The Effect of Human Capital on Income Inequality

An Econometric Analysis

Cand.merc. in Applied Economics and Finance
Department of Economics
Copenhagen Business School 2014

Author
Anette Løndal Johansen

Supervisor
Battista Severgnini

Date: 09-10-2014

Number of characters (with spaces): 125.337

Executive Summary

The question of how inequality is generated and evolves over time has been a main concern of social scientists for more than a century. The world is experiencing a global income inequality at a high level, whereby the richest eight per cent of the world's population earn half of the world's total income, while the remaining 92 per cent are left with the other half. The dimension of global inequality is likely to become even more relevant as the world becomes more integrated. At the same time income inequality is seen as an important component for a country's overall development, which makes it important to understand the sources that are affecting it.

The purpose of this thesis is to investigate and analyse the effect of human capital on income inequality. Human capital is important because it is the knowledge and competencies that can be used to produce economic value, but also for its relation to economic growth and the distribution of income. This thesis uses educational attainment as a proxy for human capital to investigate its effect. Income inequality is presented as the Gini coefficient, which measures the degree of inequality in the distribution of income in a country. This thesis outlines the importance of education because it influences the skills and competencies of individuals as well as peoples productivity. Therefore, improved education can be important for the wage people will receive, thus impact the income they will hold. The relation between education and income inequality has been the motivation for the empirical investigation in this thesis.

The results presented are taking advantage of new and broader compiled datasets that help explain the effect of human capital on income inequality. The final dataset used for the empirical investigation contains data on 123 countries from 1960 to 2010. This dataset are making it possible to use econometric methods that among others are able to address the problem of reverse causality i.e. are people more educated because they have a higher income or do people have a high income because they have a higher education? A two-least square estimation is used to address this problem of endogeneity with the use of parents' education as an instrument. The result of the instrumental variable estimation is presenting a positive and significant relation between improved educational attainment and income inequality. The empirical results presented in this thesis are all supporting this positive relation, that is, improved educational attainment is reducing income inequality.

Table of Contents

Chapter 1 ..... **5**

1. Introduction **5**

1.1 Problem statement **7**

1.2 Limitation **7**

1.3 Structure **8**

Chapter 2 ..... **9**

2. Literature Review **9**

2.1 Economic Development **9**

2.2 The Distribution of Income **11**

2.2.1 Why do we have income inequality? 12

2.3 Definition of Human Capital **13**

2.4 Human Capital and Income Inequality **15**

2.5 Partial Conclusion **18**

Chapter 3 ..... **19**

3. Methodology **19**

3.1 Panel Data **19**

3.2 The Effect of Educational Attainment on Income Inequality **21**

3.2.1 OLS Estimation 21

3.2.2 The Fixed Effects Model 22

3.2.3 Instrumental Variables 23

3.3 Discussion of the Models **26**

3.4 Partial Conclusion **27**

Chapter 4 ..... **28**

4. Data and Descriptions **28**

4.1 Data **28**

4.1.1 The Gini Coefficient as the measure of Income Inequality 31

Figure 1 – The Gini coefficient presented graphically **32**

4.2 Discussion of the Data **32**

4.3 Descriptive Statistics **34**

4.4 The development in Income Inequality **36**

4.5 The development in Educational Attainment **39**

4.6 Partial Conclusion **40**

Chapter 5 ..... **41**

5. Empirical results **41**

5.1 The Effect of Educational Attainment on Income Inequality **41**

5.1.1 Ordinary Least Square estimation 41

5.2 The Effect of Education on Income Inequality using Fixed Effect and Instrumental Variables **44**

5.2.1 Fixed effects 44

5.2.2 Instrumental variables 47

5.3 The effect of education on income inequality when additional variables are added **48**

5.3.1 OLS estimation 48

5.3.2 Instrumental variables estimation 52

5.3.3 Fixed effects 54

5.4 Summary of the results **55**

5.6 Partial Conclusion 59
Chapter 6 60
6. Comments 60
6.1 The relationship between human capital and income inequality 60
6.2 Implication and limitations of the results 61
6.3 Ideas for further research 62
Chapter 7 64
7. Conclusion 64

Bibliography

Appendix

Chapter 1

1. Introduction

Income inequality is today an important economic fact and has long been a topic of interest among economists. Income is unequally distributed in the world, both within and between countries. The data on income inequality shows that between-country inequality is accounting for the larger part of global income inequality, but despite the predominance of between-country inequality, the within-country inequality is an important contributor to total income inequality. Income inequality is by many seen as an important component for a country's overall development, which makes it important to understand the sources that are affecting it. The positive effect from human capital has been widely recognised in the literature, which suggests that human capital is important for economic growth and favourable for individuals and societies. The literature highlights education as one of the factors affecting the level of income inequality, which is the focus of this thesis.

Education is important because it increases the skills and competencies of individuals as well as their productivity. A workers' productivity is essential for the wage he will be receiving, thus a higher productivity will allow him to earn a higher wage and increase his income. A higher income is allowing people access to better food, proper health care, and clean water among others. A higher income can therefore improve some of the things people value as having a good life. This indication of a positive effect on people's income resulting from improved education is interesting when looking at income inequality and is the motivation for the empirical investigation.

This thesis uses educational attainment data from Barro and Lee (2013) to proxy human capital and income inequality measure from the Standardizing World Income Inequality Database, which is presented as the Gini coefficient. The Gini coefficient measures the extent to which the distribution of income among countries deviates from a perfect equal distribution. A Gini coefficient equal to zero expresses perfect equality and a Gini coefficient equal to 1 expresses maximal inequality.

All over the world the development in educational attainment has increased from 1960 to 2010. The world population aged 15 and above is in 2010 estimated to have an average of 9.2 years of schooling, which have been steadily increasing from 4.9 years in 1960. Income inequality has in the same period been fairly stable. In 1960 the average Gini coefficient was 38 compared to 36 in 2010. So does improved educational attainment have a positive effect when it comes to decreasing income inequality?

In recent years there has been renewed interest in the understanding of the dynamics and determinants of income distribution. This is in part motivated by the availability of new datasets and advances in the theories of economic growth and development. As the purpose of this thesis is to investigate the effect of human capital on income inequality, this thesis takes advantage of new and broader compiled datasets that can help explain the purpose outlined for this thesis. The empirical investigation uses econometric techniques such as ordinary least square, fixed effects and instrumental variables in order to fully explore the data and control for biases that may occur when performing the econometric analysis. The use of fixed effects allows us to adjust for unobserved country specific effects and to control for the heterogeneity that may exists among countries in the world. The instrumental variable estimation is allowing us to address the endogeneity problem, which might be a potential problem in the ordinary least square regression.

The causation between educational attainment and income inequality is interesting when the outline of this thesis is pursued. Are people more educated because they have a high income or do people have a high income because they have a high education? The instrumental variable regression is successfully used to address the problem of reverse causality and is allowing the results of the estimation to only show the effect coming from improved educational attainment and not the other way around. More specifically this thesis uses parents' education as an instrument and is able to present results that deal with the endogeneity that may occur when it comes to explain the effect of educational attainment on income inequality.

1.1 Problem statement

Based on the areas of interest presented in the introduction the problem statement of this thesis is given as follows:

The purpose of this study is to investigate the effect of human capital, in the form of educational attainment, on income inequality in the world. The study will investigate the effect by using different econometrics methods such as ordinary least squares, fixed effects and instrumental variable estimations to obtain a more robust result.

The problem statement of this thesis will be answered by investigating why we have income inequality in the world and why we should pay attention to it. It will be investigated which role human capital plays when looking at income inequality and what effect we see in a panel data analysis. Also it will be investigated if the effect of educational attainment is persistent when using parents' education as an instrument and when individual country dimensions are taken into account. The analysis will in conclusion look at the effect on income inequality when life expectancy, population growth, GDP, openness and government consumption are included in the regression.

1.2 Limitation

The problem statement has outlined the purpose of this study, which is to investigate the effect of human capital on income inequality. Human capital is in this thesis presented as educational attainment and it uses average years of schooling to determine the impact of education. Implicitly, the assumption in this study is that the quality of schooling does not vary among countries. To measure the quality of schooling you could look at the inputs into education such as textbooks and teachers, and the output from education. If the difference in the amount and quality of schooling differ among countries the measure of human capital as the average years of schooling could understate the true difference in the level of human capital of workers in those countries, but this will not be further discussed in the study.

The data used for the empirical investigations has been collected from different sources which all can be subject to different forms of measurement errors. This problem of measurement errors can be mitigated by instrumental variables, which will be used in this thesis. The data will be discussed but it is out of the scope of this paper to further test the quality of the data.

Further limitations will be presented throughout the paper when appropriate.

1.3 Structure

The study is divided into seven chapters. Chapter 2 presents the definitions of income inequality and human capital. The chapter describes why we have income inequality and the relationship between income inequality and human capital. This chapter should give the reader a better understanding and knowledge of why human capital is interesting when looking at income inequality. Chapter 3 presents the methodology of the empirical study. The chapter discuss the methodology used to investigate the effect of human capital on income inequality. The advantages and limitations of the chosen empirical methods are also presented. Chapter 4 presents the data used in the empirical study. The different data sources are presented and discussed. The chapter also presents descriptive statics and outlines the development in income inequality and educational attainment. Chapter 5 presents the empirical results. In this chapter the effect of educational attainment on income inequality is presented using OLS estimation, fixed effects estimation, and instrumental variables estimation. Chapter 6 presents the limitations and implications of the study and the results. Ideas for further investigations are also proposed. Chapter 7 is the final chapter, which concludes the main findings of the study.

Chapter 2

2. Literature Review

This chapter presents and describes income inequality and human capital. The first two sections present the economic development followed by a description of the distribution of income and why we have income inequality in the world. Section 2.3 defines human capital and section 2.4 outlines the relationship between income inequality and human capital. Throughout the chapter there will be made references to existing literature and results from previous studies will be presented.

2.1 Economic Development

Looking back at the economic development in the world we have seen an explosion of economic growth over the last two centuries unlike anything in the previous history of the world. In the richest countries, income per capita today is at least 10 times larger than 200 years ago. But the growth in income has not been even around the world. Among the countries that started growing first, including parts of Western Europe and offshoots as the United States and Canada, relatively slow growth, compounding over almost two centuries, was responsible for the change in living standards[1]. Countries, such as Japan, started growing later but more quickly and caught up in terms of income by the end of the 20th century. After World War II, the average rate of growth of world income increased, as the contagion of growth spread to most of the world.

The uneven distribution of growth among countries has led to a vast widening in the income gaps between rich and poor countries. The differences in income among countries are not the only contributor to inequality. In every country some people are better off and some are worse off than average. Thus income inequality in the world is a result of both within-country and between-country inequality. Data show that inequality has increased since 1820 where most of

[1] Wiel (2005)

the increase took place before World War II. The rise in inequality was seven times higher between 1820 and 1950 then between 1950 and 1992. Following 1980, world inequality declined[2]. Data also shows that between-country inequality is the most important source of inequality in the world today. Specifically, between-country inequality explains 60% of overall world inequality. Even though this between-country inequality accounts for the majority of world inequality today, we cannot forget about within-country inequality. This is also an important determinant of the variation in world income. The degree of income inequality within a country may itself be an important determinant of that country's economic success, and thus may affect the level of income. This unequal distribution of income between countries is one of the most important economic facts in the world today. These differences in income are relevant for at least two reasons. The first is that income distribution is regarded as an important determinant of growth and economic development. The other is that the level of income inequality is an indicator of the access to economic opportunities and about the extent to which development is shared by different sectors of the population. Another implication caused by the large gap between rich and poor is the potential for alleviating poverty.

Over the last few decades, we have seen that the role of technology has become more important, poverty-rates have declined all over the world, and emerging market countries have experienced remarkable growth[3]. The world is experiencing a global income inequality at a high level, whereby the richest eight per cent of the world's population earn half of the world's total income, while the remaining 92 per cent are left with the other half[4]. As the world becomes more integrated the dimension of global inequality is likely to become even more relevant. The movements of factors of production across borders will increase, and the influence of other people's standard of living and way of life is becoming of greater influence when people are looking at their own income position and aspirations. The differences in resources such as wealth and education will influence the way people see themselves and others. The high level of inequality is at the same time keeping countries from realizing development outcomes and expanding the opportunities and abilities of people.

The high level of inequality is not only an ethical concern, but it is also important for a country's economic development. Therefore, it is possible to find more than one reason for

---

2 Weil (2005)

3 United Nations Development Programme (2013)

4 Milanovic (2013)

why countries need to focus on closing the gaps. There are many factors, which will have an impact when it comes to addressing income inequality. Therefore, the next sections will introduce the distribution of income and explain why we have income inequality. As the focus of this thesis will be to outline the role of human capital its definition and its relation to income inequality will be presented.

2.2 The Distribution of Income

As stated above, there is more than one reason to pay attention to income inequality; one is its relation to poverty. The more unequal the distribution of income, the more people will live in poverty. Another reason is that the distribution of income is interesting as it is tied to the process of economic growth. The conventional textbook approach is that equality is good for incentives and therefore good for growth, even though incentives and growth considerations might sometimes be traded off against equity or insurance goals. On the other hand, development economists have long expressed counterarguments. The literature on this matter is substantial but inconclusive; some find that a higher level of inequality is good for growth in some stages of development and bad in others. Simon Kuznets was one of the first to hypothesize this theory in 1955. Using both cross-country and time series data, he found an inverted U-shape relationship between income inequality and GDP per capita. This result was interpreted as describing the evolution of income over the transition from rural to industrial economies: income inequality should increase during the early stages of development (due to urbanization and industrialization) and decrease later on (as industries would already attract a large fraction of the rural labour force). The work of Kuznets deals with the question of how the level of income affects income distribution. Kuznets' theory has been empirical tested in the literature, where some have found support of an inverted U-shape relation and others have not[5].

