS), asbl

CHARLES UNIVERSITY IN PRAGUE

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

BACHELOR THESIS

2011 **Simona Tóthová**

CHARLES UNIVERSITY IN PRAGUE

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

BACHELOR THESIS

Income inequality and redistribution in Europe

Author: Simona Tóthová

Consultant: PhDr. Martin Gregor PhD.

Academic Year: 2010/2011

Bibliographic Evidence Card

Tóthová, Simona, *Income inequality and redistribution in Europe*, Prague: Charles University, Faculty of Social Sciences, Institute of Economic Studies, 2011, pages 53, Supervisor: PhDr. Martin Gregor PhD.

Abstract

This thesis is an empirical investigation of the relationship between income inequality and redistribution and their impact on the poorest segment of population. It is examined by two models in which we use transformed Gini coefficient as a measure of income inequality and sharegain (difference between disposable and factor income share of given decile) as a measure of redistribution. Relationships are tested on a sample consisting of 24 European countries in the period 1990 -2005. This sample is deeply analyzed with special emphasis put on post-communist countries. First model points out that redistribution can be a luxury social good. Second model shows the inverse process and supports the hypothesis of positive relationship between redistribution and overall income equality.

Keywords

income inequality, redistribution, factor income, disposable income, Gini coefficient, sharegain

Abstrakt

Táto práca empirický skúma vzťah medzi nerovnosťami príjmov a redistribúciou a ich vplyvom na najchudobnejšie vrstvy obyvateľstva. Daný vzťah je skúmaný pomocou dvoch modelov. V oboch prípadoch je nerovnosť príjmov meraná Gini koeficientom a redistribúcia pomocou sharegain (rozdiel medzi podielom čistých a hrubých príjmov v danom decile) . Nastolené závislosti sú testované na vzorke 24 krajín obdobi rokov 1995-2005 s hlbším zameraním na postkomunistické krajiny. Prvý z týchto modelov naznačuje, že redistribúcia môže byť luxusným sociálnym statkom, druhý model sa zameriava na inverzný proces a podporuje hypotézu pozitívneho vzťahu medzi redistribúciou a celkovou rovnosťou v príjmoch.

Kľúčové slová

nerovnosť príjmov, redistribúcia, hrubý príjem, čistý príjem, Gini koeficient, sharegain

Rozsah práce

65 671 znakov

Prehlásenie

Prehlasujem, že som predkladanú bakalársku prácu spracovala samostatne a použila len uvedené pramene a literatúru.

Prehlasujem, že práca nebola použitá na získanie iného titulu.

Súhlasím s tým, aby práca bola sprístupnená pre študijné a výskumné účely.

Praha, 31. júla 2011 ____________________

Simona Tóthová

Acknowledgments

I would like to express my gratitude to my thesis advisor PhDr. Martin Gregor PhD. for the time and effort, he devoted to constant encouragement and essential guidance in the process of writing this work.

Moreover, I would like to thank to my parents and the closest friends for their unlimited support, love and patience they were dealing me with in hard times of making this thesis and during my studies.

UNIVERSITAS CAROLINA PRAGENSIS
založena 1348

Univerzita Karlova v Praze
Fakulta sociálních věd
Institut ekonomických studií

Opletalova 26
110 00 Praha 1
TEL: 222 112 330,305
TEL/FAX: 222 112 304
E-mail: ies@mbox.fsv.cuni.cz
http://ies.fsv.cuni.cz

Akademický rok 2009/2010

TEZE BAKALÁŘSKÉ PRÁCE

Student:	Simona Tóthová
Obor:	Ekonomické teorie
Konzultant:	PhDr. Martin Gregor PhD

Garant studijního programu Vám dle zákona č. 111/1998 Sb. o vysokých školách a Studijního a zkušebního řádu UK v Praze určuje následující bakalářskou práci

Předpokládaný název BP:

Income inequality and redistribution in post-communistic EU countries.

Charakteristika tématu, současný stav poznání, případné zvláštní metody zpracování tématu:

Struktura BP:

<u>Cieľ práce:</u>

Cieľom tejto bakalárskej práce je zhodnotiť prerozdelenia príjmov v spoločnostiach desiatich postkomunistických štátov Európskej Únie. Tieto stále silno sa rozvíjajúce štáty EU majú veľmi špecifickú históriu, čo je hlavný dôvod môjho výberu. Inšpirovala ma práca Branka Milanoviča z roku 2000 a následné kritiky na túto prácu. Milanovic previedol empirický test, kde použil regresnú analýzu na dátach z 24 štátov sveta prevažne krajín OECD. Skúmal hypotézu mediánového voliča, nerovnosť a prerozdelenie príjmov. Ja by som chcela použiť jeho metodiku a postupy, no aplikovať to práve iba na 10 postkomunistických štátov EU. Tento test vyzerá veľmi stručne nasledovne. Jednotlivci sú označený podľa tzv. trhového príjmu od najchudobnejšieho po najbohatšieho a rozdelený do skupín podľa decilov. Potom sledujeme ako sa u každého jednotlivca podiel decilu trhového príjmu mení pri prechode ku podielu disponibilného príjmu (príjem po vládnych transferoch a daniach) a zistíme ako vyzerá ich rozdiel, kedy je kladný, kedy je záporný a mojim cieľom je zistiť prečo a ktoré faktory na jeho zmenu pôsobia. Ďalším mojím cieľom je pokúsiť sa v tejto analýze nájsť odpovede na nasledujúce otázky: Ktorá skupina najviac benefituje z tejto redistribúcie? Ktorej skupine vyhovuje rovnosť v redistribúcii príjmov najviac? V druhej časti by som sa chcela zamerať na otázku mediánového voliča a pokúsiť sa zistiť ako jeho pozícia vplýva na redistribúciu príjmov a kto vlastne tento mediánový volič je.

Osnova:

1 . Úvod

2 . Vymedzenie základných pojmov

  A. nerovnosť a redistribúcia

  B. teória mediánového voliča

3 . Empirické testovanie hypotéz

  A. Opis databázy a jej špecifiká

  B. Testovanie hypotéz

4 . Analýza výsledkov testu

  A. Diskusia o rovnosti príjmov

  B. Hľadanie mediánového voliča

5. Záver

Seznam základních pramenů a odborné literatury:

Milanovic, 2000 B. Milanovic, The median voter hypothesis, income inequality and income redistribution: an empirical test with the required data, European Journal of Political Economy 16 (2000), pp. 367–410.

Milanovic, 2010 B. Milanovic, Four critiques of the redistribution hypothesis: An assesment, European Journal of Political Economy 26 (2010), pp. 147-154.

Meltzer and Richard, 1981 A.H. Meltzer and S.F. Richard, A rational theory of the size of government, Journal of Political Economy 89 (1981), pp. 914–927. Full Text via CrossRef

Luiz de Mello and Erwin R. Tiongson, Income Inequality and Redistributive Government Spending, *Public Finance Review* 2006; 34; 282

Scervini, F., 2009. The empirics of the median voter : democracy, redistribution and the role of the middle class. Collegio Carlo Alberto Working Paper. Available at http://fileserver.carloalberto.org/id2009/scervini.pdf.

Datum zadání:	
Termín odevzdání:	

Podpisy konzultanta a studenta:

V Praze dne

Contents

List of figures

List of tables

Introduction

Traditionally, in each country there are three main groups of people sorted by their factor incomes - rich, middle and poor. But in each country, naturally, their share of income differs. This constitutes a big opportunity for economists to study these differences (income inequality) and their changes after taxes and transfers (income redistribution).

The aim of this thesis is to look closely at the last group – the group with the lowest incomes and to study its gains and losses in sense of income redistribution. Income inequality, growth and redistribution are concepts, which have very often been used in recent economic studies. One of pioneers of this approach was Simon Kuznets [1], who published the study "*Economic growth and income inequality*" yet in 1955 and his hypotheses about U-shape trend of inequality over time, in a developing country is used hitherto. To name just a few other interesting authors, who will be mentioned in this paper later is Perotti [2], [3], Benabou [4] or Mello and Tiongson [5]. We decided to expand their work and enrich this topic by testing new hypothesis and results.

Specifically, our thesis is based on Milanovic's paper [6]: "*The median-voter hypothesis, income inequality, and income redistribution: an empirical test with the required data.*" We partly use his method and variables with the aim to empirically investigate relationships between income inequality and redistribution.

In this thesis we would like to compose and later estimate two models. The first will be adjusted Milanovic's [6] model and the next we call "New income inequality model". This model is built on inverse relationships to Milanovic's. Also, we decided to compare two European regions separately - Western Europe, represented by highly developed democracies and Central and Eastern Europe (CEE).

The first chapter introduces theoretical background of the problems. In particular, it contains a brief survey of the literature, introduction to inequality study and different ways of inequality measurement, with a focus on Gini coefficient. In the second part of this chapter, variables are constructed and we will present both our models.

Chapter two clarifies datasets used, pointing out Luxembourg Income Study as the only source of data used. Emphasis will be put on differences between countries and a deep descriptive analysis of main statistics will be discussed here as well.

The main objective of the last chapter before conclusion is to provide explanation of the results. First, we reveal assumptions of econometric method which have to be satisfied, and then we focus on both models. We will run several regressions, test the assumptions and run Chow's test for structural differences.

1 Theoretical background

In this chapter, we start with a brief survey of scholarly literature on the subject. Subsequently, theoretical background needed for successful understanding of our model, will be presented.

1.1 Survey of the literature

Most of researchers concerned with income inequalities use equation of the type:

$$T = f(Id, Z) \tag{1.1}$$

where T are usually for taxes, social transfers or Perotti's marginal transfers [3], Id denotes index of inequality of disposable (i.e., post-tax and post-transfers) income and Z denotes other variables. The main idea of this thesis is based on Milanovic's paper [6], which used different equation form. He argued that relationship which should be tested can be written in the form of equation where index of redistribution stands on the left side, and index of inequality of factor incomes on the right side. Based on this equation, Milanovic supposed that [6]: *"More market income unequal situations are associated with greater redistribution and an increase in the market income share of given decile is associated with a lower sharegain*[1] *for that decile."*

However, Perotti [2] used the first equation. The reason is that there were no sufficient data, which could have provided factor incomes of households at that time. He studied the relationship between growth and income inequality and asked what influence democracy had on income distribution and growth. The paper dealt with four approaches of growth and income distribution, often used in the literature, namely, it inspected fiscal policy, socio-political instability, imperfect capital markets and "human capital, investments and fertility" [2]. His paper is discussing reliability of data and also reveals his five steps reduced form estimation strategy. Reduced form equation is typical for this type of studies, but as we mentioned earlier, this approach is criticized by Milanovic [6], who used a structural equation. Unlike Milanovic [6],

[1] difference between disposable and market income share- will be discussed later in the chapter

Perotti [2] has constructed different equality variable called "MID", a combined share in disposable income of the middle class (third and fourth quintiles) around 1960. Perotti [2] came up with conclusion, that there is a positive reduced form relationship between income equality and growth, but it is insignificant and weak for poor counties. Concerning democracy affection on income Perotti ([2] p. 23 and 38) mentions: *"Because of high concentration of democracy in rich countries, it is virtually impossible to distinguish an income effect from a democracy effect in the relationship between income distribution and growth. ...the data do not support the idea that more equal societies, particularly those with democratic institutions, grow faster because they generate less demands for redistribution and therefore less distortions."*

Moving to the literature after Milanovic, there is an interesting work composed by Mello and Tiongson [5]. Contrary to Milanovic, they put Gini's coefficient as an indicator of inequality to the right side of equation. Results show that in both estimated equations (linear and nonlinear) income inequality in negatively related to redistribution. What is more income redistribution seems to act as luxury social good in this case. By testing hypothesis on sample containing only low inequality countries or only high inequality countries they confirmed the assumption that relationship between income and redistribution variables indeed depends on the initial level of inequality.

Empirical literature altogether usually uses cross-country average or panel data to investigate the relationship between redistributive government spending and inequality (Mello and Tiongson [5]). Measures of inequality are typically modifications of Gini coefficients, measures of redistribution are most often spending-related or taxes-related variables. Summary of recent studies according to Mello and Tiongson [5] may be found in Appendix A.

1.2 Income Inequality

"The worst form of inequality is to try to make unequal things equal."

Aristotle

Within a frame of this section, we explain the concept of inequality together with explanations why it occurs and how it can be efficiently measured.