The goal of reducing income inequality is often pursued by government economic policy as income inequality increases socio-political instability creating uncertainty in the politico-environment, which reduces investment. Alesina and Perotti (1995) find that income inequality

[5] See among others Barro (2000), Banerjee and Duflo (2003), Gregorio and Lee (2002)

and investment are inversely related. As investment is a primary source of growth they show an inverse relation between income inequality and growth. As policy uncertainty and threatening property rights has a negative effect on investment it reduces growth. Others, such as Alesina and Rodrik (1993), Alsenia and Rodrik (1994) and Persson and Tabellini (1994) present results that support the hypothesis that income inequality is harmful for growth. The result supports the notion that in more unequal societies, the demand for fiscal redistribution financed by distortionary taxation is higher, causing a lower rate of growth. Aghion, Caroli and Garcia-Penalose (1999) analyse the relationship between inequality and economic growth from two directions. First they look at the effect of inequality on growth, showing that when capital markets are imperfect, there is not necessary a trade-off between equity and efficiency. This explains the negative impact of inequality and the positive effect of redistribution upon growth. In their second part they analyse several mechanisms whereby growth may increase wage inequality, both across and within education cohorts. They find that technical change stands as a crucial factor in explaining the recent upsurge in wage inequality. Barro (2000) provides results from a panel of countries showing little support for an overall relationship between income inequality and the rates of growth and investment. He finds that inequality hinders growth in poor countries but encourages growth in rich countries. The literature presented finds different results but is suggesting that by reducing income inequality is can be beneficial for economic growth. Therefore, we have to look at what drives income inequality.

2.2.1 Why do we have income inequality?

Income inequality exists because people are different from each other in many ways that are relevant for their income. This could be in human capital (both education and health), in the way people live (city vs. countryside), in their ownership of physical capital, in their skills or even in their luck. These differences can be translated into differences in income. To understand the reasons for these differences and the reason why inequality in countries differ, we should think about the distribution of different economic characteristics among a population and about how different characteristics translate into different levels of income. A country might have a high degree of inequality because of a large disparity in characteristics, this could be that only some people in the population have an education and some have no schooling at all. Another reason is that the characteristics might be rewarded differently. Nine years of education might give a significant higher wage then eight years of education. Inequality in a country will differ over time in the same way because it matter how these characteristics are distributed and rewarded.

There are other reasons why we see differences in the income distribution, which are not related to the individual person. We can distinguish between drivers that are largely exogenous (i.e. outside the purview of domestic policy) and the ones that are endogenous (i.e. mainly determined by domestic policy). Though, it can be difficult to draw a clear line because some drivers might seem to be exogenous at first sight, but may be the outcome of policy decisions in the past or the outcome of a political decision to create certain institutions. An example could be the creation of the World Trade Organization, which is created to establish trade liberalization or the decision to invest in technical progress. As we experience a further increased globalization, the exogenous drives gain in importance. Trade and trade openness has mainly been given attention as important drivers, but, more recently, global finance and technical change has also been the focus of much attention[6].

As just presented income inequality exists because people are different from each other, which are relevant for their income but also for reasons that are not related to the individual person. The next section will introduce human capital and explain why this is important in relation to income and income inequality.

2.3 Definition of Human Capital

Human capital is the knowledge, competencies, values, and social and personal attributes that are represented in the ability to perform labour so as to produce economic value. In other words, it can be defined as a measure of the economic value of an employee’s skill set. Some of this is acquired through investment in education. The amount of investment a person places in education will have an impact on the income that the person receives after his education.

[6] United Nations Development Programme (2013)

Human capital can also be in the form of health. As a country develops economically, the health of its population improves. This improvement in health can be seen as direct evidence that people are leading better lives. Improvement in health will also have a productive side as healthier people can work harder and longer. Also, healthier students can learn better. Therefore, the better the well-being of a country's population is, the higher the income will be[7].

The dimensions of inequality that matter for human well-being can be looked at with two perspectives; inequality of outcomes and inequality of opportunities. Inequality of outcomes can be the level of income or the level of educational attainment and inequality of opportunities can be such as unequal access to employment and education. Unequal outcomes, particularly income inequality, are argued to play a key role in determining variations in human well-being. This is made evident by the strong association between income inequality and inequalities in health, education, and nutrition[8]. If a high income provides people with opportunities to secure their well-being and to get ahead in life, then the income each person has will matter. Having a meaningful equality of opportunity, income inequality needs to be moderate so that people start their lives from roughly equal starting points. The perspective of inequality of opportunities looks at the fact that certain individuals and groups face consistently inferior opportunities – economic, political and social – then their fellow citizens. It is argued that individuals can hardly be held responsible for the circumstances of their birth: their race, sex or urban or rural locations. Yet this make a difference for the lives they lead. Not surprisingly, unequal opportunities lead to unequal outcomes[9]. The two perspectives differ when it comes to the causality between outcomes and opportunities. Will higher incomes lead to improved opportunities or will greater opportunity lead to improved outcome in human well-being? The two perspectives are highly interdependent, therefore, equal outcomes cannot be achieved without equal opportunities, but equal opportunities cannot be achieved when households have unequal starting points. The perspective presented should help understand that even though people are offered the same opportunities the outcome is not likely to be the same for each individual. The unequal distribution of income matters as people will have different starting points which can distorts the effect coming from improved opportunities. Therefore, more education is not likely to have the same effect for all individuals.

[7] Weil (2005)
[8] WHO (2008)
[9] World Bank (2006)

As stated in the beginning, human capital is also the competencies and knowledge that are embodied in a person. People work with their minds as well with their bodies. Indeed, in developed economies, intellectual ability is far more important than physical ability in determining a person's wage. Therefore, investment in improving people's intellect has become the most important form of investment in human capital. To measure the return on our investment in human capital is a bit complicated because human capital is always attached to its owner. We cannot separate a part of a person's education from the rest of his body and see how much it gets paid. To get around this, economists infer the return to human capital from data on people's wage. The fact that people with higher education earn a higher wage can be taken as evidence that the market values their human capital. Therefore, as a person receives one more year of schooling it will result in an increase in that persons wage, which is defined as the return to education.

2.4 Human Capital and Income Inequality

So why is human capital important when we look at income inequality? First is the contribution to having a good life, as described in the previous sections. Second is its relation to economic growth and the distribution of income. In the theory of human capital, Gary Becker[10] showed that acquiring education increases the skills and competencies of individuals and their productivity. Since in a competitive labor market wages equal workers' productivity, a higher productivity will therefore lead to higher wage. This means that a more educated society holds greater welfare. Since the conception of this theory, it has been the focus of increasing research. It has encouraged the production of many empirical and theoretical studies. The acknowledged of a causal relation between education and earning is now a well-established result, but it is less clear-cut when analysing the link between income inequality and educational attainments.

On one hand, rising wage inequality should encourage investments in education mainly because it raises the return to education. Topel (1997) observes a faster skill accumulation as a

[10] Becker (1962)

result of rising returns. This increase in the supply of skills should eventually mitigate the increase in inequality.

On the other hand, when income inequality is increasing it also affects the resources that households have available to finance education. The intergenerational theory claims that there exists a perfect correlation between income and education distributions. This causes that barriers, e.g. liquidity constraints and family background, might prevent the investment in education for that part of the population belonging to the bottom of the income distribution. If this is persistent then the same part of the population will be trapped in low levels of education and income for more then one generation.

The accumulation of human capital has also been shown to be essential for economic growth, and favourable for individuals and societies[11]. The positive effect on the individual is that the more educated people are, the better labour market term in form of wage and employability people will have. Lochner and Moretti (2004) find that other positive effects will be better health, fertility, well-being and, lower chance of engaging in crime.

In the presence of imperfect credit markets Galor and Zeira (1993) show that the wealth distribution affects investments in human capital. By developing an overlapping generation model with intergenerational transmissions, they suggest that the initial distribution of wealth is crucial to determine individuals' education choices and the aggregated output in both the short and the long run. Banerjee and Newman (1993) end up with similar conclusions. Their theoretical model suggests that the pattern of occupational (educational) choice is shaped by the initial distribution of wealth. Filmer and Pritchett (1999) preform an empirical analysis using household surveys for 35 countries, where they use the poverty index as their proxy for economic status of the household. They find that the poverty index is correlated with reduced school attainment in the poorest 40 per cent of the population. Checchi (2003) analyse the issue by using an unbalanced panel for 108 countries for the period 1960-1995. His main finding is a robust negative correlation between income inequality and secondary education enrolment. The effect is stronger when looking at females' access to any level of education. This further supports the result that families are prevented from accessing school when they have low income. The literature just presented lack in properly addressing the endogeneity of the

[11] See smong others Hanushek and Kimko (2000), Krueger and Lindahl (2001), de la Fuente and Domenech (2006)

inequality variable, that is, when other omitted factors are correlated with both the education and inequality measures or when the causation goes the other way around.

Another group of studies is concerned with the effect of family income on children' education outcomes. The idea behind this type of research is that rich parents can spend more or have unconstrained access to credit on their children's education than poor parents and these investments can lead to better outcome for their children. This hypothesis has not found clear empirical evidence in the literature. The effect of parent's income on children's educational attainment has been found moderate or non-existing. Though, it is worth noting that these studies have dealt with the endogeneity of the income variable in the education equation. The income variable is endogenous since other factors, such as parent's schooling and parents' ability might determine both family income and children's outcome[12].

Gregorio and Lee (2002) present empirical evidence on how education is related to the income distribution in a panel data analysis. Their findings indicate that educational factors – higher education attainment and more equal distribution of education – play a significant role in making income distribution more equal. Castelló-Climent and Doménech (2014) find that in spite of a large reduction in human capital inequality around the world, the inequality in the distribution of income has hardly changed. They find evidence that increasing returns to education and exogenous forces such as skill-biased technological progress or globalization have offset the effect of the fall in educational inequality, which is explaining the low correlation between the changes in income and education inequality.

Knight and Sabot (1983) highlight the complicated effect of human capital accumulation on income distribution due to "composition" and "wage compression" in a dual economy. They argue that an expansion of education has two different effects on the earnings distribution. The "composition" effect increases the relative size of the group with more education and tends initially to raise income inequality, but eventually to lower it. The "wage composition" effect decreases the premium on education as the relative supply of educated workers increases, thereby lowering income inequality. Consequently, the effect of increased education on the dispersion of income is ambiguous.

[12] See among other (Acemoglu and Pischke 2001) , (Akee, Copeland, Keeler, Angold and Costello 2010) , (Cameron and Heckman 2001)

The empirical studies presented help to understand the link between human capital and income inequality. This study has acknowledged the results presented in the literature and will be focusing on addressing the endogeneity problem in order to successfully investigate the relationship between income inequality and educational attainment.

2.5 Partial Conclusion

This chapter has introduced and described the distribution of income, why we have income inequality in the world and the relationship between human capital and income inequality. Income inequality has been presented as being important for growth and economic development, but also for individuals and societies. Income inequality exists because people are different from each other in many ways that are relevant for their income. These differences in income are affecting the way people are living and can affect the resources that people have available to finance education. Therefore, when looking at income inequality, human capital is important as it is contributing to the population's standard of living and to what opportunities that might be available for them. Human capitals positive effect has been widely recognized in the literature and has been shown to be essential for economic growth and for the fact that more educated people will have better labour market terms. The chapter has given the reader an understanding and knowledge of the importance of human capital and income inequality and its relationship. This will be the basis for further empirical investigation.

Chapter 3

3. Methodology

This chapter presents the methodology used in the empirical investigations in chapter 5. Section 3.1 presents panel data and section 3.2 presents the methodology used to analysis the effect of educational attainment on income inequality. In this section the different econometric models, OLS, instrumental variable, and fixed effects are also presented.

3.1 Panel Data

The purpose of this study is to investigate the effect of educational attainment on income inequality by using panel data. Panel data contains observations of multiple occurrences obtained over multiple time periods for the same individuals or, in this case, countries. With panel data you are able to examine the data across and within countries over time. Previous cross-national studies[13] are very much hampered by a lack of internationally comparable data, and they therefore end up with a few data points from heterogeneous sources. When using cross-section data it is collected by observing many subjects (such as individual, firms, regions or countries) at the same point in time or without regard to differences in time. Therefore, this study utilizes newly assembled panel data on income inequality and educational attainment for a broader number of countries measured at five-year intervals from 1960 to 2010. Below the advantages and limitations of panel data will be presented.

Advantages

Panel data have several advantages, which are beneficial for this study. Baltagi (2008) lists some of the advantages. First it is possible to control for individual heterogeneity. Panel data suggest that individuals, firms, states, or countries are heterogeneous. Times-series and cross-

[13] Se among others Ram (1984), Chiswick (1971), Sylwester (2002)

section studies that are not controlling this heterogeneity run the risk of obtaining biased results. In this case when analysing the effect of education attainment on income inequality, there is a lot of other variables that may be country-invariant or time-invariant that may affect income inequality. Panel data is able to control for these country- and time-invariant variables. By combining time series and cross-section observations, panel data gives more informative data. This study will be able to take advantage of having information on income inequality and educational attainment on each country over a time period.

Panel data is also better when you want to study the dynamics of adjustment. Cross-sectional distributions that look relatively stable hide a multitude of changes. For example in measuring income, cross-section data can estimate what proportion of the population is poor at a point in time. Repeated cross-sections can show how the proportion changes over time. Only panel data can estimate what proportion of those poor in one period also remains poor in the next period. In that way panel data are better able to identify and measure effects that are simply not detectable in pure cross-section or pure time-series data.

By making data available for several thousand units, panel data can minimize the bias that might result if we aggregate individuals or firms into broad aggregates. So with the data used in this study we are able to minimize bias that would have existed if the data only had one observation for income inequality and educational attainment, respectively, at the different points in time.

Limitations

Panel data also have some limitations, which can be design and data collection problems. The problem in this study could include coverage (incomplete account of the population of interest). Also there could be distortions of measurement errors and selectivity problems. For example, people can choose not to work because the reservation wage is higher than the offered wage.

It can be a limitation if you only have short time-series dimension and if you have cross-section dependence. Macro panel on countries or regions with long time series that do not account for cross-country dependence may lead to misleading inference.

The data set used in this study is an unbalanced panel dataset, meaning that each country does not have data for all years for both the income inequality data and for the educational attainment data. By using panel data this study has a dataset of countries over time, and thus multiple observations on each country in the sample. Therefore, the notations in panel data are including *i* for the cross section unit (in this case each country) and *t* for time.

The data will be presented and discussed in chapter 4.

3.2 The Effect of Educational Attainment on Income Inequality

3.2.1 OLS Estimation

To estimate the effect of educational attainment on income inequality the study starts by using Ordinary Least Squares (OLS). The following models are estimated.

$$\text{Gini\_net}_{it} = \beta_0 + \beta_1 \text{yr\_sch}_{it} + \mu_{it} \qquad \text{(model 1)}$$

$$\text{Gini\_net}_{it} = \beta_0 + \beta_1 \text{yr\_sch\_pri}_{it} + \beta_2 \text{yr\_sch\_sec}_{it} + \beta_3 \text{yr\_sch\_ter}_{it} + \mu_{it} \qquad \text{(model 2)}$$

where *i* denoting the countries and *t* denoting the time. The *i* subscript, therefore, denotes the cross-section dimension whereas *t* denotes the time-series dimension. The Gini net coefficient (Gini_net) is the dependent variable, β_0 is a constant, and education presented as the average years of schooling for the population 15 years and above (yr_sch) in model 1 and as the average years of schooling for the population 15 years and above for each level of education, primary (yr_sch_pri), secondary (yr_sch_sec) and tertiary education (yr_sec_ter) in model 2 is the independent variable, and μ_{it} is the error term that varies across countries and across time.

The study starts by using OLS to investigate the effect of education on income inequality[14]. A problem with this model is that it does not distinguish between the various countries nor does it tell us whether the response of the Gini coefficient to the explanatory variable over time is the

[14] Among other studies using OLS are Castelló-Climent and Doménech (2014), Cohon and Soto (2007)

same for each country. In other words, by grouping the countries together at different times the model will camouflage the heterogeneity that may exist among countries. Therefore, it is quite possible that the error term may be correlated with the explanatory variables in the model. If that is the case, the estimated coefficients may be biased as well as inconsistent.

A way to manage country specific effects or the effect of time is to include dummy variables in the regression. A dummy variable is a numeric value that can take the value 0 or 1 to indicate the absence or presence of some categorical effect that may be expected to shift the outcome. Time dummies are included throughout the empirical analysis to investigate if the effect of educational attainment on income inequality is the same over time. It is also possible to use a fixed effects model when you are worried about time-invariant unobservable factors that might be correlated with the variables that are included in the regression.

3.2.2 The Fixed Effects Model

The fixed effects model is used whenever you are only interested in analysing the impact of variables that vary over time. The fixed effects model explores the relationship between the predictor and outcome variables within an entity. Each entity has its own individual characteristics that may or may not influence the predictor variables. When using the fixed effects model we are assuming that something within the individual country may impact or bias the variables and we need to control for this. Another important assumption of the fixed effects model is that the time-invariant characteristics are unique to the individual country and should not be correlated with other individual characteristics. Each entity is different. Therefore, the entity's error term and the constant should not be correlated with others. For the fixed effects model the following model is estimated.

$$\text{Gini\_net}_{it} = \beta_0 + \beta_1 \text{yr\_sch}_{it} + \gamma_i + \delta_t + \mu_{it} \qquad \text{(model 3)}$$

Model 3 uses the same variables as model 1. The differences are the term $\gamma_{i,}$ which stands for specific characteristics in every country that are constant over time, and the term $\delta_{t,}$ which is a time-specific effect.

As described under section 3.2.1, the study starts by assuming that $\gamma_i = 0$. The advantage of this assumption is that we can use econometric techniques that exploit the whole cross-country variation in the data. However, the omission of time invariant country-specific characteristics may bias the estimated coefficients. To account for this the model is estimated by assuming $\gamma_i \neq 0$.

The variables included in the OLS regression can be subject to potential bias that may come from several sources. As stated above, one source is omitted variables bias as it is plausible that some important factors that are not included as explanatory variables can influence both income inequality and education simultaneously. Using a fixed effect model we can adjust for an unobserved effect that is correlated with the covariates. If an omitted variable varies by country, but is constant over time, the inclusion of a country-fixed-effect term eliminates this source of endogeneity bias. The fixed effects coefficients soak up all the across-country action, and leave the within-country action. Like other models, the fixed effects model has some limitations, which will be discussed in section 3.3.

3.2.3 Instrumental Variables

Another potential source that can lead to a biased estimate comes from reverse causality. In this case it might be that people decide to invest more in education when their income is high. Therefore, the positive effect that education has on output may reflect this reverse causality. In principal, this can be handled with instruments. The instrumental variable estimation also addresses the problem of omitted variable that is correlated with education but cannot be included in the regression because it is unobserved. In this study the regression may suffer from a bias as to omitted ability. The fact that ability is unobserved and can have an effect on the education a person receives can cause the OLS estimates to be biased. The use of instrumental variables will help in addressing these biases, but you need to find valid instruments, which can be difficult.

An instrumental variable, Z, is uncorrelated with the error term but is correlated with your explanatory variables. With this new variable included, the IV estimator should capture only the effect on the dependent variable of shifts in the explanatory variables induced by Z whereas the OLS estimator captures not only the direct effect of the explanatory variables on the dependent variable but also the effect of the included measurement error and/or endogeneity. The IV estimation is not as efficient as OLS, especially if you have “weak-instruments”[15]. In order for a variable, Z, to serve as a valid instrument for the explanatory variables, the following must be true.

- The instrument must be exogenous, that is, $Cov(z,\mu) = 0$
- The instrument must be correlated with the endogenous explanatory variable, that is, $Cov(z,x) \neq 0$

As stated before, one problem with the IV estimation can be to find a good instrument. The same applies for education. This study will follow the approach used by Barro and Lee (2013), who adopt the methodology developed in the micro-literature and uses parents’ education as an instrument for the education variable. In order to use parents’ education as a valid instrument, it cannot be correlated with the error term and it needs to be correlated with the education variable. Barro and Lee (2013) support their choice of instrument by saying that in the population 15 and above the contemporaneous educational attainment includes a portion of educational attainment of the younger generation (e.g. the group aged 15-25 years), which may be correlated with the current income. But, considering that the past education attainment for parents’ generation was accumulated by their past investment in education it can be uncorrelated with the error term. This assumption cannot be directly tested, as we do not have an unbiased estimator for the error term. Therefore, you need to use common sense and economic theory to decide if the assumption holds true. The assumption that states that the explanatory variable and the instrument need to be correlated can be tested. By testing if $p_1 = 0$ in the regression: $yr\_sch = p_0 + p_1 Parents\_Education + v$, we can verify the assumption. The results are presented in the appendix and show that $p_1 = 0$ is rejected and the assumption holds true.

It is also possible to perform tests of endogeneity. A natural question to ask is whether a variable presumed to be endogenous in the model could instead be treated as exogenous. If the endogenous regressors are in fact exogenous, then the OLS estimates are more efficient then IV estimates. Thus, unless the instrumental-variables estimator is really needed, OLS should be used instead.

[15] A weak instrument occurs when the IV estimator is very little correlated with the explanatory variable

In the IV regression we treat average years of schooling (yr_sch) as endogenous because it is likely that other factors will affect income inequality as well. So we use parents' education as an instrument and test whether we can treat yr_sch as exogenous. The test is performed using the Durbin and Wu-Hausman test. The test evaluates the significance of an estimator versus another estimator. The test will help one to evaluate if the statistical model corresponds to the data. The null hypothesis of the Durbin and Wu-Hausmann test is that the variable under consideration can be treated as exogenous. The result of the test can be found in the appendix, which shows that both test statistics are highly significant, so we reject the null hypothesis of exogeneity. Therefore, we must continue to treat yr_sch as endogenous.

Specifically, following the method presented by Barro and Lee (2013), this study takes the 10 year-lag of average years of schooling among the population of 40 years and over (40-75 years old) to represent parents' education and use it to instrument for the average years of the education variable.

The two-least square estimation

The instrumental variable regression is performed using the two-stage least squares estimation. This contains two stages; in the first stage the average years of schooling is regressed on the instrumental variable that makes up the element of Z using OLS. This first stage regression, which is often called a "reduced form equation" is:

$$Yr\_sch_{it} = \beta_0 + Z_{it}\beta_1 + \mu_{it}$$

The OLS coefficient estimate from this first-stage regression is used to form fitted values, $\widehat{yr\_sch}$, for the education variable. In the second stage of the two-stage least square, the fitted value for the education variable is substituted for the actual values of the education variable in an OLS regression.

$$Gini\_net_{it} = \beta_0 + \beta_1\widehat{yr\_sch} + \mu_{it}$$

The second stage coefficient is the two-stage least square estimate.

In general, a parameter is said to be identified if different values of the parameter would produce different distributions of data. In the instrumental variables estimation the variables are identified depending on the relationship between the number of instruments and the number of endogenous regressors. The coefficients are said to be exactly identified when the number of instruments are equal to the number of regressors. The coefficients are overidentified if the number of instruments are higher then the number of regressors, and underidentified if the number of instruments are smaller then the number of regressors. The IV estimation used in this study is exactly identified, as the number of instruments is equal to the number of explanatory variables.

3.3 Discussion of the Models

This study uses a fixed effects model to address the problem of omitted variable bias that might vary from country to country but is constant over time. It is worth noting that even though the fixed effect model is used there might also be variables that are not time-invariant and you will still have omitted variable bias. A potential significant limitation of the fixed effects model is that you cannot assess the effect of variables that have little within-group variation. By estimating a fixed effects model, the study control for unobservable heterogeneity but at the expense of the very low within variation in the income Gini coefficient.

The use of instrumental variables addresses the problem of reverse causality that may occur in the regression being estimated. The instrumental variable can be used to avoid the bias that OLS suffers when an explanatory variable in a regression is correlated with the regression's error term. But it should be noted that good instruments can be difficult to find and it is no exception in this study. An instrument can be invalid if the instrument itself is correlated with the error term in the equation of interest. An invalid instrument yields a biased and inconsistent IV estimator that can be even more biased than the corresponding OLS estimator[16]. An instrument can be seen as weak when the IV estimator is very little correlated with the explanatory variable that in practice will not overcome the bias of OLS and yield misleading estimates of statistically significance even with a very large sample size.

[16] Murray (2006)

In this study, by following the approach by Barro and Lee (2013), the results are considered to be valid in suggesting the possible effect of education in income inequality. Hence, this still encourages us to be careful when interpreting the results.

3.4 Partial Conclusion

In this chapter the methodology used in the empirical investigation was introduced and described. The type of data used in the study was introduced together with the advantages and disadvantages of panel data. The OLS, fixed effects and instrumental variable models were presented and discussed. The reasons for choosing the different econometric approaches used in the study were elaborated and discussed.

Chapter 4

4. Data and Descriptions

This chapter describes the data used in the empirical investigation. The first section describes the data from the different sources and the method used to obtain the variables. In section 4.2 a discussing of the data will be presented. In section 4.3 the descriptive statistics and the development in the data will be described.

4.1 Data

This study added to the existing literature by using a new dataset on educational attainment from Barro and Lee (2013), which includes more countries and years, reduces some measurements errors, and solves some of the shortcomings that their previous dataset had. Also the data availability on income inequality has improved in coverage, both in countries and years. Previous cross-national studies are very much hampered by the lack of internationally comparable data, therefore as these dataset have become available for both educational attainment and income inequality this study is able to contribute to the literature with some new results.

Gini coefficient

The Gini coefficient is obtained from the Standardized World Income Inequality Database (SWIID) [17]. The SWIID provides comparable Gini indices of gross and net income inequality for 153 countries for as many years as possible from 1960 to the present along with estimates of uncertainty in these statistics. The Gini coefficients are scaled as in the WIID[18] where the Gini coefficient has a theoretical range from zero, which indicates that each reference unit

[17] Solt (2009)
[18] The World Income Inequality Database. Available online at http://www.wider.unu.edu/research/Database/en GB/database/.

receives an equal share of income, to one hundred, indicating that a single reference unit receives all income and all others receives nothing. The objective is to get greater coverage across countries and over time, which has been hindered by limitations of existing inequality datasets. The SWIID uses a custom missing-data algorithm to standardize the United Nations University's World Income Inequality Database; data collected by the Luxembourg Income Study served as the standard. By maximizing comparability for the largest possible sample of countries and years, the SWIID is better suited to broadly cross-national research on income inequality than previously available sources.

Educational Attainment

The data on educational attainment is obtained from Barro and Lee (2013) . This new panel dataset on educational attainment includes 146 countries from 1950 to 2010. The data are disaggregated by sex and by 5-year age intervals. The new version has improved the accuracy of estimation by using information from consistent census data, disaggregated by age group, along with new estimates of mortality rates and completion rates by age and education level. The estimates of educational attainment provide a reasonable proxy for the stock of human capital for a broad group of countries and should be useful for a variety of empirical work.

Additional variables

In the empirical investigation additional variables are added to the regression. The indicators for population growth and life expectancy are obtained from the World Bank Databank. Life expectancy at birth indicates the number of years a new-born infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life. The population growth is the exponential rate of growth of midyear population from year t-1 to t, expressed as a percentage.

The variable for openness, gross domestic product (GDP), and government consumption is taken from the Penn World Tables. The data being used is from PWT 7.1, which cover 189 countries and territories from 1950-2010 with 2005 as reference year. The Penn World Table (PWT) displays a set of national accounts economic time series covering many countries. Its expenditure entries are denominated in a common set of prices in a common currency so that real quantity comparisons can be made, both between countries and over time. It also provides information about relative prices within and between countries, as well as demographic data and capital stock estimates.

Description of the variables

In table 1 the different variables used in the empirical analysis is presented. The first column presents the variables as they will be presented in the tables under the empirical analysis in chapter 5. The second column provides a description of the different variables.