In general, "inequality" is "social or economic disparity between people or groups or the condition or an instance of not being equal". Turning our attention to "economic income inequality", it measures disparity between a percentage of population and percentage of income received by that population [7]. When disparity moves up, inequality rises. Concluding from the definition above there are two marginal situations: inequality minimum (each person from the population holds the same amount of income, actual equality) and inequality maximum (when population income is in the hands of single person). In this thesis we concentrate on the macroeconomic reasons of income inequality, such as tax systems, unemployment or social pressure to work [7].

1.2.1 Measurement of inequality

According to "Income inequality project" [7] measuring changes in inequality helps to determine the effectiveness of policies applied. To measure inequality and its alterations, there are six the most often used metrics [7].

- **Range** - the difference between the value of highest and lowest observations.
- **Range Ratio** - the ratio of a value at two predetermined percentiles (e.g. 95/5, 80/20 etc.).
- **The McLoone Index** - sum of all observations below median divided by the median multiplied by the number of observation below median.
- **The Coefficient of variation** - distribution's standard deviation divided by its mean.

- **The Gini Coefficient** - twice the area enclosed between the Lorenz curve end 45 degree equality curve.
- **Theil's T Statistic** - T of income inequality is given by the formula listed in the Appendix B.

In the Table 1 are baseline pros and cons of matrices. We can see that first four indexes are quite easy to understand and compute, whereas The Gini coefficient and Theil's Statistics require more mathematical knowledge. In addition, range and range ratio do not weight observations. The essential problem with the McLoone index, range and range ration is that they use only highest and lowest observations or ignore values above the median as in McLoone's case. In fine, the last three income inequality measures are complex indicators of inequality, the most often used by researchers.

	Pros	Cons
Range	-Easy to Calculate	-Ignores all but two of the observations -Does not weight observations -Affected by inflation -Skewed by outliers
Range ratio	-Easy to calculate -Not skewed by severe outliers -Not affected by inflation	-Ignores all but two of the observations -Does not weight observations
The McLoone Index	-Conveys comprehensive information about the bottom half	-Ignores values above the median -Relevance depends on the meaning of the median value
The Coefficient of Variation	-If data is weighted, it is immune to outliers -Incorporates all data -Not skewed by inflation	-Requires comprehensive individual level data -No standard for an acceptable level of inequality
The Gini Coefficient	-Generally regarded as gold standard in economic work -Incorporates all data -Allows direct comparison between units with different size populations	-Requires comprehensive individual level data
The Theil's T Statistics	-Can effectively use group data -Allows the researcher to parse inequality into within group and between group components	-No intuitive motivating picture -Cannot directly compare populations with different sizes or group structures -Comparatively mathematically complex

Table 1: Income inequality measures

Source: University of Texas Income Inequality Project Tutorials [7]

1.2.2 Gini Coefficient

Following Milanovic [6] we have chosen Gini coefficient as the most suitable measure for this thesis. Therefore we will define and derive the Lorenz curve, and afterwards we will try to illustrate Gini's coefficient graphically and mathematically as well.

To explain a full meaning of Gini Coefficient we have to start with a definition of the Lorenz curve.

> ***Definition*** *(according to Farris [8]): The Lorenz curve for a resource Q (income in our case) is the curve $y = L(p)$, where the Q − poorest fraction ($p \in \langle 0,1 \rangle$)) of the population has a fraction $L(p)$ of the whole.*

In other words, Lorenz curve says how much of the total income is in the hand of bottom 10, 20, …, 100 percent of households. Now, imagine that we have a dataset (for example Belgium97 as we will use later on), where h_j stands for number of units (households), x_j is the average income of each unit, and n is the number of units, where $j = 1, \dots, n$. What is more, the condition is that $x_j < x_k$ if $j < k$. We denote the size of the population by N. Furthermore T will stand for the total amount of income Q and p_j was mentioned before, but in general these mean to be a points along the p − axis between 0 and 1. [8]

To sup up, we have 3 basic equations:

$$N = \sum_{i=1}^{n} h_i \tag{1.2}$$

$$T = \sum_{i=1}^{n} x_i h_i \tag{1.3}$$

$$p_j = \frac{1}{N} \sum_{i=1}^{j} h_i \tag{1.4}$$

Now, we can easily express Lorenz curve as :

$$p: L(p_j) = 0 \quad if\ j = 0$$

$$L(p_j) = \frac{1}{T}\sum_{i=1}^{j} x_i h_i \quad if\ 1 \le j \le n \tag{1.5}$$

To simplify calculation of Gini, according to Farris [8] we can rewrite the equation (1.5) (with a help of probability density function) to Riemann integral:

$$L(p_j) = \sum_{i=1}^{j} \frac{x_i}{\frac{T}{N}}(p_i - p_{i-1}) = \int_0^{p_j} s(p)dp \tag{1.6}$$

Where,

$$s(p) = \frac{x_j}{\bar{X}} \quad for\ p_{j-1} < p \le p_j \tag{1.7}$$

According to Farris [8] share of density s(p) tells us, *"what share of the whole is owned by the portion of the population that falls in a given percentile range"* where $\bar{X}$ is a mean income.

After derivation of the Lorenz curve, we can now finally explain Gini's coefficient. Foremost, we start with theorem and afterwards we explain mathematical and graphical meaning of Gini coefficient.

> ***Theorem 1*** *[8]*: *Suppose G is the Gini index associated with the Lorenz curve* $L(p)$ *and the share density is defined by* $s(p) = L'(p)$ *almost everywhere. Let* $\bar{p}$ *be the expected value of the random variable on [0, 1] whose density function is* $s(p)$. *Then* G *and* $\bar{p}$ *are related by*

$$G = 2\bar{p} - 1 \quad and \quad \bar{p} = \frac{G+1}{2} \tag{1.8}$$

Expected value of the random variable ($\bar{p}$) is expressed as :

$\bar{p} = \int_0^1 ps(p)dp$, where $s(p)$ is constant

$$\bar{p} = \sum_{j=1}^{n} \frac{x_j}{\bar{X}} \int_{p_{j-1}}^{p_j} p\, dp = \sum_{j=1}^{n} \frac{x_j}{\bar{X}} \frac{p_j + p_{j-1}}{2} \Delta\, p_j \qquad (1.9)$$

By using Theorem 1 and equation (1.9) we can finally disclose formula for calculating Gini's coefficient:

$$G = \left(\frac{1}{T} \sum_{j=1}^{n} x_j\, (p_j + p_{j-1})\, h_j \right) - 1 \qquad (1.10)$$

Up to now we successfully defined Lorenz curve and derived Gini's coefficient mathematically. Last but not least is to show Gini's graphical definition presented by Figure 1.

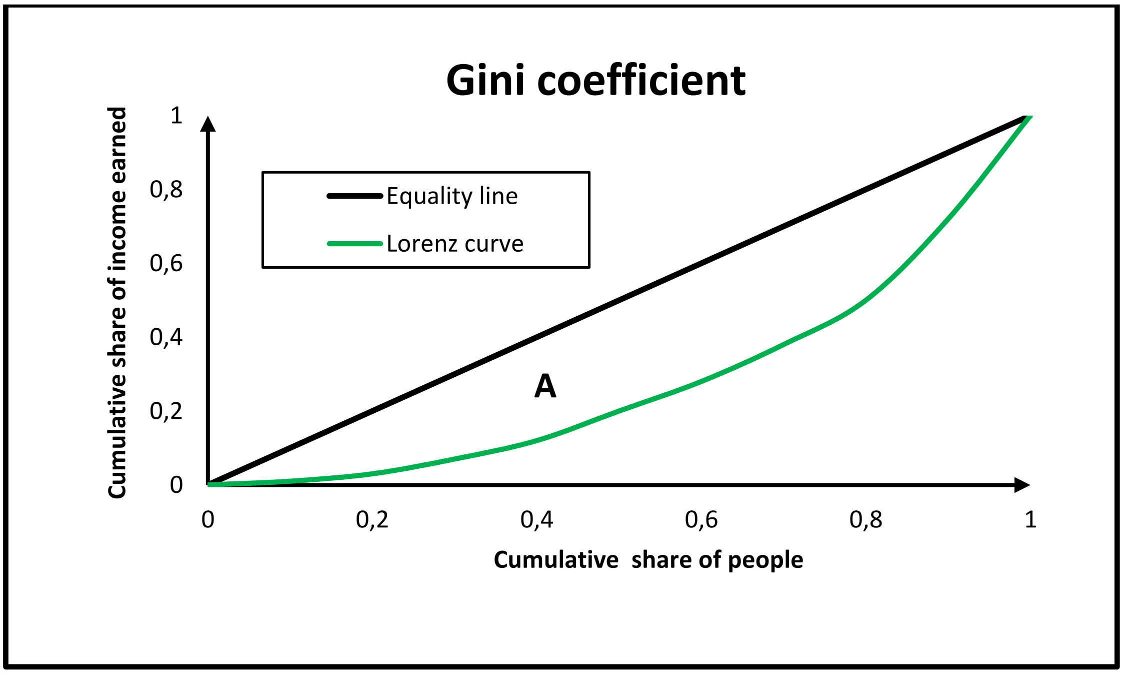


Figure 1: Gini coefficient
Source: Author's graph

In the Figure 1, we see the Lorenz curve (the green line) and the 45 degree line of perfect equality (the black line). On the y-axis a cumulative share of income earned is marked out and on the x –axis we find a cumulative share of people from lowest to highest. The Gini coefficient equals twice the area between equality line and Lorenz curve: $G = 2A$. In the Figure 1 we can see, that if $G = A = 0$ then the Lorenz curve is identical to perfect equation line. On the other hand, when the Lorenz curve is identical to x-axis, $A = 0.5$ and$G = 1$. This means that all income is in the hands of a single person (or a household). [8]

To keep the outline of this chapter we move to the second part. The next section deals with a composition of variables and models.

1.3 Income Inequality Model

We have mentioned above that for this work we were motivated by Milanovic's paper [6]. That is why we decided to except from our pre-transformed model, tested part of his hypothesis as well. However we have used different dataset and modify his model by adding new variables.

Milanovic [6] was a pervasive user of the Luxembourg Income Study (LIS). This is the only one database in the world, publicly offering cross-country micro-level data on income. More information about the LIS and the dataset will be provided in Chapter 2. Both of us are using household level data drawn from household income surveys to test our hypothesis. Starting with composition of variables the most important besides Gini coefficient is a sharegain.

1.3.1 Sharegain

Sharegain was used as a redistribution proxy in both models. At first we will define sharegain, and then we will introduce whole process of creating.

> ***Definition*** *[9]****:*** *Sharegain is a difference between disposable and factor income share of a given decile.*

As it was already referred we have disposable household level income surveys and for each country and year observation we have *"the average per capita income in local currency by decile"* [6]. In this thesis we worked with factor and disposable income. Factor-income is income before taxes and transfers are applied. It is exactly defined as earning plus cash property income (rent, dividends, royalties, but no capital gains, inheritances, etc.), where earnings consist of gross wage and salaries and income from self-employment [10].

Further, net disposable-income is the income after application of taxes and transfers. It is therefore a gross income without income tax and mandatory (employee) contributions like social insurance, unemployment insurance, etc. By the gross income we mean factor income with added occupational pensions and transfer incomes (all benefits from government including – child benefits, sickness benefits, military benefits etc.). [10] These two forms of incomes are our indicators of redistribution and will be transformed to the actual sharegain later in the chapter. As we can see, factor-income shows allocation of various kinds of assets [2] and their relative prices, meanwhile net disposable income *"shows differences in purchasing power among individuals"* [6].

Now, we defined the basic elements of our most important variable. In this thesis we work with decile or percentiles of population. While Milanovic studied all ten deciles, we looked closely only at the second and the fifth decile. However, arising from the definition of sharegain, we are particularly interested in "*disposable and factor income share of given decile".* So, our question is: How much of factor or net disposable income (in given country and year) is kept by bottom half (bottom quintile, etc.) of the poorest people rank by their incomes?

Before answering this question we define sum of all factor incomes in a given country in a given year as:

$$sFI_{ky} = \sum_{i=1}^{n} x_{i,ky} \tag{1.11}$$

where x_i represents individual's factor income, n is the number of individuals in the dataset, k stands for country, with y representing year. Identically for net disposable income:

$$sDI_{ky} = \sum_{i=1}^{n} z_{i,ky} \tag{1.12}$$

[2] Financial, physical and human

In this case z_i is individual's net disposable income. We continue by defining the sum of the factor and disposable income of bottom half (or 20 %, bottom quintile, etc.) of the poorest inhabitants of a country in a given year.