Variables	Description
Gini_net	Gini coefficient of net disposable household income
yr_sch	Average Years of Total Schooling
yr_sch_pri	Average Years of Primary Schooling
yr_sch_sec	Average Years of Secondary Schooling
yr_sch_ter	Average Years of Tertiary Schooling
lpc	Primary School Completed (% of the population aged 15 and over)
lsc	Secondary School Completed (% of the population aged 15 and over)
lhc	Tertiary School Completed (% of the population aged 15 and over)
life_expect	Life expectancy at birth indicates the number of years a new-born infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life
pop_growth	Population growth (annual %) is the exponential rate of growth of midyear population from year t-1 to t, expressed as a percentage.
rgdpl	PPP Converted GDP Per Capita (Laspeyres), derived from growth rates of consumption, government consumption, investments, at 2005 constant prices
openc	Openness at 2005 Constant Prices (%). Calculated as export plus import divided by GDP [rgdpl]
kg	Government Consumption Share of PPP Converted GDP Per Capita at 2005 constant prices [rgdpl]

Table 1 – Description of the variables

4.1.1 The Gini Coefficient as the measure of Income Inequality

There are different ways to measure inequality in a country. A measure that is frequently used is the Gini coefficient. The Gini coefficient measures the extent to which a frequency distribution (in this case levels of income) among individuals or households within an economy deviates from a perfectly equal distribution. The Gini coefficient can theoretically range from 0 (complete equality) to 1 (complete inequality). It is sometimes expressed as a percentage ranging between 0 and 100. In practice both extreme values are not quite reached. If negative values are possible, such as negative wealth of people with debts, then the Gini coefficient could theoretically be more than 1. Normally the mean is assumed positive, which rules out a Gini coefficient less than zero. Therefore, a low Gini coefficient indicates a more equal distribution, while a high Gini coefficient indicates a more unequal distribution.

The Gini coefficient is derived from a Lorenz curve, which plots the cumulative percentages of total income received against the cumulative numbers of recipients, starting with the poorest individual or household. The Lorenz curve plots the proportion of the total income of the population (y axis) that is cumulatively earned by the bottom x% of the population. This is presented in figure 1. The 45 degree line in figure 1 represents perfect equality of incomes. The Gini coefficient can then be thought of as the ratio of the area that lies between the line of perfect equality and the Lorenz curve (marked A in figure 1) over the total area under the line of equality, which is marked A and B in the figure. The calculation of the Gini coefficient can therefore be expressed as:

Gini = A / (A+B)

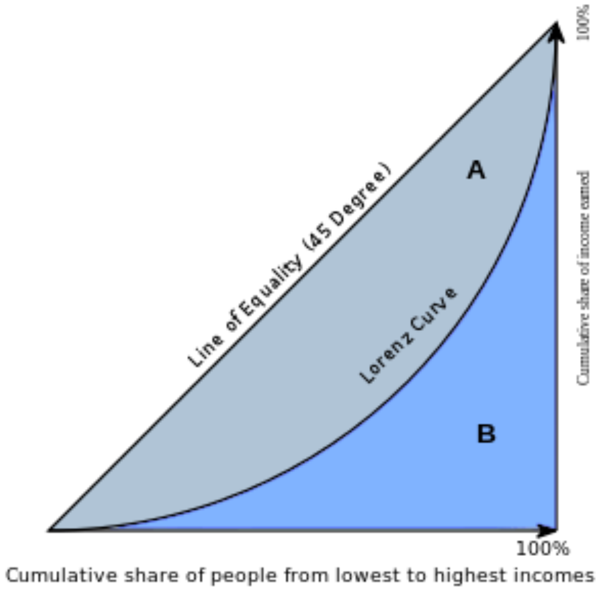

Figure 1 – The Gini coefficient presented graphically

4.2 Discussion of the Data

The data used in this study is the newest and most comprehensive dataset available for world income inequality and educational attainment. In these new versions the quality of the data has been improved, but it remains important to be careful when interpreting the results found in the empirical investigation. The data on educational attainment has been criticised by Cohen and Soto (2007) and Fuente and Domenech (2006), who found that previous data set of Barro and Lee (1993, 2001) showed implausible time-series profiles of education attainment for some countries. This has been resolved in the new data set by Barro and Lee (2013) used in this study. The data on income inequality are providing a good coverage of as many countries as possible and it allows this study to be able to explore the econometric approaches, which make the results more reliable.

As stated in the beginning of this chapter, previous cross-national studies are very much hampered by the lack of internationally comparable data. The data on education and income inequality has always been questioned, and even though the data used in this study has improved, it can still be questioned. The data used is only covering few time periods, as the data on educational attainment is divided in five years intervals, which also have some limitations.

The purpose of this study is not to provide a detailed discussing of how the data from the different sources are obtained, but to investigate the relationship between data on educational attainment and income inequality just described, which intensify the importance of taking these uncertainties into account when the results are interpreted.

4.3 Descriptive Statistics

In this section the descriptive statistics are presented to quantitatively describe the main features of the data used in the study. The descriptive statistics are shown in table 1 and include the mean, the standard deviation, the minimum observation, the maximum observation and the number of observations for all the variables presented in the study.

The statistics show that the Gini coefficient on average is 38.23 and have a standard deviation of 11.25. The lowest Gini coefficient at 15.79 belongs to Mauritius and the highest at 76.62 to Panama. The table shows a sizeable difference in the Gini coefficients, which confirm a large between-country inequality in the world.

The average years of schooling for the population 15 years old and above also have a big difference between the minimum and maximum observation. When looking at the different levels of schooling the data show that the average years of primary school is 4.55 years, the average years of secondary school is 2.26 years, and for tertiary school the average is 0.31 years. On average approximately 20 per cent have completed primary school, 19 per cent have completed secondary school and 5 per cent have completed tertiary school. The development in education will be further elaborated in section 4.5.

Life expectancy is on average approximately 67 years with a standard deviation of 10. The standard deviation is quite high indicating that the distribution of life expectancy in the world is widely dispersed.

Table 1

Descriptive Statistics

	Mean	Standard Deviation	Minimum	Maximum	Number of observations
All Countries					
Gini_net	38.23	11.25	15.79	76.62	784
yr_sch	7.12	2.87	0.38	13.09	784
yr_sch_pri	4.55	1.64	0.29	8.83	784
yr_sch_sec	2.26	1.40	0.30	7.48	784
yr_sch_ter	0.31	0.28	0.004	1.56	784
lpc	20.43	12.23	0.69	81.30	784
lsc	18.95	13.75	0.03	69.75	784
lhc	5.54	5.22	0.06	26.36	784
life_expect	67.49	9.99	31.24	82.82	774
pop_growth	1.41	1.15	-2,.5	6.68	774
rgdpl*	3.600.000	2.620.000	18.300	9.980.000	747
kg*	4.700.000	3.050.000	82	9.960.000	747
openk*	6.170.000	2.330.000	6.662	9.970.000	747

Note: the values are presented in thousand

Table 1 shows a high standard deviation for the GDP and government consumption which indicates a large differences between the countries, which would be expected when looking at the whole world. The data on government consumption vary between the minimum (82) and maximum (9.960.000) observation. Therefore, it is very different how much countries use in government consumption.

The data on income inequality and school attainment will be further discussed in the next two sections, which will contribute to the empirical investigation presented in chapter 5.

4.4 The development in Income Inequality

Table 2 presents a summary of data for the Gini coefficient and the average years of schooling for the years 1960 and 2010 for all countries and for the different regions in the world. The countries are divided into seven regions[19], which include the Advanced Economies (AVE) 25 countries, East Asia and the Pacific (EAP) 20 countries, Europe and Central Asia (ECA) 22 countries, Latin America and the Caribbean (LAC) 26 countries, Middle East and North Africa (MENA) 20 countries, South Asia (SA) 7 countries and, Sub-Saharan Africa (SSA) 35 countries. The data has been shown for the different regions to get a better picture of how income inequality is divided in the world. The table shows that the Gini coefficient has been fairly stable over time. In 1960 the average Gini was 38 compared to 36 in 2010. Overall the data shows that the standard deviation of the Gini coefficient is lower in 2010 than in 1960, indicating that the dispersion in the Gini coefficient has decreased.

The regions with the lowest inequality are the Advanced Economies and Europe and Central Asia with a Gini of 29 and 28 in 1960 and a Gini of 30 in 2010. Latin America and the Caribbean, Sub-Saharan Africa and South Asia are the regions with highest inequality were the Gini coefficient is ranging from 50 to 55. Middle East and North Africa is the region where inequality has improved the most. In 1960 the Gini was 55 compared to 2010 where it was 36. However, one has to be careful when interpreting these results, since we in the whole world only had 23 countries with data available in 1960, whereas we had 89 in 2010. Therefore, new

[19] The classification of the regions are taken from the Barro and Lee (2013) dataset.

data, which is becoming available, might tend to come from countries with greater inequality then the previous average.

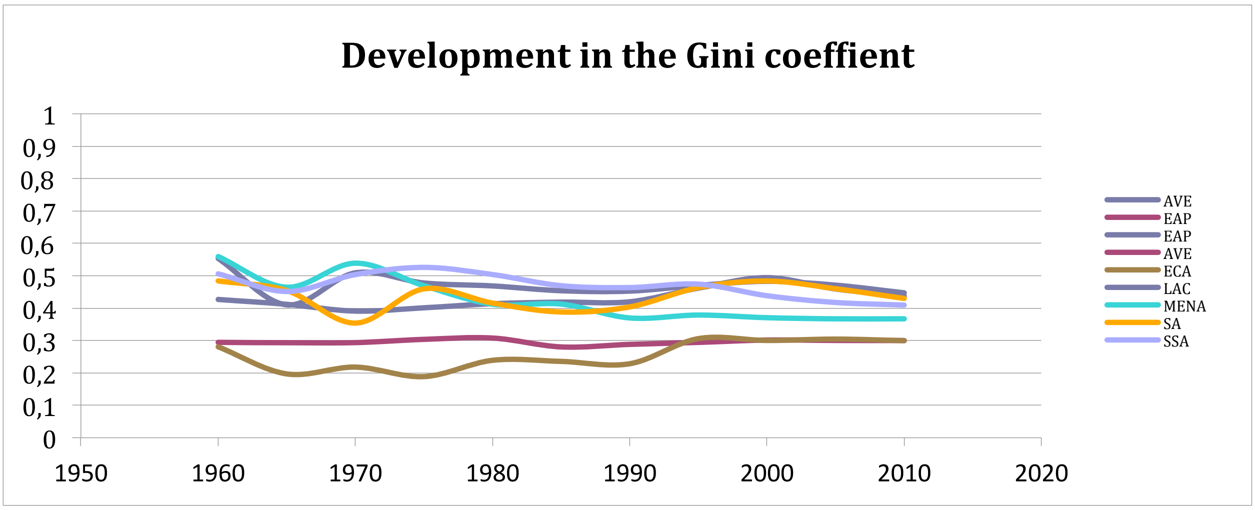

Figure 2 – The development in the Gini coefficient. The horizontal axis presents the years and the vertical axis presents the Gini coefficient ranging from 0 to 1. The regions are the Advanced Economies (AVE), East Asia and the Pacific (EAP), Europe and Central Asia (ECA), Latin America and the Caribbean (LAC, Middle East and North Africa (MENA), South Asia (SA) and, Sub-Saharan Africa (SSA).

In figure 2 the statistics are shown graphically for each region. In the beginning of the period we see some jumps, which can be misleading. As just described the data from the beginning of the period do not cover the same number of countries as in the end of the period. For some regions the years only contain a few number of countries that are not representing all the regions variation. However, the graph still gives a useful graphical picture of the development of the Gini coefficient.

Table 2

Summary of data

	Average yr_sch 1960	Average yr_sch 2010	Gini 1960	Gini 2010
All Countries				
Mean	4.88	9.17	39.55	36.36
Standard deviation	2.97	2.31	14.23	8.86
Maximum	10.16	13.09	76.62	59.25
Minimum	0.38	2.38	26.16	23.09
SSA				
Mean	1.68	5.38	50.66	40.95
Standard deviation	1.16	2.03	13.58	11.30
*SA**				
Mean	2.41	6.64	35.39	43.00
Standard deviation	2.39	3.08	8.99	10.08
*MENA**				
Mean	2.90	9.48	53.89	36.71
Standard deviation	2.06	1.43	10.08	7.94
LAC				
Mean	3.94	8.49	55.37	44.74
Standard deviation	1.65	0.96	18.79	3.68
ECA				
Mean	6.19	10.63	28.08	29.97
Standard deviation	0.22	0.94	1.55	4.15
EAP				
Mean	3.25	8.39	42.68	43.74
Standard deviation	0.58	2.17	14.73	7.97
AVE				
Mean	7.38	10.67	29.42	30.01
Standard deviation	1.88	1.41	2.65	4.20

Note: * The first period is 1970 because of data availability

4.5 The development in Educational Attainment

Table 2 also presents the data on educational attainment, which shows that the development in average years of schooling has increased from 1960 to 2010. The world population aged 15 and above is in 2010 estimated to have an average of 9.2 years of schooling, steadily increasing from 4.9 years in 1960. The table shows that the average years of schooling have been increasing in all regions. The Advanced Economies and Europe and Central Asia are not only the regions with the lowest Gini coefficient but also the regions were the population is attending school for the highest number of years. These high-income countries are estimated to have 11 years of schooling, compared to 7.1 years in low-income countries. Since 1950, the average years of schooling in developing countries increased significantly from 2.1 years to 7.1 years. In South Asia and Middle East and North Africa regions, average years of schooling have more than doubled since the 1980s.

In the advanced countries most of the improvement in years of schooling comes from higher secondary and tertiary completion and enrolment ratios, in developing countries the improvements are accounted for by higher primary and secondary completion and enrolment ratios. Developing countries have successfully reduced illiteracy rates, especially among the younger cohorts. The proportion of the uneducated in the total population over age 15 in developing countries has declined significantly over the past six decades since 1990, from 64.9% in 1950 to 20.1% in 2010[20].