Let l be a boarder person of given percentile or decile p. The person with the highest factor (disposable) income from a given group (e.g., defined by the 20 % of the poorest citizens). In this case, we cannot forget that individuals are ranked according to their incomes from the poorest to the richest. We can define sum of all factor and disposable incomes of the given decile as:

$$sFIp_{ky} = \sum_{i=1}^{l} x_{i,ky} \qquad (1.13)$$

$$sDIp_{ky} = \sum_{i=1}^{l} z_{i,ky} \qquad (1.14)$$

Getting to the end, we can answer our question by defining share of the disposable and factor incomes:

$$shFI_{p,ky} = \frac{\sum_{i=1}^{l} x_{i,ky},}{\sum_{i=1}^{n} x_{i,ky}} \qquad (1.15)$$

$$shDI_{p,ky} = \frac{\sum_{i=1}^{l} z_{i,ky}}{\sum_{i=1}^{n} z_{i,ky}} \qquad (1.16)$$

Finally, according to definition and taking to the account all equations defined in this section, we define sharegain as follows:

$$\begin{aligned} sharegain_p &= shDI_{p,ky} - shFI_{p,ky} \\ &= \frac{\sum_{i=1}^{l} z_{i,ky}}{\sum_{i=1}^{n} z_{i,ky}} - \frac{\sum_{i=1}^{l} x_{i,ky}}{\sum_{i=1}^{n} x_{i,ky}} \end{aligned} \qquad (1.17)$$

1.3.2 Other Variables

➢ **GINI**

In previous section we defined factor and disposable income and derived sharegain as proxy for redistribution. According to the first part of this chapter, we can calculate Gini coefficient for both types of incomes without any problems. In this thesis, except of Gini coefficient for factor and disposable income will be used also Gini in the following form:

$$GINI = Gini_d - Gini_f \tag{1.18}$$

d is an index for the disposable income and f for the factor income. This variable is used only in our model, because Milanovic uses $Gini_f$ (eventually$shFI_{p,ky}$) as proxy for income inequality while testing the redistribution hypothesis.

➢ **Democracy index**

There are a lot of indexes and definitions of democracy in the literature. For instance according to Perotti [2]: "Jodice and Taylor [11], who assigns a value of 1 to a democracy, 0.5 to a "semi-democracy", and 0 to dictatorships for each year in the 1960 – 85 period." Perotti than constructed his own "democracy" dummy based on Jodice and Taylor [11]. He assigned 1 to all countries, in which average value of the Jodice and Taylor's definition over the 1960 – 85 is greater than 0.5, and 0 otherwise [2]. We used Democracy index 2010 by Economist Intelligence Unit (EIU). EIU index has a scale from 0 to 10 and divides countries into 4 main groups [12]:

1. Full democracies – scores of 8 to 10.
2. Flawed democracies – scores of 6 to 7.9.
3. Hybrid regimes – scores of 4 to 5.9.
4. Authoritarian regimes – scores below 4.

This index is based on 60 indicators. Indicators are questions asked in surveys like for example: "Is media coverage robust? Is there open and free discussion of public issues, with a reasonable diversity of opinions?" There are always three possible answers: “Yes” for 1 point, “No” for 0 and “middle way” for 0.5. These indicators are grouped into five categories [12]:

1. Electoral process and pluralism – whether national elections are free and fair.
2. Civil liberties – the security of voters.
3. The functioning of government – influence of foreign powers on government.
4. Political participation – the capability of the civil service to implement policies.
5. Political culture.

In each category, a country can score from 0 to 10, and the overall score (index) is computed as an average of the five category indexes [12].

Urbanization

Urban population refers to people living in the urban areas as defined by national statistical offices. It is calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects [13]. We have chosen this variable, because income distribution may depend on urbanization. According to Kuznets [1], urban areas are usually more unequal as rural areas as well as they have higher per-capita income.

Dummy variables

Dummy variables have been included to specify country and year differences and minimize heterogeneity problems in dataset. Firstly, we have a dummy for post-communistic countries. There are included countries from middle and east Europe. All of them apart from Russia have already jointed European Union. Next dummy indicates countries which were part of Austria-Hungary Empire in 19^{th} century. “Early90” is giving information that this data are from the period between 1990 and 1993. This time can be called a transition period of post-communistic countries, what could have an effect on western or high developed countries as well. In the end, we included categorical variables for region, which distribute countries according to their geographic positions to West, South, North and CEE Europe with Russia.

Herewith, we introduced essential variables. All of them apart from democracy index are kept in percent. Only now we are able to move forward and set up the models.

1.3.3 Models Introduction

In the following lines we will finally introduce income inequality and redistribution models. At first we will discuss Milanovic's model, afterwards we turn to our hypothesis and model. The main characteristics of both models are similar but hypotheses are reversed.

As pointed out in the survey of the literature, Milanovic [6] tested equation:

$$R = f(Im, Z) \tag{1.19}$$

According to this, "R" is indicator of redistribution represented by depend variable $sharegain_{50,ky}$or $sharegain_{20,ky}$ and is positively related to "Im". For factor income inequality (Im) Milanovic [6] put three different variables: $Gini_f$, $shFI_{50,ky}$ and $shFI_{20,ky}$. What is more in his regressions he distinguishes two cases. In the first set of regressions he works with the dataset, where are excluded pension transfers. However he instead adds a variable with information about share of a population over the age of 65. In second set, Milanovic excludes this variable, but uses "factor P income". He also compares results for all countries (24 OECD countries) to only established democracies. The redistribution hypothesis in his paper is tested by running panel data regressions with fixed effects. To bring it together:

> ***Redistribution hypothesis:*** *"The countries with more unequal initial incomes redistribute more." [6]*
>
> ***Tested equations:***

$$sharegain_{p,ky} = \beta_0 + \beta_1 Gini_f + \beta_2 Age65_{ky} + \sum_k \beta_k dummy_{ky} + \varepsilon_{ky}$$

$$sharegain_{p,ky} = \beta_0 + \beta_1 shFI_{,ky} + \beta_2 Age65_{ky} + \sum_k \beta_k dummy_{ky} + \varepsilon_{ky}$$

$$p \in \{20; 50\}$$

> ***Results:*** *"The results show strong support for the redistribution hypothesis. More unequal factor-income countries redistribute more toward the poor and very poor." [6]*

In addition to our own model this thesis tests also first equation of Milanovic's model. Thought, there are few differences. At first, we used different dataset of countries (only European countries). Secondly, because of lack of data for our countries, we used Ordinary Least Squared (OLS) regression instead of panel data. Also, we only worked with factor P income, so we did not need a variable for age. Further, we added last section mentioned variables to the regression.

Now, we will concentrate on our model. We already discussed two possible equations, to be concrete equation: (1.1) and (1.19). We constructed third one. We retained the notion that there should be an index of redistribution (R) and an index of inequality (Im). But in our model the dependent variable is an index of inequality. Therefore, the equation looks as follows:

$$Im = f\ (R, Z) \tag{1.20}$$

There is one more change. While, Milanovic uses Gini for factor income (or shares of it) as an index of inequality, we decided to use$GINI$. In general, our income inequality measure is the change (reduction) in inequality when we move from factor to disposable income. What we ask and want to test is:

Does income redistribution to the poorest population, occurred by government taxes and transfers, leads to greater overall income equality? Is this effect more or less visible in post-communistic European Union countries?

By "overall income equality" we want to highlight the fact that as dependent inequality variable will be used Gini coefficient, measuring overall income inequality, and not share of decile. Both variables are very close but only their differences are allowing us test this hypothesis.

For testing this hypothesis we will use OLS regression method. Our baseline model looks as follows:

$$GINI = \beta_0 + \beta_1\, sharegain_p + \beta_2 democracy + \beta_3 urban^3 + \beta_4 unem^4 + \sum_k \beta_k dummies + \varepsilon \quad (1.21)$$

As already mentioned, this is only the basic model. In Chapter 3 we will run several regressions to find out the most suitable model and relationships between the variables. Furthermore, we will test this model for basic OLS assumptions and Chow's test will be applied to support our hypothesis, that there is a need for splitting Europe dataset. We assume that there are visible differences between post-communistic countries and rest of Europe, what could have impact on results.

Thereby, we finished explanation of theoretical background needed for this work. In the next chapter, we will continue with description and statistical analysis of dataset.

[3] urbanization
[4] unemployment

2 Data sample description

In this chapter, we will describe data sample and data sources. After that we will move to the statistical analyses.

We have already mentioned that in this thesis are primarily used data from Luxembourg Income Study (LIS). LIS is the Cross national data center in Luxembourg and simultaneously a data archive and research center dedicated to cross-national analysis. Their purpose is to enable cross-national comparative research on income and poverty to students and researchers with non-commercial purposes [10]. The harmonized micro-datasets from upper and middle income countries are collected and placed on server. The database consists of household and personal level data on market and government income, demography, employment and expenditures from (mostly democratic) 36 countries since 1968 [10]. Because LIS is the only institution offering this type of data in the world, they are available only to registered researchers usually through$LISSY$. That is a remote-execution data access system for the LIS micro-data, allowing users to submit programs using statistical software packages like Stata, SAS or SPSS. [10]

For this case study, we have used Stata software to draw $GINI$ and sums of factor and disposable incomes from LIS to compose our core variables. We extracted the remaining variable from The World Bank database.

We used data from 24 countries for period from 1990 to 2004. There are 68 observations in total. Each country is a democracy, but eight of them are post-communistic European Union members and Russia. Five observations were in the time of survey newly established post-communistic countries[5]. According to democracy index used in this paper, the most democratic is Norway (9.8). Russia is with the lowest score (4.26) at the bottom of “Hybrid Regimes” (the only country from dataset). In general, Scandinavian countries together with Ireland and Denmark reached

[5] Hungary 1991, Slovakia 1992, Czech Republic 1992, Russia 1992, Poland 1992.

the highest results. The only post-communistic country placed between "Full Democracies" is Czech Republic.

When we compare Czech and Slovak Republic in terms of democracy, the difference is 0.84 point (in spite of the fact, that they were one country until 1993). They both have the lowest score in political participation. This measure can indicate problems in implementation of policies. However, there is a big difference in political culture. Whereas Czech Republic scored 8.13, Slovakia got only 5 points. Such a low score can point out problems with sufficient democracy support in Slovakia.

2.1 Income inequality and inequality reduction- analysis of dataset

In terms of factor income, most income equal countries, with $Gini_f$ around 20 %, are Slovakia and Czech Republic in 1991 – 92. However, the reason is probably due to the fading influence of communistic regimes. But Russia - another post-communistic country, was the most unequal country in Europe first five years after the revolutions (see Table 2). Moreover, Russia has a very high Gini's coefficient in factor and also disposable income, but reduction of inequality was only 5.5 % in 1995. In addition, difference between Russian disposable income inequality and European inequality average is almost 16 %. That shows that Boris Yeltsin's government did not optimized government transfers and taxes to inequality reductions in country. It would be very interesting to see how situation changed over the last 15 years. Unfortunately, required data have not been aired yet.

ALL	Max	Min	St. Dev[6]	Average
$Gini_f$	50.4 (Russia95)	29.2 (Slovakia92)	5.05	40.8
$Gini_d$	44.9 (Russia95)	18.9 (Slovakia92)	5.12	28.06
GINI	24.5 (Sweden95)	2.9 (Greece95)	5.13	12.74

Table 2: Descriptive statistics of Gini coefficients for Europe
Source: LIS [10], World Bank [14] and author's calculations

[6] Standard deviation

In the opposite side of the chart is Sweden. During the decade their initial inequality starts around 45 %, almost the same as in Russia, but as a result of different government policies, they have managed to reduce inequality the most of all European countries – permanently around 22 % each available year.

To forward we now focus on statistics of post-communistic countries. First, interesting is the value of their standard deviation. For factor and disposable income the standard deviation is quite high (Table 3), but for $GINI$ it is the lowest from the whole sample (2.4 %). We explain it by the similar environment of the countries and similar development of their tax systems and social policies after revolutions. From this sample, Poland is in a similar position like Sweden in the sense of reduction inequality. When we are not concerning Russia, Poland has the highest factor income inequality. However (Table 3), $GINI$ is there highest among this sample (in 2004). There are a few facts, which can explain this situation and will examine them in the next section.

CEE	Max	Min	St. Dev.	Average
$Gini_f$	50.4 (Russia95)	29.2 (Slovakia92)	6.25	38
$Gini_d$	44.9 (Russia95)	18.9 (Slovakia92)	7.01	29,5
GINI	12.4 (Poland04)	3.5 (Russia92)	2.4	8,5

Table 3: Descriptive statistics of Gini coefficients for CEE
Source: LIS [10], World Bank [14] and author's calculations

2.1.1 Case study – Poland

We start this case study by looking at the Table 4. High $GINI$ index shows that government deals with the problem of inequality among population. Overall situation in Poland 2004 is in Table 4.

	GINI	$Gini_f$	$Gini_d$	Unemp[7]	Sharegain50	$shFI_{50}$	Income Tax	Gov.[8]
Poland04	12.5	44.5	32	19	19.6	6.78	progressive	SLD[9]

Table 4: Descriptive statistics of Gini coefficients for Poland
Source: LIS [10], World Bank [14], [15], [16] and author's calculations

High factor income inequality is logically accompanied with low factor income share of bottom half of population. Combined with high unemployment it probably leads to government participation and needed redistribution. However, unemployment does not have to be a reason of high redistribution, but can be also a result. Because of high social and unemployment benefits, and high taxes, low income population will choose not to work. What is more, Poland has a progressive tax system and Leszek Miller from Social Democratic Party of Poland served as the Prime Minister in 2004 [15]. All of these are reasons for income inequality reductions.

In general, high share of government expenditures on transfers and high share of government revenues from direct taxes on total revenues, lead to the increase in redistribution. According to IMF[10] report [17] share of government expenditures on transfers and subsidies was 54.8 % and the share of government revenues on direct taxes was 54.8 % in 2004. These data support the hypothesis and obviously lead to high reductions in income inequality. Poland can be a social state with fast growing and developing economy. However, they had permanently high unemployment rate, what can be their major problem and reason of high inequality reductions. We can deduce that unemployment can be an important proxy for income inequality analysis.

[7] Unemployment
[8] Government- party with the highest number of mandates in parliament
[9] Social Democracy of Poland
[10] International Monetary Fund

Non-CEE	Max	Min	St. Dev.	Average
$Gini_f$	48.7 (UK04)	32.1 (Luxembourg94)	4.1	41.8
$Gini_d$	35.2 (Uk04)	20.9 (Finland91)	4.36	27.7
GINI	24.2 (Sweden95)	2.8 (Greece95)	5.05	14.1

Table 5: Descriptive statistics of Gini coefficients for non-CEE
Source: LIS [10], World Bank [14] and author's calculations

As a last thing in this chapter, we would like to go back to Kuznets and his hypothesis. As mentioned in previous chapter, he predicted that urban areas are more unequal as rural, as they have higher income per capita [1]. According to our data, United Kingdom together with Belgium, Sweden, Denmark and Luxemburg are the most unequal countries in Europe. Moreover, Gini coefficient of factor income shows that they belong also between the most unequal countries in Europe. That is why, we can conclude that our data support Kuznets hypothesis [1].

To conclude, income inequality is in Europe reduced on average by 12.7 % (Table 2) as measured by Gini coefficient. Despite the fact that factor and disposable income inequality is approximately the same, reduction is on average higher in non-CEE countries (14.1 % - Table 5) than in still developing CEE countries (8.5 % - Table 3) However, between CEE countries belongs also influential, but feebly income inequality reducing Russia what can bias these results. Anyway, on average, third of European initial income inequality is reduced by governments. While distribution of human, physical and financial assets is the most unequal in Russia and United Kingdom, the smallest differences in purchasing power among individuals were in Slovakia 1992 and Finland 1991.

2.2 Redistribution analysis

Now we proceed to another part of this thesis, where we are going to analyze descriptive statistics of redistribution proxy.

We start this section by discussion of results shown in Table 6. Norway's average gain of bottom quintile is low, moreover, gain of bottom half poorest is the smallest in Europe. Therefore, Norway obviously does not redistribute very much. Nevertheless, its factor income share of bottom half is the highest – 22 % and at the same time, share of bottom quintile is only 2.35 %.

ALL	Maximum	Minimum	St. Dev.	Average
$sharegain_{50}$	22.3 (Poland95)	7.4 (Norway91)	3.4	14.7
$sharegain_{20}$	9.5 (Poland95)	2.75 (Russia95)	1.17	6.3

Table 6: Descriptive statistics of sharegain in Europe
Source: LIS [10], World Bank [14] and author's calculations

What is more, a similar situation exists in the shares of disposable income (Appendix C). This alerts that Norway has a strong middle class, which earns approximately one fifth of total income. We have already discussed Poland in this paper. Poland reduced income inequality by 12.5 % (see Table 4) through government, what (see in Table 6 and Table 7) led to the extra high redistribution.

CEE	Maximum	Minimum	St. Dev.	Average
$sharegain_{50}$	22.3 (Poland95)	8.9 (Romania97)	4.1	16.39
$sharegain_{20}$	9.5 (Poland95)	2.75 (Russia95)	1.74	6.39

Table 7: Descriptive statistics of sharegain for CEE
Source: LIS [10], World Bank [14] and author's calculations

As mentioned before, the redistribution in Sweden follows a similar pattern. Focusing on non-CEE countries the highest redistributors to bottom quintile are Benelux countries (Table 7). They redistribute on average around 8 % of total income to the poorest. Of course Russia is at the bottom of the chart as the least redistributive together with Greece. However, opposite to Russia, factor income share of bottom half was more than 5 % higher in Greece in 1995. Compared to the sharegain average, this number is really low.

Non-CEE	Maximum	Minimum	St. Dev.	Average
$sharegain_{50}$	19.7 (Sweden95)	7.4 (Norway91)	3.1	14.1
$sharegain_{20}$	8.2 (Netherland004)	4.45 (Greece95)	0.9	6.3

Table 8: Descriptive statistics of sharegain for non-CEE
Source: LIS [10], World Bank [14] and author's calculations

In this analysis we are apart from sharegain of bottom half and quintile, interested also in sharegain of middle class and share of factor incomes. We defined middle class sharegain as the share of third, fourth and fifth decile. According to statistics the worst redistributors to middle class are Luxembourg and Switzerland. They both redistribute to middle-class less than two percent of overall income, but their disposable and factor income shares are around 20 % (Appendix C). Moreover, their unemployment is extremely low.

Move to the share of factor income of bottom quintile now. Unfortunately, there was a problem with the share of factor income data for the poorest quintile. When we called for sum of factor incomes in Slovakia, Czech Republic (for all provided years), and for few more observations[11], database issued zeroes. We were not able to find out whether the zeroes are inaccurate, so sharegain may be overestimated.

Data shows that one third of observations are in negative numbers. We have already said that Poland is a great redistributor to bottom deciles. Their factor income share of last quintile was -4.1 % in 1995 and only -1.25 % in 2004 and it is the lowest share among our observations. However, the reason for these low numbers can be hide in high unemployment. But we have to highlight that the share is negative for one third of our observations.

The greatest share of factor incomes was in Romania in 1997 – almost 2.4 %. Romania belongs to Southeastern post-communistic countries. After the revolution in 1989, they struggled with problems. However, the "Big Bang", with a help of the IMF and the World Bank, took place in 1997 [18]. Romania led by a right-wing government implemented a package of reforms, pursued privatization and adopted more

[11] CZ92/96/04, HU91/94, Pl92, Sk92/96, Be92, Ir04,

than hundred new laws in 1997 [18]. Although the Romanian average gain of bottom half and bottom quintile of poorest population is under the European average, the level of the shares of factor and disposable income is still in the better half of our observations. Furthermore the unemployment was only 5.5 % in 1997.

By the description of Romania we finished our analysis of income redistribution. We went through all key points. In addition, we tried to look closely to post-communistic countries and analyze their political environment and possible reasons of inequality.

So far we managed to explain theoretical background and we introduced our empirical models in first chapter. In this chapter we analyzed data sample, because it was necessary for a full comprehension of this thesis.

3 Empirical part

Finally, we move to the most important part of this work – empirical modeling and testing. In this chapter we will use concepts and data introduced in previous chapters to run our regressions.

The outline of this chapter is as follows. We start with the examination of Ordinary Least Squares (OLS) assumptions as they are crucial for correctly run regressions and have to be fulfilled. Then, we move to adapted Milanovic and our "new income inequality" models. They both will be provided by detailed explanation of results. The last section of this chapter will contribute with testing hypothesis separately on post-communist countries and the rest of Europe. We will run the regressions on both samples and if it is possible, we will try to illustrate the results.

3.1 OLS Assumptions

For all models in this chapter Ordinary Least Squares estimation of coefficients in linear regression models with cross-sectional data will be used. As is known, this method has 6 basic assumptions. When all the assumptions are satisfied, we can say that our OLS estimator $\hat{\beta}$ is a normally distributed random variable, unbiased and efficient. What is more, it is the best linear unbiased estimator (BLUE) [19].

ASSUMPTIONS:

- ✓ Disturbances are random variables drawn from a normal distribution. [19] Doornik-Hansen's test for normality of residuals will be applied and we will look at the Q-Q plot as a graphical normality test.
- ✓ Mean of this distribution is zero, in other words $\boldsymbol{E}[\varepsilon_i] = 0$ [19].
- ✓ Variance of this distribution is constant, homoscedasticity is present: $Var[\varepsilon_i] = \sigma^2$ [19]. After each regression White's test will be used as a test of this assumption. When heteroscedasticity found, we either try to eliminate it or use a different model.
- ✓ Disturbances are not autocorrelated: $Cov[\varepsilon_i \varepsilon_j] = 0$ [19]

- ✓ Disturbances are not correlated with the explanatory variable: $Cov(x_{ik}, e_i) = 0$ [19].
- ✓ Explanatory variables are not linearly dependent, so no multicolinearity is presented. [19] We will check this assumption now, by examining the correlation matrix (Appendix E). Correlation higher than 0.8 shows problems and is marked as multicolinearity. Correlation equal to 1 is perfect multicolinearity, when one variable is linearly related (multiplied) to another. Fortunately, our correlation matrix does not prove any signs of multicolinearity, except of $logGDP$ and GDP. However, this correlation was expected and we will not use these variables together in one model.

By now, all OLS assumptions should be explained. These assumptions will be checked after each estimated equation. Finally, we can move to regressions of our models. At first, we will start with adapted model put forward in Milanovic's paper [6].

3.2 Milanovic's model

At first place, we shall recall Milanovic's hypothesis. He proposed that the countries with more unequal initial incomes redistribute more [6]. In general, he expected positive relationship between $Gini_f$ and $sharegain_p$. We will regress equations with dependent variables $sharegain_{20}$ and $sharegain_{50}$ and the core independent variable $GiniFI$ (positive sign expected). As other independent variables we use urbanization (positive sign expected), GDP per capita (positive sign expected), unemployment (positive sign expected), democracy (positive sign expected), early90 (negative sign expected), post-communistic (positive sign expected) and AH (positive sign expected). Estimated equation is as follows:

$$\begin{aligned} sharegain_{20(50)} &= \beta_o + \beta_1 Gini_f + \beta_2 urban + \beta_3 logGDP \\ &+ \beta_4 unemp + \beta_5 democ + \beta_6 early90 \\ &+ \beta_7 postcom + \beta_8 AH + \varepsilon \end{aligned} \tag{3.1}$$

Depend variable	const.	$Gini_f$	urban	logGDP	unempl	democ.	early90	postcom	AH	R^2
$sharegain_{20}$										
(1)	-3,55	-9.223***	0.038***	0.988***	0.082**	0,006	-0,115	1.197*	1.131**	49.5 %
	(-0.484)	(-0.465)	(0.415)	(0.902)	(0.341)	(0.007)	(-0.115)	(1.197)	(1.131)	
(2)	-4,11	-8.79***	0.037***	1.038***	0.083**	-0,003	-	1.262**	1.149**	49.3 %
	(-0.530)	(-0.444)	(0.404)	(0.947)	(0.342)	(-0.0038)	-	(1.262)	(1.149)	
(3)	-4,1	-8.78***	0.037***	1.034***	0.083***	-	-	1.259**	1.149**	49.3 %
	(-0.529)	(-0.443)	(0.404)	(0.943)	(0.342)	-	-	(1.259)	(1.149)	
(4)	2.78	-9.49***	0.044***	-	0.065**	0.413***	-0.432	0.11	1.08**	42.4 %
	(-0.14)	(-0.409)	(0.415)	-	(0.231)	0.384	(-0.369)	(0.10)	(0.923)	
$sharegain_{50}$										
(1)	-12,13	6,38	0.189***	0,32	0.325***	0,406	-0,806	1,837	4.95***	51.6 %
	(-0.39)	(0,093)	(0.594)	(0.084)	(0.389)	(0.128)	(-0.233)	(0.533)	(1.43)	
(2)	-16.04**	9,37	0.182***	0,669	0.328***	0,335	-	1,195	5.079***	50.4 %
	(-0.48)	(0.137)	(0.57)	(0.177)	(0.39)	(0.106)	-	(0.665)	(1.473)	
(3)	-17.04**	8,14	0.18***	1.11*	0.33***	-	-	2,62	5.03***	50.5 %
	(-0.49)	(0.119)	(0.57)	(0.29)	(0.396)	-	-	(0.759)	(1.458)	
(4)	-10.09*	6.29	0.191***	-	0.32***	0.537	-0.908	1.49	4.93***	51 %
	(-0.36)	(0.097)	(0.601)	-	(0.382)	(0.169)	(-0,263)	(0.437)	(1.431)	

Table 9: Econometric results for Milanovic approach
Source: Output from program GRETL
Data Source: LIS [10], World Bank [14]

We used OLS regression with $N = 68$ observations and the results are shown in Table 9.