Table 2 shows that the standard deviation of educational attainment in almost all regions has increased. This indicates that the dispersion of schooling has increased, when looking at the different regions.

Although the data are suggestive, further statistical analysis is required to examine their robustness and obtain order of magnitude for the importance of educational factors in explaining differences in income distribution across countries. This is presented in the next chapter.

[20] Barro and Lee (2013)

4.6 Partial Conclusion

In this chapter the data used in the empirical investigations has been presented. The different sources of data have been introduced and discussed. The data used has been improved in coverage allowing this study to contribute to the literature with new results. The data do have some limitations and shortcomings, which have been discussed, and it has been addressed how to carefully handle them.

Chapter 5

5. Empirical results

This chapter presents and analyses the empirical results based on the methodology in chapter 3. Section 5.1 presents the results found when testing the effect of educational attainment on income inequality using ordinary least squares. Section 5.2 presents the fixed effects model and the instrumental variable estimation. In section 5.3 additional variables are added to the regression and the results are presented using ordinary least squares, instrumental variables and the fixed effects model.

5.1 The Effect of Educational Attainment on Income Inequality

5.1.1 Ordinary Least Square estimation

The empirical analysis starts by investigating the effect of educational attainment on income inequality using OLS estimation. Table 3 shows the results for model 1 and 2, which were presented in chapter 3. The results suggest that as the average years of schooling increase, the income inequality will decrease. These results are supporting the positive effect from human capital, which has been recognised in the literature. The coefficients of the education variable in column 1 of table 3 show that improved educational attainment has a positive and statistically significant impact on income inequality. Column 1 and 2 present the estimate of average years of schooling with and without time dummies. Column 3 and 4 present the estimates of the different levels of schooling; primary, secondary, and tertiary with and without time dummies.

Table 3

Dependent variable: Gini Net Coefficient					
	OLS				
	1	2	3	4	5
yr_sch	-2.04***	-2.33***			
	(0.11)	(0.12)			
yr_sch_pri			-1.98***	-1.95***	
			(0.25)	(0.24)	
yr_sch_sec			-2.95***	-3.52***	
			(0.33)	(0.38)	
yr_sch_ter			2.58	1.51	
			(1.69)	(1.92)	
lpc					-0.27***
					(0.03)
lsc					-0.33***
					(0.03)
lhc					-0.39***
					(0.08)
Constant	52.77***	50.94***	53.13***	50.41***	52.20***
	(0.90)	(2.35)	(1.03)	(2.45)	(0.94)
R^2	0.27	0.33	0.28	0.34	0.28
Obs.	784	784	784	784	784
Country dummies	*No*	*No*	*No*	*No*	*No*
Time dummies	*No*	*Yes*	*No*	*Yes*	*No*

*Notes: Robust standard errors in parenthesis, ***,** and * are 1,5, and 10 per cent significance level. The dependent variable is the Gini net coefficient. The explanatory variables are average year of schooling, average years of schooling for each level of schooling, primary, secondary and university education of the population 15 and above. The different levels of schooling are also shown as schooling completed for the population 15 and above.*

All the estimates presented in table 3 are significant at a one per cent level, except the coefficient for tertiary education. The result in column 1 shows that as the average years of schooling is increased with an additional year the income inequality is suggested to decrease by 2.04 per cent. Looking at the data in year 2000 it would mean that Norway with a Gini coefficient of 24 would decrease to the level of Denmark, which have a Gini coefficient of 22. The same would apply for Guatemala with a Gini coefficient of 50, which would decrease to the level of Ghana with a Gini coefficient of 48 in 2000. Column 3 presents the results for the different levels of schooling; primary, secondary and tertiary schooling, which indicates that for one additional year of primary or secondary year of schooling the income inequality is decreased. An additional year of tertiary education has the opposite effect causing income inequality to increase, which makes sense as a part of the population, that may not be as big as the part that receives primary and secondary education, receives a higher degree and therefore increases income inequality. The fact that people who have a higher level of education can earn higher wages will encourage people to invest more in education. The increased investment in education will increase ones wages, thus ones income. The descriptive statistics in section 4.3 showed that on average only about 5 per cent complete tertiary school compared to about 20 per cent completing primary and secondary school.

In column 5 the estimates for school completion at each level is shown. All the coefficients are negative and highly significant indicating that income inequality is decreased when a larger percentage of the population is completing school.

The regression has also been estimated using one and two year lag. The coefficient for primary school is slightly decreased and the coefficients for secondary and tertiary school are increased. All results are highly significant, but because this does not change the results significantly, the results are not reported.

A problem with the model estimated in column 1 and 3 is that it does not distinguish between the various countries nor does it tell us whether the response of the Gini coefficient to the explanatory variable over time is the same for each country. In other words, by grouping the countries together at different times the model will camouflage the heterogeneity that may exist among countries. Therefore, it is quite possible that the error term may be correlated with the regressors in the model. If that is the case, the estimated coefficients may be biased as well as inconsistent. Therefore, the study starts by including time dummies into the regression. In column 2 and 4 the regression is estimated with time dummies as this might explain some of the variance in the Gini coefficient. When including time dummies in column 2 the coefficient for education is increased by 0.29 and remains statistically significant, indicating that when controlling for the time dimension the effect from an additional year of schooling is decreasing income inequality by 2.33 per cent. I column 4 each level of schooling is estimated with time dummies and we see that the results do not change significantly. The effect of the time dummies should be interpreted with care as the dataset only contains data from 1960 to 2010 in a five-year interval. Therefore, the data might lack in observations to tell if time has an effect on the results. Throughout the empirical analysis the results will be presented both with and without time dummies to investigate the effect of time on the outcome.

The results of the OLS estimation in table 3 show a positive relationship between human capital, which is presented as educational attainment, and income inequality. The results are highly significant and support the assumption that by improving and investing in human capital, the effect on income inequality is positive. The results presented in table 3 are taking advantage of the whole cross-country variation in the data and show that by holding everything else constant the effect of improved education will contributed to a more equal income distribution. But the results from the OLS estimation can be subject to omitted variable bias and at the same time the results do not tell us if the improved equality in the distribution of income actually comes from improved educational attainment or if educational attainment is improved because income inequality has decreased. Nor do the results in table 3 tell us if the outcome is the same when controlling for the individuality of each country. This will be the focus of the next section.

5.2 The Effect of Education on Income Inequality using Fixed Effect and Instrumental Variables

This section is addressing the problem of endogeneity bias that may exist in the estimation of model 1 and 2. With the use of panel data this empirical investigation is able to control for variables that cannot be observed or measured like cultural factors or variables that change over time but not across countries. This is, it accounts for individual heterogeneity. Section 5.2.1 present model 3 using fixed effect and section 5.2.2 presents the model using instrumental variables.

5.2.1 Fixed effects

The fixed effects model has been estimated to eliminate the possibility of time invariant unobserved effects. The fixed effects explore the relationship between educational attainment and income inequality within each country. In the fixed effects model, the individual-specific effect is a random variable that is allowed to be correlated with the explanatory variables. By estimating the fixed effects model the assumption is that something within the individual country may impact or bias the educational variable and income inequality and we need to control for this. Therefore, the results are adjusted for effects that are country specific and can have biased the OLS estimates.

The results are presented in table 4 and are slightly different from the OLS estimates. The effect on income inequality by increasing the average years of schooling remains positive and the results are significant at a five per cent level. In column 3 of table 4, model 3 is estimated. The results show that the education coefficient is decreased to -0.33 and the coefficient remains statistically significant at a five per cent level. When estimating the fixed effect model we get the within country effect from improved educational attainment. The results confirm that improving educational attainment within a country also has a positive effect on income inequality. Time dummies have been added in column 4. This increases the coefficient for average years of schooling and the significance level is maintained.

In chapter 4 the development of the Gini coefficient was described as having been fairly stable during the time period, which is being analysed. This large stability of the income Gini coefficient in the long term suggests that the variability of the income Gini coefficient is mainly cross-sectional. The R^2 in a regression where the dependent variable is the Gini coefficient and the explanatory variables are country dummies the value is equal to 0.80, and the adjusted R^2 is equal to 0.79. In a similar regression where the explanatory variables are time dummies the R^2 is equal to 0.03 and the adjusted R^2 is equal to 0.02[21]. The R^2 is a statistical measure of how close the data is to the fitted regression line, so it tells how well the data fits the model. The closer to the value 1, the better the fit. An important thing to keep in mind is that the R^2 cannot decrease when more variables are added to the model. So by adding more variables to the model it will increase R^2 but it may also increase the variance of the forecast error. The adjusted R^2 is on the other hand penalizing for adding more regressors. So for comparative purposes, the adjusted R^2 is a better measure than R^2. The results confirm that most of the variability in the income Gini coefficient comes from variability across countries. Thus, econometric techniques that exploit the cross-sectional or the within country variation in the data might give different results.

[21] The results are presented in the appendix

When using the fixed effects model the coefficients soak up all the across-country action, and leave the within-country action. As just described the dataset is experiencing a low variation in the income Gini coefficient. Using the fixed effect model we can control for unobserved heterogeneity, but at the expense of the very low within-country variation in the income Gini coefficient. This is one side effect of the features of fixed effects models; it cannot be used to investigate time-invariant causes of the dependent variable. Substantially, fixed effects models are designed to study the cause of changes within a person or entity. A time invariant characteristic cannot cause such change, because it is constant for each person. This might be an explanation to why we get a different result when using fixed effects.

Table 4

Dependent variable: Gini Net Coefficient

	IV	IV + FE	FE	FE	IV	IV + FE
	1	2	3	4	5	6
yr_sch	-2.23***	0.05	-0.33**	-1.09**		
	(0.10)	(0.16)	(0.15)	(0.38)		
yr_sch_pri					-2.59***	-1.02
					(0.25)	(0.55)
yr_sch_sec					-3.68***	-0.05
					(0.32)	(0.57)
yr_sch_ter					7.71***	3.94
					(1.61)	(2.20)
Constant	54.13***	37.88***	40.58***	46.84***	55.95***	41.77***
	(0.90)	(1.19)	(1.12)	(1.96)	(1.06)	(2.01)
R^2	0.27	0.27	0.27	0.31	0.27	0.06
Obs.	784	784	784	784	784	784
Time dummies	*No*	*No*	*No*	*Yes*	*No*	*No*

*Notes: Standard errors in parenthesis, ***,** and * are 1,5, and 10 per cent significance level. The dependent variable is the Gini net coefficient. The explanatory variables are average year of schooling, average years of schooling for each level of schooling, primary, secondary and university education of the population 15 and above.*

5.2.2 Instrumental variables

The OLS results suggested that improved educational attainment would decrease income inequality, but the results were not able to tell us anything about the causation of the variables. This will be addressed in this section. Table 4 also presents the results for the instrumental variable regression. In column 1, 2, 5 and 6 parents' education is used as an instrument. The results suggest that the effect of an additional year of schooling remains positive and highly significant in column 1 and 5 where average years of schooling and the different levels of schooling are included in the regression. The effect of average years of schooling does not change significantly under the IV estimation compared to the OLS estimation, but in this estimation we can be sure that the effect we are seeing is a result from increased educational attainment and not from variables that have been omitted from the regression. The effect from each level of education has a more sizable change and all the coefficients become significant at a one per cent level. Column 5 shows that tertiary schooling has a large impact on income inequality, indicating that as a larger part of the population receives a tertiary education the income inequality is increasing. Overall we see that more years of schooling is essential when the goal is to move towards a more equal income distribution, but the large impact of higher education could be decreasing the positive impact.

The results have also been estimated with the inclusion of time dummies, but those results do not change significantly and are not reported. With the use of parents' education as an instrument the results presented in table 4 are able to address some of the concerns that was acknowledged under the OLS estimation. These results support the assumption that by improving and investing in human capital the distribution of income will become more equal. The results in table 4 are suggesting that the cause of effect is going from education to income inequality as the instrumental variable regression is addressing the problem of reverse causality. In other words, as the average years of schooling are increasing the distribution of income is becoming more equal and not the other way around.

The results change when using fixed effects in the instrumental variable regression. The effect on income inequality when the average years of schooling are increased with an additional year is negative; an increase in average years of schooling will increase inequality by 0.05%. At each level of schooling the effects remain unchanged but none of the coefficients are statistically significant. These results are showing that when looking within each country the effect from improved educational attainment does not have a positive impact on income inequality, but as the results show the coefficient is not statistically significant. The limitations of the fixed effects model, which have been presented, may explain the different results.

5.3 The effect of education on income inequality when additional variables are added

Section 5.2 has been concentrating on eliminating some of the bias the study has brought to attention. So far, the results have suggested that as the average years of schooling is increasing the impact on income inequality is positive. In order words the income inequality is decreased when people are attaining more years of schooling. In section 4.5 the development of education was presented and the data showed that the number of years of schooling has been increasing, so we are seeing an overall increase in the amount of education among the population. The same positive development has not been seen in income inequality, which has been fairly stable during the time period being analysed. To further mitigate concerns of omitted variable bias, additional variables have been added for further empirical testing. The variables added are; life expectancy, population growth, GDP, government consumption and openness. The reasoning behind the inclusion of these variables will be presented during the analysis.

5.3.1 OLS estimation

Table 5 presents the OLS estimation of the regression for the additional variables. Note that the variables are presented in the natural logarithm, which changes the interpretation of the coefficients from the previous estimations. The estimation under section 5.1 and 5.2 is a level-level model, where a change in average years of schooling by one, is expected to change the income inequality by the coefficient for average years of schooling. The regression in table 5 is a log-log model where the interpretation is; with a change in the explanatory variable by one per cent, it is expected that the income inequality is changed by the coefficients per cent. One of the attractive features of the log-log-model is that the slope coefficients measure the elasticity of the dependent variable with respect to the explanatory variables. That is, the percentage change in income inequality for a given (small) percentage change in average years of schooling.