The main numbers in the Table 9 show values basic estimated coefficients. The numbers in parenthesis represent standardized beta coefficients. We have used z-score according to Wooldridge [20] to standardize each variable:

Let y_i be a dependent variable with the average value $\bar{y}$, and let $x_{i1}, \dots, x_{ik}$ be explanatory variables with average values $\overline{x_1}$,..., $\overline{x_5}$. Let $\widehat{\sigma_y}$ be the sample standard deviation of dependent variable and let $\widehat{\sigma_1}$,... ,$\widehat{\sigma_k}$ be sample standard deviations of $x_1, \dots, x_k$. Then each variable in equation (3.1) has been standardized by replacing it with its z-score [20]:

$$\frac{(y_i - \bar{y})}{\widehat{\sigma_y}} = \frac{\widehat{\sigma_1}}{\widehat{\sigma_y}} \widehat{\beta_1} \left[\frac{x_{i1} - \overline{x_1}}{\widehat{\sigma_1}} \right] + \dots + \frac{\widehat{\sigma_5}}{\widehat{\sigma_y}} \widehat{\beta_k} \left[\frac{x_{ik} - \overline{x_k}}{\widehat{\sigma_k}} \right] + (\frac{\widehat{\varepsilon_i}}{\widehat{\sigma_y}}) \quad (3.2)$$

$$z_y = \widehat{b_1} z_1 + \widehat{b_2} z_2 \dots + \widehat{b_k} z_k + error \quad (3.3)$$

The new coefficients are: $\widehat{b_j} = \left(\frac{\widehat{\sigma_j}}{\widehat{\sigma_y}} \right) \widehat{\beta_j}$ for $j = 1, \dots, k$. [20]

In the first model with dependent variable $sharegain_{20}$, two variables are insignificant (democracy and early90). Because their p-values are high (0.97 and 0.69) we do not reject the null hypothesis that the coefficients β_5 and β_6 are zeroes. Therefore we have estimated the models without them (3)[12]. Now, we can see that all variables (except constant) are significant on 0.05 significance level. The same process was applied to the model with $sharegain_{50}$. There is still insignificant $GiniFI$ but as it is a core variable, we will not omit it. However, we have only managed to explain approximately 45% of dependent variable variance included in the data in each model (R^2 in Table 9). Considering OLS assumption, test for the normality of residuals is satisfied for all estimated models. According to White's and Breusch-Pagan test we do

[12] Means equation (3) in Table 9

not reject the null hypothesis that data is homoscedastic (conducting from high p-values) in any of the estimated models. All test's results are visible in Appendix F.

Focusing on the results of regressions, following Milanovic [6] we hypothesized a positive relationship between factor-income inequality and redistribution. Results confirm this hypothesis for a redistribution to bottom half of the poorest, but not to the bottom quintile. Moreover coefficient of factor income inequality for bottom half estimates is insignificant. Increase in factor-income inequality by 1 % increases redistribution to the bottom half by 0.0629 percentage point, but it causes 0.095 percentage point loss of the poorest (equations (4) in Table 9). One standard deviation increase in $Gini_f$, (5.05 Gini points), decrease the share of the poorest in disposable income by 2.06 %[13]. Furthermore, the unemployment increased by one percentage point, raises redistribution by 0.33 percentage point. Moreover, the data show that the same relative movement of urbanization has a larger effect on redistribution for the bottom quintile then unemployment does (equation (1)). The largest standardized effect on redistribution GDP per capita has. If GDP per capita rises by one percentage point, redistribution to the poorest in country increases by 1.034 %. However, nominal GDP used is not very effective index, as far is influenced by inflation (especially in post-communistic countries). Because of that, we run also equation (4). Fortunately, in the Table 9 we can see, that after excluded GDP per capita our coefficients does not change significantly except from already mentioned post-communistic dummy apparently influenced by inflation, and democracy index, which decreased.

Results for democracy are interesting as well. Before omitting nominal GDP per capita, that democracy influenced much more bottom half then bottom quintile. What's more in the case of bottom quintile was in model (2) negative. However, results of equation (4) show that it was influence very much by nominal GDP and probably hided inflation. These results are hard to explain since factor income inequality is highly influenced by institutions, democracy including. It is very difficult to distinguish democracy effect from other effects in small heterogeneous sample like ours.

[13] Standard deviation of $Gini_f$ times standardized estimate of $Gini_f$.

To end explanation of this model we return to the relationship between factor income inequality and redistribution to the poorest. Although, we expected this relationship to be positive, results of model show opposite situation. In other words, further increase in factor income inequality leads to decrease in redistribution to the poorest. We presume that bottom quintile of poorest (according to their income) consist primarily of unemployed and low wages population and is most affected by changes in factor income inequalities and redistribution. First affect after increase in factor income inequality is connected to more than proportional increase in redistribution to the poorest. However, in short time share of government expenditure will be higher than share of government revenues, and so redistribution to the poorest start to decrease.

In survey of the literature we mentioned, that Mello and Tiongson [5] deduced from their equation that redistribution is a luxury social good. We can conclude that our adjusted Milanovic's model applied on European dataset supports this hypothesis and also shows that redistribution to the poorest quintile of population is a luxury social good as deduction about showed.

3.3 New income inequality model

While the last model analyzed how redistribution depends on factor income inequality, now we will focus on the question, whether redistribution to bottom half and bottom quintile of poorest leads to greater reduction of income inequality. In other words, we expect positive relationship between $GINI$ and $sharegain$. Actually, positive relationship is expected among all explanatory variables used in the regression. However we expect lower reduction of overall income inequality in post-communistic countries than in the rest of Europe. The essential estimated equation is:

$$\begin{aligned} GINI = \beta_0 + \beta_1 sharegain_{20(50)} + \beta_2 urban + \beta_3 unempl \\ + \beta_4 democracy + \beta_5 logGDP + \beta_6 AH \\ + \beta_7 postcom + \beta_8 North + \beta_9 South + \beta_{10} West \\ + \beta_{11} early90 + \varepsilon \end{aligned} \quad (3.4)$$

We have again used Ordinary Least Squares estimation method with 68 observations. When estimating this equation, we obtained results presented in Table 10:

GINI	hc[14](1)	hc(2)	(3)	(4)
Constant	-18.287** (-0.252)	-5.792* (0.1651)	-16.35* (0.145)	1.167 (0.082)
Sheregain$_{20}$	0.984*** (0.224)	0.838** (0.191)	-	-
Sheregain$_{50}$	-	-	0.353** (0.237)	0.395** (0.265)
Urbanization	0.132*** (0.279)	0.157*** (0.333)	0.083 (0.176)	0.079 (0.168)
Democracy	0.295 (0.062)	-	0.746* (0.158)	-
Unemploym.	0.278*** (0.223)	0.324*** (0.261)	0.157** (0.126)	0.045 (0.035)
logGDP	0.935 (0.166)	-	1.128 (0.201)	-
AH	5.147*** (1.003)	4.749** (0.925)	1.902 (0.370)	2.42 (0.471)
Post-comm.	-5.980*** (-1.165)	-9.939*** (-1.936)	-4.074* (-0.794)	-6.504*** (-1.267)
North E.	6.625*** (1.291)	5.449*** (1.062)	4.122*** (0.8032)	6.038*** (1.176)
South E.	-4.838** (-0.942)	-6.348** (-1.237)	-4.120** (-0.8028)	-3.236** (-0.631)
West E.	-1.134 (-0.221)	-2.343 (-0.456	-2.758** (-0.537)	-1.8 (-0.351
Early90	-	-0.21 (-0.041)	-0.604 (-0.117)	-1.201 (-0.234)
R^2	86.2%	83.8%	72%	68.2%

Table 10: Econometric results for our income inequality model
Source: Output from program GRETL, Data Source: LIS [10], World Bank [14]

[14]Means heteroscedasticity corrected equation.

Let us focus on the OLS assumptions presented in the introduction of this chapter. We have already checked multicolinearity by the correlation matrix (see Appendix E). Secondly, we checked assumption of homoscedasticity. Unfortunately, White's test proved presence of heteroscedasticity in the models where sharegain$_{20}$ was an explanatory variable. This can possibly be caused by inconsistency in data mentioned in Chapter 2. We tried to correct this problem by Feasible GLS method as follows:

Firstly, we ran OLS regression of our model and recorded residuals. Then we calculated log of square residuals and used it as a dependent variable in auxiliary regression which generated fitted values of log of square residuals. Moving on we calculated weights in the form:$\widehat{h(x_i)} = \exp(\widehat{\log(e_i^2}))$. Log transformation was used to ensure that $h(x_i)$ is positive. In the end, we estimated the original equation by Weighted Least Squares using reciprocal $h(x_i)$ as weights [21].

The coefficients resulting from this method are in column (1) and (2). We used Feasible GLS method instead of robust standard errors because in small samples as is ours robust t-statistics might be distributed differently resulting from the condition n→∞ [21], where n is a number of observations in the sample. For the models (3) and (4) where sharegain$_{50}$ was used, we do not reject null hypothesis that homoscedasticity is present. Furthermore, according to Doornik-Hansen test, residuals are normally distributed in each model. As far we presume that mean of distribution of disturbances is zero, all OLS assumptions should be fulfilled.

Comparing to Milanovic model from previous section, "New income inequality model" has coefficient of determination two times higher. Models with redistribution to poorest managed to explain more than 80 % of our data. This number is slightly lower for the redistribution to poor, but our model can be still considered as qualitative. Good sign is that our core variable $sharegain_{20\ (50)}$ is according to the p-value significant at the least on 95 % significance level. Similarly to the last model, democracy appears to be insignificant in first model. However, in model (3) is significant on 10 % significance level. The surprise is, that GDP per capita (in logarithm) is not significant in any model as well as the variable "Early90". We also included geographical dummy variables- North, South and West Europe (Central Europe was excluded as we have to drop one region to explain results correctly). They appeared to be significant, except of West Europe. However, these are a categorical variables connected together, so we cannot omit West Europe from our regression, even if it is insignificant.

In the hypothesis we assumed positive relationship between redistribution and reduction of inequality, and our data support this implication. If redistribution to the poorest quintile increases in sharegain by one percentage point, disposable income increases as well. Further it causes decrease in disposable income inequality (↓ $Gini_d$). Thus, in condition that factor-income inequality does not change significantly, $GINI$ rises by 0.98 % according to results of equation (1). Table 10 shows that hypothesis holds for all tested equations.

Let us focus on first tested equation. We have already discussed influence of sharegain in previous paragraph. Urbanization is as we predicted in positive relationship with$GINI$. When urbanization increases by 1 %, inequality decreases by 0.132 %. One of the possible explanations is connected to unemployment. In urban areas employment opportunities and higher wages are usually better, so when we assume that in rural areas there is a higher unemployment and lower wages, the poorest people tend to move to urban areas. After the movement and possibly new employment, their wages increase together with their disposable income. Thus disposable income inequality of the poorest decreasing and reduction of inequality is increasing. However, this reduction is caused more by labor market than by government transfers and taxes. Though, supposition about positive relationship between unemployment and inequality reduction at the start of this subsection, was right.