The OLS results in table 5 show that life expectancy and population growth have a significant impact when they are added into the regression. The results suggest that as life expectancy is improved the income inequality is decreased. One of the most basic indicators of well-being is life expectancy. Analysts have long recognized the powerful association between personal income and expected life spans [22]. People with higher income tend to live longer than people with lower incomes. The coefficient for life expectancy in table 5 is suggesting that as people live longer the income distribution will become more equal.

Population growth has the opposite effect, that is, as a population grows the higher is income inequality, which makes sense. As the population increases it will have an effect on the income per capita in a country. If the quantity of capital does not change an increase in the population will mean less capital available among the people who are working, and it will cause people to have less income. When people have less income they become poor, and poorer people will then have less money to buy more and better food resulting in bad nutrition, which generally results in bad health. People who are poor also have less money to buy vaccines, get clean water and be able to demand safe working conditions. By looking at the coefficient for both variables, we see that when population growth is increased by one per cent, income inequality is increased by 0.12 per cent. Life expectancy decreases income inequality by 0.39 per cent when it is increased by one per cent. The variables add to the effect coming from human capital on income inequality and can support the importance of improving human capital when addressing income inequality.

[22] Deaton (2003)

Table 5

Dependent variable: ln Gini Net Coefficient

	OLS					
	1	2	3	4	5	6
lnyr_sch	-0.08**	-0.06**	-0.06**	-0.06**	-0.27***	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
lnlife_expect	-0.39***	-0.42***	-0.42***	-0.42***		-0.45***
	(0.09)	(0.09)	(0.09)	(0.09)		(0.09)
lnpop_growth	0.12***	0.11***	0.12***	0.12***		0.12***
	(0.01)	(0.01)	(0.01)	(0.01)		(0.01)
lnrgdpl		0.02**	0.02**	0.02**	0.03**	0.02**
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
lnkg			-0.01	-0.01	-0.02**	-0.03***
			(0.01)	(0.01)	(0.01)	(0.01)
lnopenk				0.002	0.01	0.01
				(0.01)	(0.01)	(0.01)
Constant	5.28***	4.94***	5.60***	5.09***	4.97***	5.40***
	(0.36)	(0.40)	(0.36)	(0.47)	(0.34)	(0.04)
R^2	0.41	0.40	0.40	0.40	0.33	0.37
Obs.	712	686	686	686	747	686
Time dummies	Yes	Yes	Yes	Yes	Yes	No

*Notes: Standard errors in parenthesis, ***,** and * are 1,5, and 10 per cent significance level. The dependent variable is the natural log of the Gini net coefficient. The explanatory variables are average year of schooling of the population 15 and above, life expectancy, population growth, GDP, government consumption, and openness. All the explanatory variables are presented in log.*

The variable GDP is also added into the regression to capture the impact of economic development. The results suggest that as the GDP is increasing, so is income inequality. The GDP coefficient is statistically significant at a five per cent level and the coefficients for education, life expectancy and population growth do not change significantly. The theory by Simon Kuznets suggests that as countries experience economic growth the income inequality first increases and then later decreases. The results in table 5 do indicate that as the income of a country grows larger the disparity in income is increased. To test Kuznets theory the GDP squared was added to capture non-linearites. The results did not support the theory presented by Kuznets, it did instead show a u-shaped relationship. The scarcity of data is making it difficult to properly test the economic development, which also can explain some of the different results found in the literature[23].

Beside the GDP, government consumption is added to the regression. This does not change the coefficients and significance of the other variables. By including government consumption we can examine its effect on income inequality. Government consumption includes the expenditures used on education and health[24], which can have different effects. First, a part of social expenditures consists of direct transfers to the poor, increasing their income and redistributing income from rich to poor. The second is that social expenditures may promote access for the poor to education and other human-capital-enhancing activities, such as health care, thereby contributing to future income equality, especially when credit markets are imperfect. The coefficient for government consumption indicates that an increase in the government consumption is lowering income inequality. The coefficient is not significant when all variables are added.

The results do not change when openness is added into the regression. The more open an economy is the results suggest the higher is income inequality. As openness is calculated as export plus import divided by GDP, the openness variable shows the effect of having a more open economy, which can indicate a higher level of globalization.

In column 5 the regression is estimated without life expectancy and population growth. This changes the significance of the coefficient for average years of schooling as well as the effect. The coefficient for GDP remains significant and the coefficient for government consumption becomes significant. The results suggest that when both life expectancy and population growth are added into the regression the effect of an additional year of schooling keeps having a positive but smaller effect on income inequality. The inclusion of life expectancy and population growth can therefore be important variables when estimating the effect of educational attainment. The two variables add to the positive impact from improvements in human capital on income inequality, and these results do not change when adding the other variables.

[23] See among others Banerjee and Duflo (2003) and Barro (2000)

[24] The consumption concept in PWT 7.0-7.1 includes government expenditures on education and health, which is a major difference from previous versions. See "Brief note on the tables for PWT 7.0-7.1" under Data and Documentation (https://pwt.sas.upenn.edu/php_site/pwt_index.php)

In column 6 the regression is run without time dummies and the results show that the education coefficient become insignificant and the coefficient for government consumption becomes significant.

As addressed previously the OLS results are inconsistent when one or more of the explanatory variables in a regression are correlated with the regressions error term. The correlation between the explanatory variables and the error term can be caused by omitted variables, mis-measured explanatory variables, or an endogenous explanator among the explanatory variables. Therefore, the OLS results do not tell us anything about the causation of the variables added. It does not tell us if a higher life expectancy is causing income inequality to decrease or if the life expectancy is improved because of improved income inequality. In order to cope with this the results just described in table 5 are also estimated using instrumental variables and fixed effects.

5.3.2 Instrumental variables estimation

In this section the results are presented with the use of instruments. The average years of schooling are instrumented with parents' education as previously presented. The other explanatory variables are instrumented with their one-year lag. The instrumental variable estimation is used to overcome the possible bias resulting from the OLS estimation but it is important to keep in mind the possibility of weak instruments. Using the lagged value as an instrument can be questioned, as the lagged variables might not meet the assumption underlying the use of instrumental variables. It can be that some of the lagged variables belong in the regression and can be correlated with the error term. Another problem is that when the values are lagged one year, the estimation will be missing one observation for each country. Therefore, we will not have an instrument for the year 1960, as this is the first year in the dataset. In this estimation, it is assumed that the lagged values are not correlated with the error term and are correlated with the explanatory variable, and are therefore sufficient as instruments[25].

Table 6 presents the results for the two-least square estimation. The results do not change dramatically compared to the OLS results. The results suggest that the effect of education on

[25] See the apendix for the test of endogeneity

income inequality is positive. With a one per cent increase in the average years of schooling the income inequality is decreased by 0.06 per cent when all variables are included. The life expectancy and population growth remain highly significant in all specifications, the same applies for government consumption. The coefficient of GDP is positive in all specifications but only significant when some of the variables are not included.

The results have been estimated both with and without time dummies and are presented in column 4 and 6 respectively. With the inclusion of time dummies the coefficient for average years of schooling is negative and statistically significant at a five per cent level. The coefficient for GDP is positive and insignificant. The results change when the time dummies are removed from the regression in column 6. The coefficients for average years of schooling remain negative but are now insignificant and the coefficient for GDP is significant at a five per cent level. The results for openness are also different when time dummies are added, but the coefficients are not significant in any specification.

Table 6

Dependent variable: ln Gini Net Coefficient

	IV					
	1	2	3	4	5	6
lnyr_sch	- 0.11***	- 0.08**	-0.06*	- 0.06**	- 0.25***	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
lnlife_expect	-0.27**	- 0.31***	- 0.28**	- 0.28**		-0.45***
	(0.10)	(0.10)	(0.10)	(0.10)		(0.09)
lnpop_growth	0.14***	0.12***	0.11***	0.11***		0.12***
	(0.01)	(0.01)	(0.01)	(0.01)		(0.01)
lnrgdpl		0.05**	0.04	0.04	0.05*	0.02**
		(0.01)	(0.03)	(0.27)	(0.03)	(0.01)
lnkg			-0.07**	-0.08**	-0.10***	-0.03***
			(0.01)	(0.03)	(0.03)	(0.01)
lnopenk				-0.01	-0.02	0.01
				(0.02)	(0.03)	(0.01)
Constant	4.98***	3.89***	5.77***	6.14***	5.69***	5.40***
	(0.37)	(0.66)	(0.96)	(1.27)	(1.32)	(0.04)
R^2	0.40	0.37	0.30	0.30	0.15	0.37
Obs.	678	648	648	648	709	686
Time dummies	Yes	Yes	Yes	Yes	Yes	No

*Notes: Standard errors in parenthesis, ***,** and * are 1,5, and 10 per cent significance level. The dependent variable is the natural log of the Gini net coefficient. The explanatory variables are average year of schooling of the population 15 and above, life expectancy, population growth, GDP, government consumption, and openness. All the explanatory variables are presented in log.*

The results from the IV estimation are still supporting the positive effect from improved educational attainment on income inequality. The results remain significant and positive when additional variables are added. The additional variables are allowing us to control for other factors whose omission could bias the results. The results indicate that when all variables are added, the time dimension has a significant importance.

5.3.3 Fixed effects

In this section the fixed effects model with additional variables has also been estimated. The results presented in table 7 are quite different from the results found under the OLS and IV estimations. The results suggest that in all specifications the coefficient for average years of schooling is negative and highly significant as expected. The coefficient for average years of schooling remains the same in column 1 to 4 when additional variables are added. This suggests that the education variable is not affected by the inclusion of the additional explanatory variables.

The coefficient for life expectancy is positive and has the opposite effect on income inequality as previously described. The coefficient of government consumption also becomes positive and is not significant in any specifications. The coefficient of openness remains positive and become significant at a 5 per cent level when time dummies are not included. The coefficient of population growth and GDP remains positive but none of them are statistically significant.

The results from the fixed effect estimation show that as long as we are looking within the same country the effect of improved educational attainment will have a positive impact on income inequality whereas no support for the other variables are found. The important thing to notice is that the education variable in all specification is significant and has the expected sign.

Table 7

Dependent variable: ln Gini Net Coefficient

	Fixed Effects					
	1	2	3	4	5	6
lnyr_sch	- 0.17***	- 0.17***	- 0.17***	- 0.17***	- 0.16***	- 0.13***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
lnlife_expect	0.15	0.19	0.19	0.20		0.18
	(0.13)	(0.13)	(0.13)	(0.13)		(0.12)
lnpop_growth	0.003	0.003	0.003	0.004		0.002
	(0.01)	(0.01)	(0.01)	(0.01)		(0.01)
lnrgdpl		0.00	0.00	0.001	-0.006	0.004
		(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
lnkg			0.001	0.002	0.00	0.003
			(0.01)	(0.01)	(0.00)	(0.01)
lnopenk				0.01	0.01	0.01**
				(0.02)	(0.01)	(0.02)
Constant	3.30***	3.16***	3.13***	2.80***	3.88***	2.66***
	(0.38)	(0.53)	(0.54)	(0.56)	(0.21)	(0.52)
R^2	0.80	0.79	0.79	0.80	0.77	0.80
Obs.	712	686	686	686	747	686
Time dummies	Yes	Yes	Yes	Yes	Yes	No

*Notes: Standard errors in parenthesis, ***,** and * are 1,5, and 10 per cent significance level. The dependent variable is the natural log of the Gini net coefficient. The explanatory variables are average year of schooling of the population 15 and above, life expectancy, population growth, GDP, government consumption, and openness. All the explanatory variables are presented in log.*

5.4 Summary of the results

The results of the empirical analysis suggest that the effect from improved educational attainment on income inequality is positive and statistically significant. The effect has been tested with the use of different econometric methods which all support the positive effect on income inequality resulting from improved educational attainment. The econometric methods applied in the analysis are testing the robustness of the results and are addressing the causation question: does human capital have an effect on income inequality or is it the other way around? By using parents' education as an instrument the results confirm the positive effect going from improved educational attainment to a more equal distribution of income.

The results presented in the OLS regression do not tell us anything about the causality of the education variable. The results suggest that a higher level of education is decreasing the level of income inequality, but with these results we cannot make a final conclusion that the effect is actually coming from improved education. The fixed effects estimations are suffering under the same problem as the OLS, it does not tell us anything about the causality. The fixed effects results are presenting the relationship between educational attainment and income inequality within each country. These results are able to eliminate the possibility of time invariant unobserved effects that could have biased the results. Therefore, when using the fixed effects we are assuming that something within each country may impact or bias the results if we do not control for it. The results presented using the fixed effects support the effect described that a higher level of education is decreasing income inequality. Last the IV results, which are able to address the problem of reverse causality, further support that the causality is going from education to income inequality and not the other way around and that a higher level of education is decreasing income inequality.

Section 4.5 described the development in educational attainment in the world, which showed that average years of schooling have increased from 1960 to 2010. The development shows that the world is moving towards a situation where more people are getting educated, which improve the knowledge and competencies that are placed within each person. The theory suggests that improving the skills and competencies of the population it may impact the income people receive after their education. When people are able to earn a higher income they will also have the possibility to retain a better standard of living and have the opportunity to get a proper health care. Therefore, the improvement in education might increase people's chances of getting a job and the possibility of increasing their income. The results in this study do suggest that income inequality is reduced when educational attainment is improved. The positive effect is although not directly reflected in the development of income inequality. The data shows that the level of income inequality from 1960 to 2010 has been fairly stable and the world is still experiencing inequality at a high level. Castelló-Climent and Doménech (2014) show with their results that the fall in education inequality has reduced income inequality, but its effects have been offset by an increasing demand for skilled workers and the effect of globalization. They also suggest that the return to education is convex, which can be another force that offset the positive effect on income inequality resulting from improved educational attainment. The results in this empirical analysis present educational attainment at the different levels of education. The results show that tertiary education has a sizeable effect compared to secondary and primary education. In the world today we are experiencing that the role of technology is important and the demand for people with higher education is likely to increase. If the supply of highly educated people is low, the wage which firms are willing to pay will be higher. In advanced countries most of the improvements in years of education come from higher secondary and tertiary completion and enrolment ratios, in developing countries the improvements are accounted for by higher primary and secondary completion and enrolment ratios. So when people complete a higher education they will be able to receive a higher wage and increase their income. If there exists a large difference between the amount of people who receive a high education and the amount of people who receive a lower education the distribution of income will be less equal.

Aghion, Caroli, and Carcia-Penalosa (1999) show that given that capital market imperfections are the root of the relationship between inequality and growth, transfers or subsidies to borrowers are an important policy tool, which is also particularly relevant in the case of investments in human capital. Increased access to education would also reduce inequalities between dynasties, as it would diminish the effect of family wealth upon individuals' investment possibilities, thereby increasing ex-ante equality. However, they find that increasing the supply of skills has a counteracting impact on wage inequality because it is itself a cause of skill-biased technical change. In their analysis of biased technical change, they find that its effect on earnings inequality is nonlinear. With the arrival of a new "General Purpose Technology" the skill premium increases because of the high demand for skilled "experimentation" labour during the first stages of social learning. Thereafter the skill premium starts tapering off. They show that the mechanisms tend to generate a kind of alternative Kuznets curve, with inequality first rising and then falling during the transition to a new technological paradigm. Therefore, the impact of education on the dispersion of wage is ambiguous.

The results presented show that the improvements in education and human capital inequality observed in many countries have not resulted in significant increases in income inequality due to forces in the opposite direction or because the effect is ambiguous. Thus, the results presented in this study do not find a consistent and significant support for the effect of globalization when openness was added into the regression. The results in some specifications did suggest that the more open an economy is, the higher is income inequality, but the coefficient for openness was only significant in one specification. The GDP was also added to capture the impact of economic development and the results indicated that some countries might still be developing and therefore the coefficient was positive and causing the income inequality to increase.

The empirical results show that most of the variability in the income Gini coefficient comes from variability across countries. Thus, econometric techniques that exploit the cross-sectional or the within country variation in the data might give different results. We see that when using the fixed effects model the coefficients leave the within-country action, which has been quite stable. Therefore, when using the fixed effect model we can control for unobserved heterogeneity, but at the expense of the very low within country variation in the income Gini coefficient. When looking at the whole world it is most likely that the countries have individual country specific effects that might impact the results. When these are omitted it can bias the results, but the fixed effects model do have a limitation, as it cannot assess the effect of variables that have little within-group variation. Therefore, the low variation in the Gini coefficient can be an explanation to why we see a different result when using fixed effects. It should be mentioned that the fixed effects estimations all confirm the positive relation between educational attainment and income inequality. The results suggested that, both when educational attainment was included as the only explanatory variable and when additional variables were added as explanatory variables, by increasing education the coefficient for average years of schooling was negative, indicating that income inequality is decreased when education is improved.

The purpose of this thesis has been to outline the effect of between human capital on income inequality and be able to present robust results that can explain the relationship. The results presented in the empirical analysis are successfully addressing the problem of endogeneity of the inequality variable, that is, when other omitted factors are correlated with both education and inequality measures, or when the causation goes the other way around. The results therefore succeed in addressing the cause of effect and are able to confirm a positive effect from improved education on income inequality. One of the concerns acknowledged with education is the unobservable ability factor. Improvements in educational attainment will not necessarily have the same outcome for all people because the ability factor is individual to each person. In section 2.4 a group of studies was mentioned, which focused on the effect of family income on children's education outcome. The idea behind these studies is that rich parents can spend more or have access to credit on their children's education than poor parents and these investments can lead to better outcome for their children. Other factors, such as externalities, which are an incidental effect of some economic activity for which no compensation is provided, can have a relation to education. Giving one person more education will raise not only her own output but also the output of those around her. There are other positive externalities, for example a more educated population is more likely to have an honest and efficient government. Another positive effect of human capital can be that governments often get involved in improving human capital (in form of public education or mandatory schooling).

With the use of instruments this study has made it possible to mitigate some of the bias presented. All the factors just described could be influencing income inequality and education, which is why the IV estimation is enabling us to give a more robust result when trying to answer the problem outlined in the thesis. The results presented in the IV estimation are only showing the effect resulting from improved educational attainment, which confirms a positive relationship i.e. improved educational attainment decreases income inequality. The effect does not change when more variables are added into the regression in table 6. The coefficients for the education variable remain negative and significant.

5.6 Partial Conclusion

This chapter has presented the empirical investigation and highlighted the main findings of the study. The different econometric approaches have presented different results, which have all been discussed. The main findings conclude that the effect of improved educational attainment have a positive and statistically significant effect on income inequality. The empirical results also present some differences, which might be an effect of low variation in the income inequality data, which have been taken into account in the conclusion of the results.

Chapter 6

6. Comments

This chapter discusses some of the implications and limitations of the results of the empirical investigation. Section 6.1 compares the literature about human capital and income inequality presented in chapter 2 with the empirical results found in chapter 5. Section 6.2 presents the implications and limitations of this study and section 6.3 presents ideas for further research.

6.1 The relationship between human capital and income inequality

This section will reflect on some of the relations between human capital and income inequality, which were presented in chapter 2, and the results found in the empirical analysis. As already described Castelló-Climent and Doménech (2014) are presenting results, which suggest that the positive effect from improved education on income inequality has been offset by forces in the opposite direction and that the return to education is convex. The theory presented by Kuznets suggests that as an economy develops income inequality first increases and then decreases later on. Therefore, some countries might still be developing and will have an income inequality at a high level. This could be some of the explanations to why we have not seen a large improvement in income inequality in the world. Section 2.4 also presented some of the literature concerning the problem of this study, which suggests that the distribution of wealth is crucial to determine individuals' education choice. The results suggest that families are prevented from accessing school when they have a low income. These studies are lacking in properly addressing the endogeneity of the income variable, which make it difficult to say if the effect is actually going in that direction. This reverse causality has been highlighted throughout the paper and is one of the main contributions from this study. This study succeeds in addressing the causation of the effect; income inequality is reduced when human capital is improved.

Knight and Sabot (1983) argue that an expansion of education has two different effects on the earning distribution. One is the "composition" effect, which increases the relative size of the group with more education and tends initially to raise income inequality, but to eventually lower it. The data presented in this study shows that the overall education attainment has increased. Developed countries have improved in higher secondary and tertiary education and developing countries in primary and secondary education. The proportion of uneducated people in the developing countries has declined significantly from 64.9% in 1950 to 20.1% in 2010. This shows that the group with more education gets bigger and can be causing income inequality to increase. The second effect is the "wage composition" effect, which decreases the premium on education as the relative supply of educated workers increases. Therefore, as the size of the group with more education increases the supply of educated workers increases, which might decrease the premium on education. These effects of increased education on the distribution of income can therefore be ambiguous, which can be another explanation to the results presented in this study.

6.2 Implication and limitations of the results

The results of the empirical investigation should be interpreted with caution. The scarce availability of data is without doubt a limitation for the empirical investigation. In chapter 4 the data was presented, which showed that when looking at the whole world only 23 countries had data for both educational attainment and income inequality available in 1960 compared to 89 countries in 2010. The implications of having this unbalanced dataset are that the new data, which is becoming available, might come from countries with greater income inequality then the previous average.

The time period being analysed is also limited by the data on educational attainment, which is only available in five years intervals. This can make it difficult to properly test the effect of the time dimension and it limits the dataset for having more variation in the observations opposed to if data has been available for every year in the period.

The literatures that have analysed the effect of income inequality and human capital on economic growth have found mixed results. It might be that while human capital inequality data have enough signal in both cross-section and panel data models, the stability of the income Gini coefficient could explain the mixed evidence found in the literature. Whereas cross-section estimates show that income inequality is harmful to growth (Persson and Tabellini (1991)), the panel data models indicate that the relationship is not linear (e.g. Barro (2000), Banerjee and Duflo (2003)), positive (e.g. Forbes (2000)) or dependent on the time lag (e.g. Halter, Oechslin and Zweimüller (2014)). Conversely, Castelló and Doménech (2002) show that a more uneven distribution of human capital has a negative influence on the growth rates per capita in both cross-section and panel data models. These evidences might also explain some of the mixed results found in this study.

The choice of economistic methods can have some implications for the results found in the empirical investigation. As mentioned in chapter 3 a potential significant limitation of the fixed effects model is that you cannot access the effect of variables that have little within group variation. The fixed effects model allows the results to be adjusted for unobserved heterogeneity but at the expense of the very low variation in the income Gini coefficient. It can be discussed if the fixed effects model is suited for this empirical investigation, but it has been found important to include the results as the study is looking at global income inequality and the fact the countries might differ from each other, these effects have been found important to adjust for.

The instrumental variable estimation is very useful in allowing the researcher to avoid the bias that the ordinary least squares model suffers from when an explanatory variable in a regression is correlated with the regression disturbance term. But the instrumental variable estimation requires that the instrument is not invalid and not weak. The choice of instrument is essential for receiving a valid result for one's instrumental variable estimation. The validity of the use of parents' education as an instruments can of course be questioned because how can you be sure that you are matching the correct parent and child. This study has chosen to use this approach presented by Barro and Lee (2013) together with tests and explanations and allowed this to serve as the argumentation for the validity of the instrument.

6.3 Ideas for further research

This section provides some ideas for further research of the relationship between human capital and income inequality.

The results of this study have been limited by data availability and this can have had an effect to the results presented. It might be that the results will change if the study was adjusted to only look at a specific region or country where you could obtain a balanced dataset and be able to investigate the effect further.

The data availability keeps improving, which make this investigation interesting not just now but also in the future when larger dataset become available.

In this study, educational attainment is used as a proxy for human capital. Human capital has other proxies which also can be interesting to investigate.

Chapter 7

7. Conclusion

The outline of this thesis has been to investigate the effect of human capital on income inequality. Income inequality has for long been an interesting subject among economists as income inequality is seen by many as an important component for a country's overall development. Therefore, it has been interesting to investigate some of the factors that have an impact on income inequality in the world. Throughout the paper human capital has been presented as an important factor when it comes to addressing income inequality. In this study educational attainment has been used as a proxy for human capital and the empirical results presented show that improved educational attainment has a positive effect when the desired outcome is to decrease income inequality.

All over the world the development in educational attainment has increased and as education has been argued to be essential because it improves peoples' skills, competencies and productivity, it can result in a higher wage hence a higher income. This positive relation would therefore indicate that a higher level of education would result in a more equal distribution of income. This relation has been investigated by carrying out OLS, fixed effects and instrumental variable estimations, which all find support of the positive effect from improved educational attainment on income inequality.

This thesis has outlined the limitations and implications of the empirical investigation, where a main focus has been on eliminating the biases that may be present in the empirical investigation. There can be multiple factors that can have an impact on the distribution of income both between and within a country, which have made it important for this thesis to mitigate the potential effect coming from other sources and be able to outline the effect coming from human capital.

The factors that have an impact on income inequality will likely be different depending on which country you look at. There can be cultural differences, different climate, different religion etc. that will have an influence when people decide to invest in more education. Therefore, the fixed effects model was estimated to adjust for these country specific characteristics. The results presented were slightly different from the OLS and the IV results, but they do support that education has a positive effect on income inequality.

The problem with the OLS estimation and the fixed effects is their lack of addressing the problem of causality. These estimations do not show if the effect is actually coming from human capital. The empirical results presented in this study have therefore successfully been addressing the problem of endogeneity. This has been made possible by the improved datasets on educational attainment from Barro and Lee (2013) and the Gini coefficients obtained from the Standardized World Income Inequality Database. By using different econometric methods it has made it possible for this study to present more robust results. As the focus has been to eliminate the problem of endogeneity, the instrumental variable estimation has been performed, where parents' education has been used as an instrument for the education variable. With the use of instruments it is possible to attack the reverse causality problem that might occur when estimating an OLS regression with the Gini coefficient as the dependent variable and the average years of schooling as the independent variable. The instrumental variable estimation only presents the effect resulting from improved education, which allows this thesis to conclude that by improving human capital it will result in a more equal distribution of income in the world.

Bibliography

Acemoglu, D. & Pischke, J.-. 2001, "Changes in the wage structure, family income, and children's education", *European Economic Review,* vol. 45, no. 4, pp. 890-904.

Aghion, P., Caroli, E. & Garcia-Penalosa, C. 1999, "Inequality and Economic Growth: The Perspective of the New Growth Theories", *Journal of Economic Literature,* vol. 37, no. 4, pp. 1615-1660.

Akee, R.K.Q., Copeland, W.E., Keeler, G., Angold, A. & Costello, E.J. 2010, "Parents' Incomes and Children's Outcomes: A Quasi-Experiment Using Transfer Payments from Casino Profits", *American Economic Journal: Applied Economics,* vol. 2, no. 1, pp. 86-115.

Alberto, A. 1993, *Income Distribution, Political Instability, and Investment,* National Bureau of Economic Research;, Cambridge, Mass.

Baltagi, B.H. 2008, *Econometric Analysis of Panel Data,* 4th ed. edn, United Kingdom: Wiley-Blackwell Etextbooks.

Banerjee, A.V. & Newman, A.F. 1993, "Occupational Choice and the Process of Development", *Journal of Political Economy,* vol. 101, no. 2, pp. 274-298.

Banerjee, A. & Duflo, E. 2003, "Inequality and Growth: What Can the Data Say?", *Journal of Economic Growth,* vol. 8, no. 3, pp. 267-299.

Barro, R.J. 2000, "Inequality and Growth in a Panel of Countries", .

Barro, R.J. & Lee, J.W. 2013, "A new data set of educational attainment in the world, 1950–2010", *Journal of Development Economics,* vol. 104, pp. 184-198.

Becker, G.S. 1962, "Investment in Human Capital: A Theoretical Analysis", *Journal of Political Economy,* vol. 70, no. 5, pp. 9-49.

Becker, G.S. 1993, "*Human capital : a theoretical and empirical analysis, with special reference to education",* Chicago The University of Chicago Press, Chicago.