Now consider unemployment. When unemployment increases, share of the poorest in country as well as share of government expenditure on transfers also increases. These expenditures have to be compensated, so government raises direct taxes. Therefore, share of the government revenues from direct taxes to total gains increases. Because of this, redistribution to poor rises and as we discussed before, our model shows that increase in redistribution (represented by sharegain) heighten reduction of income inequality. According to Table 10, one standard deviation increase in unemployment causes 0.223 standard deviation increase in $GINI$. We can conclude, that these results confirm our theory above.

First equation estimations shows, that democracy measured by EIU's democracy index has the smallest standardized effect on reduction of income inequality and is insignificant. But, in third equation estimation with sharegain_{50}, the same relative movement in democracy index has a larger effect on inequality reduction than unemployment does.

Next, GDP per capita was used in equation (1) and (3). GDP is excluded in equations (2) and (4), but as we can see other coefficients have not changed significantly. Moreover, in both cases nominal GDP was insignificant. One percentage point increase in GDP means that state is becoming wealthier or there is a high inflation (as it is in nominal value). Either way leads to decrease in inequality by 0.94 % in first estimated equation and by 1.13 % in third equation.

Now, we move to explanation of dummy variables used in the model. At first, we will focus on geographical dummies. On average, the results for all 4 estimated equations were similar. We will use equation (4) to demonstration our findings. The coefficient by Western Europe is not statistically different from zero on 10 % significance level, except of equation (3), where is significant on 5% significance level. Countries belonging to Northern Europe reduce income inequality 6 % more than Central-European countries, other things being fixed. Moreover, Southern European countries reduce inequality 3.2 % less. These results confirm our hypothesis from Chapter 2, where we discussed that Scandinavian countries are highly redistributive and significantly reduce income inequalities. Our estimates confirm that they are very strong social states. On the other hand, Southern-European countries like Spain, Italy or Greece were at the bottom of the $GINI$chart. We can say that estimations of our model confirm hypothesis from data sample analysis. Data from period 1990 – 1993 do not denote any big difference to rest of data as countries in this period reduces inequality only approximately 1 % less, other things being equal.

Last but not least, we will focus on post-communistic countries. Governments in post-communist countries reduce inequality on average 6.6 % less than rest of Europe according to our estimates. Descriptive statistics in Chapter 2 show that post-communist countries redistribute on average the same, or to the poorest even more, than the rest of Europe. Their GDP per capita is on average higher, however the reason is in high inflation, so we cannot take this statistic into the account. They have on average higher unemployment by 2 percentage points, but Democracy index is lower as well as urbanization (by 10 percentage points). We could conclude that urbanization and democracy are decisive factors.

In the end of this work we decided to run the Chow's test on equation (3), to find out whether there are any structural differences with respect to post-communistic countries in our data. Conducting from very low p-value (0.0008), the test shows that it would be gainful to divide the model into two separate models. Unfortunately, our data are not satisfactory for post-communistic countries. Therefore, estimates are heteroscedastic and biased, so not sufficiently presentable.

Conclusion

This thesis is an empirical investigation of the relationship between income inequality and redistribution and their impact on the poorest segment of population. For studying these relations we used two models, both based on Milanovic's hypothesis [6].

Milanovic first proposed that income inequality should by an explanatory variable not for taxes and transfers separately, but for full redistribution process. We agreed with this claim and investigated this relationship in the first model. However, this thesis also enriches this topic by studying inverse process as well (e. g., process, where redistribution is an explanatory variable for income inequality).

Economic income inequality is here measured by Gini coefficient as it allows direct comparison between samples with different size of population and generally it is most often used proxy. For income redistribution we defined special variable called sharegain, what is a difference between disposable and factor income (pre-tax and pre-taxes) share of given decile. Both investigated models are based on these variables. The first is adjusted Milanovic's [6] model, which studies positive relationship between factor income inequality and redistribution. We enlarged it by more variables (unemployment, urbanization, dummies, etc.) and applied on a different dataset. The second is our "New income inequality model". We put redistribution to the role of explanatory variable and overall inequality measure served as explained variable.

Both investigated models were tested on a European dataset, consisting of 24 democratic countries from which 8 are post-communist for period 1990 – 2005. We provided deep descriptive analyses of this dataset in the second chapter. Our analysis shows that the smallest differences in purchasing power among individuals were in Slovak and Czech Republic in 1991 – 1992. Russia is the worst distributor of all assets, but government policies do not reduce these differences. Opposite situation is in Sweden, the most inequality-reducing country. According to low standard deviation of income inequality reduction, CEE countries have a similar policy-environment. Norway has a very strong middle class and the most redistributive non-CEE country union is Benelux. We also provided support for Kuznets [1] hypothesis that urban areas are more unequal than rural. In this thesis special emphasis was put on post-communist countries. Based on used democracy index, Czech Republic is the only post-communistic country belonging to "full democracies". Poland is the most inequality reducing post-communist country but this is accompanied with high unemployment what can be a reason and an important proxy for our further investigation. On average, third of European initial income inequality is reduced by governments.

Empirical estimations in the third part of this thesis were computed by OLS method with 68 observations. Results of the first model, when redistribution to bottom quintile was used as a proxy, do not support hypothesis that more factor income inequality leads to higher redistribution. We are explaining this relationship by the fact that redistribution to the bottom quintile is luxury social good. The estimated parameters also do not support the hypothesis that unemployment constitutes an influential proxy. So we can deduct that in the case of Poland, high unemployment can be caused by high redistribution not vice-versa.

Since the "New income inequality model" was heteroscedastic, we corrected it by using feasible GLS method. This model supports the hypothesis that redistribution occurred by government taxes and transfers leads to a greater income equality. However, unexpected result is that urbanization has a very strong effect on reductions of income inequalities.

Concerning post-communist countries, they on average reduce overall inequality eight percent less than the rest of Europe. Chow's test showed that dataset has structural differences in respect to post-communism. However, our data are not satisfactory for post-communist countries. It would be very interesting to compare these models across two different data samples e.g., post-communist countries versus the rest of Europe. As for now, we are leaving this idea open for a further research.

Appendices

Appendix A Summary of Recent Studies

Source	Sample size	Period	Data structure	Measures		Correlation	
				Redistribution[a]	Inequality	Sign[b]	Significance
Basset, Burkett, and Putterman (1999)	Up to 54 countries	1970-1985	Cross-country average	Social security and welfare	Mostly Q3 in 1960s[c]	Generally negative	Inconsistent
Easterly and Rebelo (1993)	Not available	1970-1988	Cross-country average	Spending variables	Gini, various income shares	Positive	Significant
Figini (1998)	Up to 63 countries	1970-1990	Cross-country average	Tax rates, total revenue, and total spending	Gini coefficient in 1970	Nonlinear	Significant
Gouveia and Masia (1998)	50 U.S. states	1979-1991	Panel	Spending variables	Ratio of mean to median income	Generally negative	Generally significant
Lindert (1996)	14 OECD countries	1962-1981	Panel	Spending variables	Income gap index	Negative	Generally insignificant
Meltzer and Richard (1983)	U.S. average	1937-1977	Time series	Spending variables	Ratio of mean to median income	Positive	Significant
Milanovic (2000)	24 mostly OECD countries	1967-1997[d]	Panel	Gain by poorest quintile or poorest half	Pretransfer Gini coefficient	Positive	Significant
Panizza (1999)	46 U.S. states	1970-1980	Cross-state average	Tax, tax progressivity, and spending variables	Q3 in 1970[c]	Generally positive	Inconsistent
Partridge (1997)	48 U.S. countries	1960-1990[e]	Panel	Tax, employment and spending variables	Pretax Gini, Q3	Inconsistent	Generally significant
Perotti (1992)	40 democracies	1970-1985	Cross-country average	Spending variables	Q3 in 1970[c]	Negative	Insignificant
Perotti (1994)	52 countries	1970-1985	Cross-country average	Transfers	Q3 in 1960[c]	Negative	Insignificant
Perotti (1996)	49 countries	1970-1985	Cross-country average	Tax rates and spending variables	Q3 and Q4 in 1960	Generally positive	Generally insignificant
Perrson and Tabellini (1994)	13 OECD countries	1960-1981	Cross-country average	Transfers	Q3 in 1965[c]	Positive	Insignificant
Rodrigues (1999)	50 U.S. states	1984-1994	Time series and cross-state average[f]	Spending variables	Distribution skewness	Inconsistent	Insignificant
Tanninen (1999)	Up to 45 countries	1970-1988	Cross-country average	Spending variables	Adjusted Gini coefficient in 1970s[g]	Inconsistent	Generally insignificant

Table 11: Summary of Recent Studies: Inequality and Redistribution
Source: L. de Mello, E. R. Tiongson: Income Inequality and Redistributive Government Spending [5]

Note: See text for further discussion of these results.

a. In percentage of GDP unless otherwise indicated.
b. Negative means greater inequality is associated with less spending.
c. A higher Q3 means greater income equality. For consistency with other studies in the reporting of main results, Q3 and Q4 are taken to mean –Q3 and –Q4.
d. Total number of observations is 79.
e. Total number of observations is 144.
f. Refers to national time-series.
g. Adjusted for variations in the Gini definition.

Appendix B Theil's T Statistic

Mathematically, with individual level data Theil's T statistic of income inequality is given by [7]:

$$T = \sum_{p=1}^{n} \left\{ \left(\frac{1}{n}\right) * \left(\frac{y_p}{\mu_y}\right) * \ln\left(\frac{y_p}{\mu_y}\right) \right\}$$

where n is the number of individuals in the population, y_p is the income of the person indexed by p, and μ_y is the population's average income.