Cameron, S.V. & Heckman, J.J. 2001, "The Dynamics of Educational Attainment for Black, Hispanic, and White Males", *Journal of Political Economy,* vol. 109, no. 3, pp. 455-499.

Castelló, A. & Doménech, R. 2002, "HUMAN CAPITAL INEQUALITY AND ECONOMIC GROWTH: SOME NEW EVIDENCE", *Economic Journal,* vol. 112, no. 478, pp. C187-C200.

Castelló-Climent, A. & Doménech, R., 2014 2014, "Human Capital and Income Inequality: Some Facts and Some Puzzles", , no. Working Papers 1228, BBVA Bank, Economic Research Department.

Checchi, D. 2003, "Inequality in Incomes and Access to Education: A Cross-country Analysis (1960–95", *LABOUR,* vol. 17, no. 2, pp. 153-201.

Checchi, D. 2000, "Does Educational Achievement Help to Explain Income Inequality?", .

Chiswick, B.R. 1971, "Earnings Inequality and Economic Development", *The Quarterly Journal of Economics,* vol. 85, no. 1, pp. 21-39.

Cohen, D. & Soto, M. 2007, "Growth and human capital: good data, good results", *Journal of Economic Growth,* vol. 12, no. 1, pp. 51-76.

Deaton, A. 2003, "Health, Inequality, and Economic Development", *Journal of Economic Literature,* vol. 41, no. 1, pp. 113-158.

Filmer, D. & Pritchett, L. 1999, "The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries", *Population and Development Review,* vol. 25, no. 1, pp. 85-120.

Forbes, K.J. 2000, "A Reassessment of the Relationship between Inequality and Growth", *The American Economic Review,* vol. 90, no. 4, pp. 869-887.

Fuente, D.L. & Domenech, R. 2006, "Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?", *Journal of the European Economic Association,* vol. 4, no. 1, pp. 1-36.

Galor, O. & Zeira, J. 1993, "Income Distribution and Macroeconomics", *The Review of Economic Studies,* vol. 60, no. 1, pp. 35-52.

Gregorio, J.D. & Lee, J. 2002, "Education and Income Inequality: New Evidence From Cross-Country Data", *Review of Income and Wealth,* vol. 48, no. 3, pp. 395-416.

Halter, D., Oechslin, M. & Zweimüller, J. 2014, "Inequality and growth: the neglected time dimension", *Journal of Economic Growth,* vol. 19, no. 1, pp. 81-104.

Hanushek, E.A. & Kimko, D.D. 2000, "Schooling, Labor- Force Quality, and the Growth of Nations", *The American Economic Review,* vol. 90, no. 5, pp. 1184-1208.

Jallade, J. 1982, "Basic education and income inequality in Brazil", *World Development,* vol. 10, no. 3, pp. 187-197.

Knight, J.B. & Sabot, R.H. 1983, "Educational Expansion and the Kuznets Effect", *The American Economic Review,* vol. 73, no. 5, pp. 1132-1136.

Krueger, A.B. & Lindahl, M. 2001, "Education for Growth: Why and For Whom?", *Journal of Economic Literature,* vol. 39, no. 4, pp. 1101-1136.

Li, J. & Yu, H. 2014, "Income inequality and financial reform in Asia: the role of human capital", *Applied Economics,* vol. 46, no. 24, pp. 2920-2935.

Lochner, L. & Moretti, E. 2004, "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports", *The American Economic Review,* vol. 94, no. 1, pp. 155-189.

Mayrhofer, T. & Schmitz, H. 2014, "Testing the relationship between income inequality and life expectancy: a simple correction for the aggregation effect when using aggregated data", *Journal of Population Economics,* vol. 27, no. 3, pp. 841-856.

Milanovic, B. 2013, "Global Income Inequality in Numbers: in History and Now", *Global Policy,* vol. 4, no. 2, pp. 198-208.

Murray, M.P. 2006, "Avoiding Invalid Instruments and Coping with Weak Instruments", *The Journal of Economic Perspectives,* vol. 20, no. 4, pp. 111-132.

Persson, T. & Tabellini, G. 1991, *Is Inequality Harmful for Growth? Theory and Evidence*, Cambridge, Mass., USA.

Ram, R. 1984, "Population increase, economic growth, educational inequality, and income distribution: Some recent evidence", *Journal of Development Economics,* vol. 14, no. 3, pp. 419-428.

Ram, R. & Ural, S. 2014, "Comparison of GDP Per Capita Data in Penn World Table and World Development Indicators", *Social Indicators Research,* vol. 116, no. 2, pp. 639-646.

Ray, D. 1998, "Development Economics".

Solt, F. 2009, "Standardizing the World Income Inequality Database *", *Social Science Quarterly,* vol. 90, no. 2, pp. 231-242.

Sylwester, K. 2002, "Can education expenditures reduce income inequality?", *Economics of Education Review,* vol. 21, no. 1, pp. 43-52.

Topel, R.H. 1997, "Factor Proportions and Relative Wages: The Supply- Side Determinants of Wage Inequality", *The Journal of Economic Perspectives,* vol. 11, no. 2, pp. 55-74.

United Nations Development Programme 2013, "Humanity Divided: Confronting Inequality in Developing Countries", .

Weil David, N. 2005, *Economic growth,* , Boston.

World Bank *, World Bank Databank*. Available: http://data.worldbank.org/.

World Bank, 200,6 Equity and Development.

World Health Organization (WHO) 2008, "Closing the Gap in a Generation Health: Equity through Action on the Social Determinants".

Appendix

Instrumental Variables - *Tests of the instruments*

Regression: $yr\_sch = p_0 + p_1 Parents\_Education + v$

Dependent variable: Average years of Schooling
Independent variable: Parents' education

```
. reg yr_sch IV

      Source |       SS       df       MS              Number of obs =    1898
-------------+------------------------------           F(  1,  1896) =30039.87
       Model |  17369.8416     1  17369.8416           Prob > F      =  0.0000
    Residual |  1096.31695  1896   .57822624           R-squared     =  0.9406
-------------+------------------------------           Adj R-squared =  0.9406
       Total |  18466.1585  1897  9.73440092           Root MSE      =  .76041

------------------------------------------------------------------------------
      yr_sch |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          IV |   .9486699   .0054735   173.32   0.000     .9379352    .9594047
       _cons |    1.28306   .0297369    43.15   0.000     1.224739     1.34138
------------------------------------------------------------------------------
```

Tests of endogeneity

```
. ivregress 2sls GINI_NET (yr_sch = IV)

Instrumental variables (2SLS) regression               Number of obs =     784
                                                       Wald chi2(1)  =  330.97
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.2700
                                                       Root MSE      =  9.6078

------------------------------------------------------------------------------
    GINI_NET |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      yr_sch |  -2.231935    .122683   -18.19   0.000     -2.47239   -1.991481
       _cons |   54.13061   .9388796    57.65   0.000     52.29044    55.97078
------------------------------------------------------------------------------
Instrumented:  yr_sch
Instruments:   IV

. estat endogenous

  Tests of endogeneity
  Ho: variables are exogenous

  Durbin (score) chi2(1)          =  45.1533  (p = 0.0000)
  Wu-Hausman F(1,781)             =  47.7295  (p = 0.0000)
```

```
. ivregress 2sls log_GINI (log_yr_sch log_lifeexp log_popgrowth log_realrgdpl log_realkg log_openk = log_IV laglog_lifeexp laglog_
> gdpl laglog_realkg laglog_openk)

Instrumental variables (2SLS) regression             Number of obs =     648
                                                       Wald chi2(6)   =  328.47
                                                       Prob > chi2    =  0.0000
                                                       R-squared      =  0.2517
                                                       Root MSE       =  .25021

------------------------------------------------------------------------------
    log_GINI |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  log_yr_sch |  -.0310343   .0330854    -0.94   0.348    -.0958805     .033812
 log_lifeexp |  -.3079964   .1042215    -2.96   0.003    -.5122668    -.103726
log_popgrowth|   .1126386   .0154532     7.29   0.000     .0823508    .1429263
log_realrgdpl|   .0494573   .0278106     1.78   0.075    -.0050504     .103965
  log_realkg |  -.0872417   .0282852    -3.08   0.002    -.1426797   -.0318038
   log_openk |   .0023057   .0259579     0.09   0.929    -.0485709    .0531823
       _cons |   5.752674   1.309585     4.39   0.000     3.185934    8.319414
------------------------------------------------------------------------------
Instrumented:  log_yr_sch log_lifeexp log_popgrowth log_realrgdpl log_realkg
               log_openk
Instruments:   log_IV laglog_lifeexp laglog_popgrowth laglog_rgdpl
               laglog_realkg laglog_openk

. estat endogenous

  Tests of endogeneity
  Ho: variables are exogenous

  Durbin (score) chi2(6)          =  33.0154  (p = 0.0000)
  Wu-Hausman F(6,635)             =  5.68165  (p = 0.0000)

.
```

Explanation of the variability of the Gini coefficient

Dependent variable: GINI_NET
Independent variable: Country dummies

The regression is a screenshot that only contains some of the countries

```
reg  GINI_NET i.WBcodex

      Source |       SS       df       MS              Number of obs =    4597
-------------+------------------------------           F(172,  4424) =  106.91
       Model |  442799.756   172  2574.41719           Prob > F      =  0.0000
    Residual |  106531.802  4424  24.0804253           R-squared     =  0.8061
-------------+------------------------------           Adj R-squared =  0.7985
       Total |  549331.557  4596  119.523838           Root MSE      =  4.9072

------------------------------------------------------------------------------
    GINI_NET |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     WBcodex |
        ARG  |   10.45641   1.537755     6.80   0.000     7.441646    13.47118
        ARM  |   4.086845   1.666887     2.45   0.014     .8189139    7.354777
        AUS  |  -.6723502   1.524633    -0.44   0.659    -3.661394    2.316693
        AUT  |  -3.538108   1.637895    -2.16   0.031    -6.749202   -.3270144
    Andorra  |  -1.878619   3.143111    -0.60   0.550     -8.04069    4.283452
     Angola  |   19.56554   1.890072    10.35   0.000     15.86005    23.27102
   Anguilla  |  -2.112438   5.092422    -0.41   0.678    -12.09613    7.871257
 Azerbaijan  |  -.7599684   1.600184    -0.47   0.635    -3.897129    2.377193
        BDI  |   4.411703    1.85949     2.37   0.018     .7661726    8.057233
        BEL  |  -5.808088   1.606876    -3.61   0.000    -8.958369   -2.657808
        BEN  |   4.555067   2.805788     1.62   0.105    -.9456811    10.05582
        BGD  |   5.777267   1.571556     3.68   0.000     2.696232    8.858303
        BGR  |  -6.553455    1.53094    -4.28   0.000    -9.554863   -3.552046
        BLZ  |   21.16639   2.300522     9.20   0.000     16.65622    25.67657
```

Dependent variable: GINI_NET
Independent variable: Time dummies

The regression is a screenshot that only contains some of the years

```
reg GINI_NET i.year
```

Source	SS	df	MS
Model	18230.325	52	350.583173
Residual	531101.232	4544	116.879673
Total	549331.557	4596	119.523838

Number of obs =	4597
F(52, 4544) =	3.00
Prob > F =	0.0000
R-squared =	0.0332
Adj R-squared =	0.0221
Root MSE =	10.811

GINI_NET	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
year						
1961	-3.620472	3.000677	-1.21	0.228	-9.503259	2.262314
1962	-6.342506	2.942406	-2.16	0.031	-12.11105	-.5739591
1963	-5.88268	2.970564	-1.98	0.048	-11.70643	-.0589308
1964	-6.809781	2.942406	-2.31	0.021	-12.57833	-1.041234
1965	-6.917604	2.89123	-2.39	0.017	-12.58582	-1.249388
1966	-8.920208	3.105091	-2.87	0.004	-15.0077	-2.83272
1967	-5.809646	2.970564	-1.96	0.051	-11.6334	.0141032
1968	-2.531621	2.692744	-0.94	0.347	-7.810708	2.747466
1969	-2.642665	2.752369	-0.96	0.337	-8.038646	2.753316
1970	-2.197951	2.621019	-0.84	0.402	-7.336423	2.94052
1971	-2.23549	2.736382	-0.82	0.414	-7.600129	3.129149
1972	-7.018477	2.786847	-2.52	0.012	-12.48205	-1.554903

List of the 123 countries included in the study				
Albania	Algeria	Kazakhstan	Panama	Uruguay
Argentina	Ecuador	Kenya	Peru	United States
Armenia	Egypt	Cambodia	Philippines	Venezuela
Australia	Spain	Lao	Papua New Guinea	Viet Nam
Austria	Estonia	Liberia	Poland	South Africa
Burundi	Finland	Sri Lanka	Portugal	Zambia
Belgium	Fiji	Lesotho	Paraguay	Zimbabwe
Benin	France	Lithuania	Romania	
Bangladesh	Gabon	Luxembourg	USSR	
Bulgaria	United Kingdom	Latvia	Rwanda	
Belize	Ghana	Morocco	Senegal	
Bolivia	Gambia	Maldives	Serbia	
Brazil	Greece	Mexico	Singapore	
Barbados	Guatemala	Mali	Sierra Leone	
Botswana	Guyana	Malta	El Salvador	
Central African Republic	Honduras	Mongolia	Slovak Republic	
Canada	Croatia	Mozambique	Slovenia	
Switzerland	Haiti	Mauritania	Sweden	
Chile	Hungary	Mauritius	Swaziland	
China	Indonesia	Malawi	Togo	
Cote d`Ivoire	India	Malaysia	Thailand	
Cameroon	Ireland	Namibia	Tajikistan	
Colombia	Iran	Niger	Trinidad and Tobago	
Costa Rica	Iceland	Nicaragua	Tunisia	
Cyprus	Israel	Netherlands	Turkey	
Czech Republic	Italy	Norway	Taiwan	
Germany	Jamaica	Nepal	Tanzania	
Denmark	Jordan	New Zealand	Uganda	
Dominican Republic	Japan	Pakistan	Ukraine	