Appendix C Dataset

	$shDI_{50}$	$shFI_{50}$	$sharegain_{50}$	$shDI_{20}$	$shFI_{20}$	$sharegain_{20}$	$Gini_f$	$Gini_d$	GINI	GDP p.c.	urbanization	democracy	unemployment	early90	R-U	Middle E.	West E.	North E.	South E.	post-comm
AT94	26.32	11.27	15.05	6.22	0.01	6.21	38.6	27.9	10.7	25375.07	65.8	8.49	3.53	0	1	1	0	0	0	0
AT97	27.23	11.75	15.48	6.68	0.20	6.48	38.4	26.6	11.9	25953.48	65.8	8.49	4.20	0	1	1	0	0	0	0
AT04	27.89	11.78	16.11	7.03	0.03	7.00	43.9	26.8	17.1	35357.73	66.4	8.49	4.93	0	1	1	0	0	0	0
CZ92	29.48	12.39	17.09	7.55	0.00	7.55	29.6	20.8	8.8	2903.08	75.0	8.19	2.60	1	1	1	0	0	0	1
CZ96	26.91	10.42	16.49	6.92	0.00	6.92	34.0	25.9	8.0	6011.69	74.5	8.19	3.89	0	1	1	0	0	0	1
CZ04	27.29	7.47	19.82	7.29	0.00	7.29	36.5	26.8	9.8	10720.89	73.6	8.19	8.29	0	1	1	0	0	0	1
DE94	29.31	17.00	12.31	7.49	0.28	7.21	44.0	27.0	17.0	26329.72	73.3	8.38	8.40	0	0	1	0	0	0	0
DE00	28.53	12.77	15.76	7.21	0.11	7.10	45.4	26.6	18.8	23114.23	73.1	8.38	7.75	0	0	1	0	0	0	0
DE04	25.13	10.37	14.75	6.33	0.09	6.24	48.4	28.7	19.7	33268.77	73.3	8.38	10.29	0	0	1	0	0	0	0
EE04	23.18	9.81	13.36	5.18	-0.18	5.37	40.0	34.0	6.1	8918.70	69.4	7.68	9.97	0	0	0	0	1	0	1
HU91	24.75	8.93	15.82	5.66	0.00	5.66	38.8	28.3	10.6	3222.57	65.7	7.21	7.40	1	1	1	0	0	0	1
HU94	23.77	3.66	20.11	5.84	0.00	5.84	42.5	32.5	10.0	4014.32	65.3	7.21	10.85	0	1	1	0	0	0	1
PL92	27.04	6.32	20.72	6.86	0.00	6.86	34.5	27.4	7.1	2197.99	61.4	7.05	13.32	1	1	1	0	0	0	1
PL95	26.24	3.94	22.29	5.41	-4.08	9.49	40.4	31.1	9.3	3603.79	61.5	7.05	13.34	0	1	1	0	0	0	1
PL99	27.65	9.54	18.11	6.71	-1.02	7.74	40.6	29.0	11.6	4344.72	61.7	7.05	12.53	0	1	1	0	0	0	1
PL04	26.37	6.78	19.59	6.21	-1.25	7.46	44.5	32.1	12.4	6620.06	61.5	7.05	18.97	0	1	1	0	0	0	1
RO97	27.04	18.07	8.97	6.85	2.39	4.46	37.9	28.1	9.8	1564.51	53.8	6.60	5.52	0	1	1	0	0	0	1
SI97	29.57	16.23	13.34	7.13	0.01	7.12	32.1	25.0	7.2	10207.45	50.7	7.69	7.24	0	1	0	0	0	1	1
SK92	30.01	13.40	16.61	7.67	0.00	7.67	29.2	18.9	10.3	2689.41	56.5	7.35	10.40	1	1	1	0	0	0	1
SK96	27.65	8.60	19.05	6.74	0.00	6.74	30.3	24.1	6.2	5074.34	56.5	7.35	11.34	0	1	1	0	0	0	1
Be92	29.43	11.18	18.26	7.45	0.00	7.45	31.3	22.4	8.9	23057.84	96.6	8.05	6.70	1	0	0	1	0	0	0
Be97	27.58	8.94	18.64	7.12	0.10	7.02	45.4	25.0	20.5	24500.38	96.9	8.05	8.96	0	0	0	1	0	0	0
Ch92	25.06	13.99	11.06	4.75	0.29	4.46	39.4	29.6	9.7	36506.49	73.4	9.09	2.82	1	0	1	0	0	0	0
Ch00	29.32	21.19	8.13	7.64	1.40	6.24	38.5	28.3	10.3	34787.16	73.3	9.09	2.66	0	0	1	0	0	0	0
Ch02	29.49	19.75	9.74	7.42	0.76	6.67	38.1	27.2	10.9	38247.42	73.3	9.09	2.93	0	0	1	0	0	0	0
Ch04	29.55	19.11	10.45	7.40	0.70	6.69	38.7	26.3	12.4	49121.95	73.3	9.09	4.31	0	0	1	0	0	0	0
DK92	25.36	8.29	17.07	6.17	-0.29	6.46	43.5	23.6	19.9	29051.35	84.9	9.65	9.03	1	0	0	0	1	0	0
DK95	26.71	8.60	18.10	7.13	-0.21	7.34	43.5	22.1	21.4	34809.41	85.0	9.65	6.99	0	0	0	0	1	0	0
DK00	26.58	9.42	17.16	7.09	-0.16	7.25	42.3	22.5	19.9	29992.94	85.1	9.65	4.48	0	0	0	0	1	0	0
DK04	26.48	8.63	17.84	7.00	-0.23	7.23	42.4	22.8	19.6	45310.12	85.7	9.65	5.51	0	0	0	0	1	0	0
ES90	26.25	14.06	12.18	6.35	-0.03	6.38	35.3	30.5	4.8	13414.57	75.4	8.16	16.00	1	0	0	0	0	1	0
ES95	23.23	9.22	14.00	5.13	-0.07	5.20	43.3	35.1	8.1	15150.95	75.9	8.16	22.68	0	0	0	0	0	1	0
FI91	29.66	16.79	12.87	7.27	0.72	6.55	37.9	20.9	17.0	24978.73	61.4	9.19	6.50	1	0	0	0	1	0	0
FI95	29.85	14.76	15.10	7.59	0.24	7.35	41.9	21.7	20.2	25587.29	61.4	9.19	15.27	0	0	0	0	1	0	0
FI00	27.61	15.96	11.66	6.77	0.53	6.24	41.7	25.0	16.7	23514.46	61.1	9.19	9.73	0	0	0	0	1	0	0
FI04	26.60	13.23	13.37	6.31	0.27	6.04	44.5	26.1	18.4	36134.83	62.1	9.19	8.79	0	0	0	0	1	0	0
FR94	26.81	11.10	15.71	6.76	0.15	6.60	46.5	28.9	17.6	23039.37	74.7	7.77	12.59	0	0	0	1	0	0	0
GR95	22.47	10.94	11.53	4.45	0.00	4.45	37.7	34.8	2.9	12386.53	59.3	7.92	9.06	0	0	0	0	0	1	0
IR94	23.24	11.78	11.46	5.37	0.00	5.37	41.0	33.9	7.2	15509.77	57.7	8.79	14.58	0	0	0	1	0	0	0
IR95	22.39	11.25	11.14	5.18	0.00	5.18	41.7	34.5	7.2	18582.32	57.9	8.79	11.99	0	0	0	1	0	0	0
IR96	22.72	11.52	11.20	5.20	0.08	5.13	41.1	33.4	7.7	20339.01	58.1	8.79	11.75	0	0	0	1	0	0	0
IR04	21.86	3.46	18.40	5.44	0.00	5.44	41.2	32.0	9.2	45504.04	60.2	8.79	4.47	0	0	0	1	0	0	0
IT91	27.20	15.69	11.50	7.12	0.12	7.00	38.6	29.0	9.6	21058.72	66.7	7.83	10.10	1	0	0	0	0	1	0

	$shDI_{50}$	$shFI_{50}$	$sharegain_{50}$	$shDI_{20}$	$shFI_{20}$	$sharegain_{20}$	$Gini_f$	$Gini_d$	GINI	GDP p.c.	urbanization	democracy	unemployment	early90	R-U	Middle E.	West E.	North E.	South E.	post-comm
IT93	24.68	11.49	13.18	5.94	0.09	5.84	44.6	34.0	10.6	17964.80	66.8	7.83	10.24	1	0	0	0	0	1	0
IT95	24.77	11.24	13.52	5.87	-0.10	5.97	44.6	33.8	10.8	19808.59	66.9	7.83	11.67	0	0	0	0	0	1	0
LU91	30.00	20.49	9.51	8.36	0.57	7.79	32.2	23.9	8.3	35444.41	81.3	8.88	1.48	1	0	0	1	0	0	0
LU94	30.21	17.75	12.47	8.33	0.17	8.16	32.1	23.5	8.6	43560.94	82.5	8.88	3.48	0	0	0	1	0	0	0
LU97	29.05	16.05	13.00	8.26	0.07	8.19	35.7	26.0	9.7	44145.43	83.3	8.88	2.53	0	0	0	1	0	0	0
LU04	27.72	14.05	13.67	7.23	0.04	7.19	38.8	26.8	12.0	74419.60	83.0	8.88	5.11	0	0	0	1	0	0	0
NL91	27.54	10.83	16.71	6.26	0.00	6.26	38.0	26.3	11.7	20130.40	69.5	8.99	7.28	1	0	0	1	0	0	0
NL94	26.81	10.66	16.15	4.78	-0.02	4.80	37.9	25.6	12.3	22832.53	72.0	8.99	7.16	0	0	0	1	0	0	0
NL99	29.69	12.59	17.10	7.53	-0.04	7.57	35.3	23.0	12.2	26033.31	76.0	8.99	3.62	0	0	0	1	0	0	0
NL04	30.72	14.96	15.77	8.22	0.03	8.19	43.8	26.1	17.7	37458.54	79.5	8.99	4.65	0	0	0	1	0	0	0
NO91	29.35	21.98	7.37	7.23	2.35	4.87	38.1	23.1	15.0	28077.38	72.4	9.80	5.41	1	0	0	0	1	0	0
NO95	26.99	14.45	12.54	6.30	0.23	6.08	42.3	24.0	18.4	34155.92	73.8	9.80	4.89	0	0	0	0	1	0	0
NO00	27.22	19.03	8.19	6.27	1.18	5.09	43.9	25.9	18.0	37472.37	76.1	9.80	3.44	0	0	0	0	1	0	0
NO04	26.80	13.97	12.83	6.36	0.24	6.12	47.5	28.2	19.3	56311.49	77.1	9.80	4.37	0	0	0	0	1	0	0
RU92	16.92	7.93	8.99	3.08	-0.30	3.37	47.0	43.4	3.6	3095.08	73.4	4.26	5.22	1	0	1	0	0	0	0
RU95	16.28	4.29	11.99	2.71	-0.04	2.75	50.4	45.0	5.5	2669.94	73.4	4.26	9.49	0	0	1	0	0	0	0
SE92	28.35	16.82	11.53	6.43	0.47	5.96	44.9	22.8	22.1	30819.72	83.4	9.50	5.72	1	0	0	0	1	0	0
SE95	27.62	7.91	19.71	6.71	-0.23	6.94	46.5	22.0	24.5	28726.06	83.8	9.50	9.05	0	0	0	0	1	0	0
SE00	26.36	8.63	17.73	6.50	-0.15	6.65	45.0	25.6	19.4	27879.15	84.0	9.50	5.81	0	0	0	0	1	0	0
SE05	22.44	7.57	14.87	5.79	-0.04	5.83	43.6	23.8	19.8	41065.82	84.3	9.50	7.68	0	0	0	0	1	0	0
UK91	23.00	8.21	14.79	5.35	-0.02	5.38	45.1	33.8	11.3	18386.53	88.8	8.16	8.38	1	0	0	1	0	0	0
UK94	23.25	4.36	18.89	5.76	0.00	5.76	48.3	34.1	14.1	18328.39	88.9	8.16	9.57	0	0	0	1	0	0	0
UK95	23.24	5.73	17.50	5.61	0.00	5.61	47.3	34.3	13.0	19943.77	89.0	8.16	8.59	0	0	0	1	0	0	0
UK99	22.29	5.23	17.06	5.12	-0.10	5.22	47.9	35.0	12.9	25604.86	89.3	8.16	5.95	0	0	0	1	0	0	0
UK04	22.96	5.35	17.61	5.64	-0.10	5.74	48.7	35.2	13.5	36781.81	89.6	8.16	4.63	0	0	0	1	0	0	0
AT94	26.32	11.27	15.05	6.22	0.01	6.21	38.6	27.9	10.7	25375.07	65.8	8.49	3.53	0	1	1	0	0	0	0

Table 12: Dataset
Source: Luxembourg Income Study Database [10], World Banka Database [14]

Note:

AT	Austria	SK	Slovak Republic	IR	Ireland
CZ	Czech Republic	BE	Belgium	IT	Italy
DE	Germany	CH	Switzerland	LU	Luxembourg
EE	Estonia	DK	Denmark	NL	Nederland
HU	Hungary	ES	Spain	NO	Norway
PL	Poland	FI	Finland	RU	Russia
RO	Romania	FR	France	SE	Sweden
SI	Slovenia	GR	Greece	UK	United Kingdom

Table 13: Explanation of abbreviation

Appendix D Descriptive statistics of data

	Minimum		Maximum		Average	Standard deviation
$shDI_{50}$	16.28	(RU95)	30.72	(NL04)	26.34	2.95
$shFI_{50}$	3.46	(IR04)	21.98	(NO91)	11.63	4.53
$sharegain_{50}$	7.37	(NO91)	22.29	(PL95)	14.71	3.45
$shDI_{20}$	2.71	(RU95)	8.36	(LU91)	6.42	1.11
$shFI_{20}$	-4.08	(PL95)	2.39	(RO97)	0.08	0.74
$sharegain_{20}$	2.75	(RU95)	9.49	(PL95)	6.35	1.17
$Gini_f$	29.2	(SK92)	50.4	(RU95)	40.8	5.1
$Gini_d$	18.9	(SK92)	45.0	(RU95)	28.1	5.1
GINI	2.9	(GR95)	24.5	(SE95)	12.7	5.1
GDP p.c.	1564.51	(RO97)	74419.60	(LU04)	24017.57	14809.81
urbanization	50.7	(SI97)	96.9	(BE97)	72.8	10.8
democracy	4.26	(RU)	9.80	(NO)	8.11	1.17
unemployment	1.48	(LU91)	22.68	(ES95)	7.91	4.12

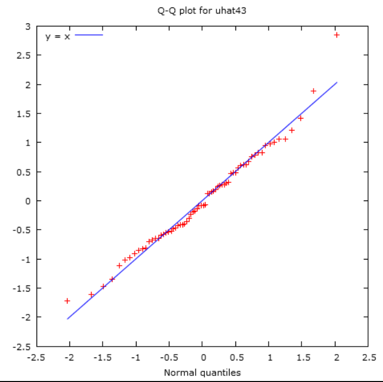
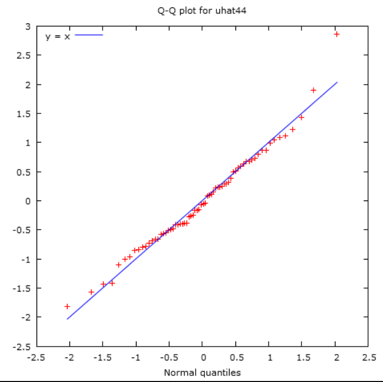
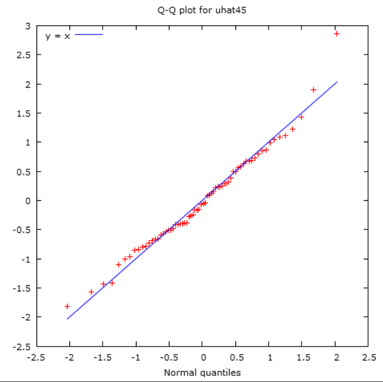
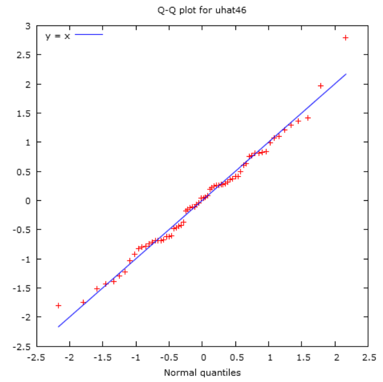
Table 14: Descriptive statistics of data
Source: Luxembourg Income Study Database [10], World Banka Database [14]

Appendix E Correlation matrix of variables

sharegain$_{50}$	sharegain$_{20}$	GINI	urbanization	democracy	unemployment	early90	A-H	Middle E.	West E.	North E.	South E.	Post comm.	Log GDP	GDP p.c.	Gini$_f$	
1.00	0.45	0.18	0.17	-0.10	0.26	-0.15	0.40	0.10	0.10	-0.09	-0.19	0.38	-0.22	-0.23	0.04	**sharegain$_{50}$**
	1.00	0.25	0.10	0.28	-0.04	-0.07	0.27	0.05	0.03	-0.01	-0.10	0.23	0.14	0.16	-0.39	**sharegain$_{20}$**
		1.00	0.40	0.60	-0.14	-0.18	-0.29	-0.27	-0.12	0.66	-0.33	-0.37	0.53	0.48	0.49	**GINI**
			1.00	0.28	-0.33	0.04	-0.48	-0.35	0.36	0.17	-0.22	-0.45	0.46	0.44	0.31	**urbanization**
				1.00	-0.32	-0.11	-0.41	-0.50	0.10	0.55	-0.15	-0.46	0.79	0.69	-0.03	**democracy**
					1.00	-0.05	0.10	-0.04	-0.12	-0.10	0.37	0.22	-0.36	-0.49	0.18	**unemployment**
						1.00	-0.00	-0.00	-0.07	-0.02	0.14	0.04	-0.22	-0.22	-0.31	**early90**
							1.00	0.68	-0.36	-0.32	-0.07	0.83	-0.67	-0.55	-0.42	**A-H**
								1.00	-0.48	-0.42	-0.25	0.54	-0.55	-0.40	-0.18	**Middle E.**
									1.00	-0.37	-0.22	-0.33	0.30	0.25	0.02	**West E.**
										1.00	-0.20	-0.21	0.33	0.31	0.24	**North E.**
											1.00	-0.05	-0.06	-0.19	-0.09	**South E.**
												1.00	-0.79	-0.65	-0.44	**post comm.**
													1.00	0.90	0.25	**log GDP**
														1.00	0.19	**GDP p.c.**
															1.00	**Gini$_f$**

Table 15: Correlation matrix of variables
Source: Output from program GRETL

Appendix F Tests for OLS assumptions

Depend variable	White's test	Breusch-Pagan's test	Doornik-Hansen's test	Q-Q plot
	p values			
$sharegain_{20}$				
(1)	0.387	0.798	0.104	
(2)	0.879	0.799	0.085	
(3)	0.878	0.728	0.085	
(4)	0.628	0.710	0.340	

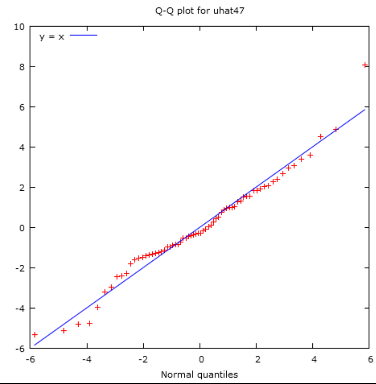
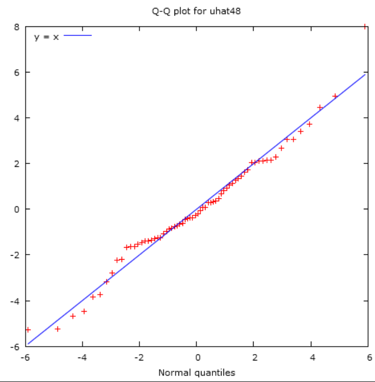
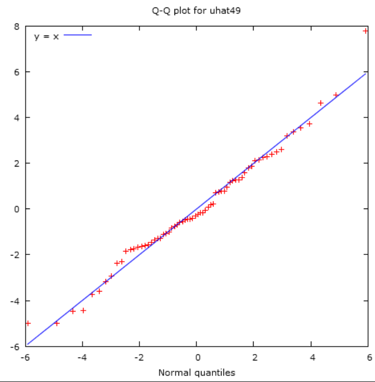
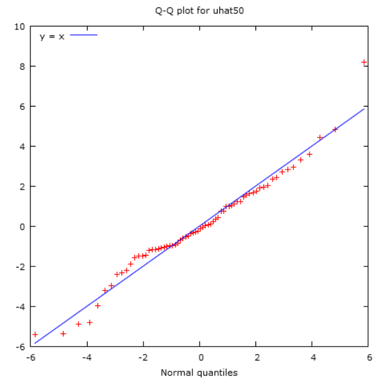
Depend variable	White's test	Breusch-Pagan's test	Doornik-Hansen's test	Q-Q plot
		p values		
sharegain$_{50}$				
(1)	0.206	0.232	0.041	
(2)	0.108	0.209	0.072	
(3)	0.104	0.313	0.194	
(4)	0.079	0.179	0.018	

Table 16: Test for OLS assumptions: Milanovic's model
Source: Output from program GRETL

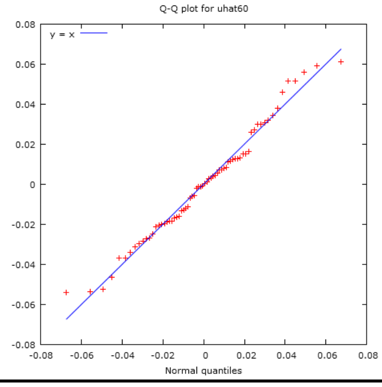
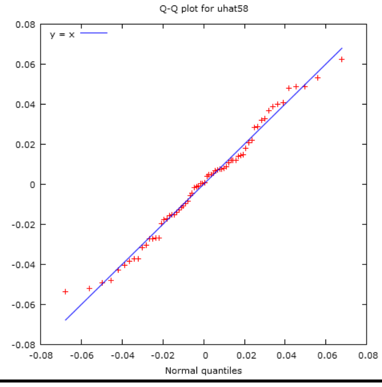
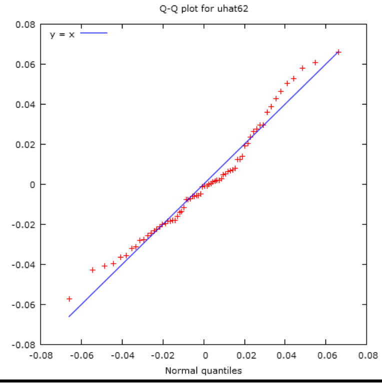
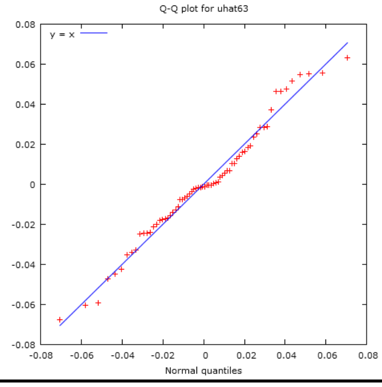
GINI	White's test	Breusch-Pagan's test	Doornik-Hansen's test	Chow's test	Q-Q plot
	p values				
(1)	0.074	0.657	0.632		
(2)	0.067	0.707	0.780		
(3)	0.133	0.735	0.134	0.0008	
(4)	0.105	0.261	0.848		

Table 17: Test for OLS assumptions: "New income inequality model"
Source: Output from program GRETL

References

[1] S. Kuznets, „Economic Growth and Income Inequality," *American Economic Review,* zv. 45, pp. 1-28, 1955.

[2] R. Perotti, „Growth, Income Distribution and Democracy: What the Data Say," Columbia University, New York, 1995.

[3] R. Perotti, „Growth, Income Distribution and Democracy: What the Data Say," *Journal of Economic Growth,* zv. 1, pp. 149-187, 1996.

[4] R. Benabou, „Inequality and Growth," Luxembourg Income Study, Luxembourg, 1996.

[5] L. de Mello a E. R. Tiongson, „Income Inequality and Redistributive Government Spending," *Public Finance Review,* zv. 34, pp. 282-305, 2006.

[6] B. Milanovic, „The median voter hypothesis, income inequality and income redistribution: an empirical test with the required data.," *European Journal of Political Economy,* zv. 16, pp. 367-410, 2000.

[7] P. Burkholder, „University of Texas Inequality Project," 2003. [Online]. Available: http://utip.gov.utexas.edu/tutorials/intro_ineq_studies.ppt. [Cit. 2011].

[8] F. A. Farris, „The Gini Index and Measures of Inequality," *The American Mathematical Monthly,* pp. 851-864, December 2010.

[9] B. Milanovic, „Four critiques of the redistribution hypothesis: An Assessment," *European Journal of Political Economy,* zv. 26, pp. 147-154, 2009.

[10] „LIS Cross-National Data Center in Luxembourg," [Online]. Available: http://www.lisdatacenter.org. [Cit. April-June 2011].

[11] D. Jodice a D. L. Taylor, World Handbook of Social and Political Indicators, New Haven: Yale University Press, 1988.

[12] The Economist Intelligence Unit, „Democracy index 2010, Democracy in retreat," The Economist, 2010.

[13] „Urban population (% of total)," The World Bank, 2010. [Online]. Available: http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS. [Cit. 2011].

[14] The World Bank, „The World Bank Indicators," [Online]. Available: http://data.worldbank.org/indicator. [Cit. April-June 2011].

[15] „Republic of Poland, Rzeczpospolita Polska," [Online]. Available: http://zhenghe.tripod.com/p/poland/facts2.html. [Cit. 2011].

[16] Ministerstwo finansów, „Tax Administration and Tax System in Poland," Warszawa, Poland, 2004.

[17] Staff Representatives for the 2004 Consultation , „Staff Report for the 2004 Article IV Consultation," International Monetary Fund, Washington, D.C., 2004.

[18] „Restructuration and Privatization Romania - The Romanian Shock Therapy 1997," Romania Central, [Online]. Available: http://www.romania-central.com/economy-of-romania/the-economy-of-romania/32-macroeconomic-development-in-romania-1989-2006/321-sub-periods-of-the-romanian-transition-process/3214-1997-2000-restructuration-and-a-new-wave-of-privatizations-in-romania/. [Cit. 2011].

[19] B. Petold-Gebicka, „Violation of basic assumptions, Heteroskedasticity," Institute of Economic Studies, Charles University in Prague, 16. November 2010. [Online]. Available: http://ies.fsv.cuni.cz/default/file/download/id/14913. [Cit. 2011].

[20] J. M. Wooldridge, Introductory Econometrics: A Modern Approach, South-Western

College Pub; 2 edition, 2002.

[21] B. Petold-Gebicka, „Violation of basic assumptions, Heteroskedasticity," Institute of Economic Studies, Charles University in Prague, 23. November 2010. [Online]. Available: http://ies.fsv.cuni.cz/default/file/download/id/14912. [Cit. 2011].