Why can’t we agree on the effect of education on income inequality?

A meta-analysis

Ida Marie Støp Meland

Master’s thesis

Spring 2022

Department of Comparative Politics

University of Bergen

Abstract: Does education influence income inequality? Despite being a widely examined subject within political, sociological, and economic research over the past 70 years, researchers have not reached unanimity on whether higher levels of education lead to a decline in societal income inequality. This thesis contains a meta-analysis of existing research to investigate this relationship, reviewing 61 studies conducted between 1971 and 2020, with 917 individual estimates and approximately 71.500 datapoints. The results show that there has been a change in research findings on this topic over time, as recent research reports a stronger negative effect of education on national income inequality compared to research published more than 10 years ago. These findings are in line with the traditional hypothesis which holds that education has an equalizing effect. Additionally, primary education is found to have the strongest effect on income inequality, compared to secondary- and tertiary education. There is no indication of publication bias within the selected research. Heterogeneity between the various studies from the past 50 years of research is in part explained by variation in study characteristics such as statistical framework, choice of conceptualization and measurements, and source of data. Researchers producing empirical research on the relationship between education and income inequality are advised to consider the revealed sources of heterogeneity from this study.

Acknowledgements

I want to thank my supervisor Michael Alvarez for his guidance, support, and impeccable effort to answer my many emails. A humble thanks to Chris Doucouliagos who let me use the data collected by him and his colleagues Abdul Abdullah and Elizabeth Manning for their article "Does Education Reduce Income Inequality? A Meta-Regression Analysis", published in *Journal of Economic Surveys* in 2013. A further thanks to Troy Broderstad, for lending me his notes the work with the impressive master thesis "A Meta-Analysis of Income and Democracy" and providing me with guidance when I have felt lost about how to conduct this analysis.

A thanks to my friend and study-partner Alex, without you the past two years would have been much, much duller. Thank you Erlend and Andrea for the mornings at Nordnes sjøbad, and your infinite support.

Lastly, a thank you to my greatest educators, Mum and Dad.

Contents

List of figures

List of tables

Abbreviations

FAT-PET-MRA	Funnel-asymmetry-test and precision-effect test for meta regression analysis
FE	Fixed effects
GMM	Generalized Method of Moments
G-to-S	General to specific
IV	Instrumental variables
MEML	Mixed effects multilevel model
MRA	Meta regression analysis
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary least squares
RE	Random effects
SLS	Stage last square
WLS	Weighted least squares
WLS CR	Weighted least squares with cluster-robust standard errors
WLS R	Weighted least squares with robust standard errors

1 Introduction

1.1 Subject of investigation - Education and Income Inequality

Growing economic inequality has become a controversial topic amongst political, economic and sociological researchers during the past decade. A type of economic inequality that is thought to be especially important is *inequality in income.* Despite being a much researched topic, there is little knowledge on why we observe greater income inequality in some societies and time periods than in others – and what factors might counteract this type of inequality (Piketty 2014, 304; Cowell 1998, 1; Neckerman and Torche 2007, 337).

A wealth of research proposes *education* as a factor in reducing income inequality (Sylwester 2002, 43; 2003, 250; Rodríguez-Pose and Tselios 2009, 414; Piketty 2014, 305). An increase in formal education at different levels has been perceived as an "equalizer" within central political, economic, and sociological theories (De Gregorio and Lee 2002; Barro 2000; Ahluwalia 1976a; Checchi 2001). In general, these theories forward the expectation that increasing levels of education amongst the population within a country should be followed by an observable decline in income inequality. Educational spending is often justified as a policy measure in accordance with this linear causal argument, with schools serving as "engines of social mobility" (Croix 2020, 403; Partridge, Mark D. Partridge, and Rickman 2007, 282; De Gregorio and Lee 2002, 395; Tselios 2009, 414; Gerring 2012, 226).

However, several empirical studies suggest that the relationship between education and income inequality is far from clear (Partridge, Mark D. Partridge, and Rickman 2007, 282; De Gregorio and Lee 2002, 395; Rodríguez-Pose and Tselios 2009, 412; Ram 1981, 253; Piketty 2014). Acknowledging the fact that education is suited to reduce poverty and improve people's lives in many ways, researchers still question whether increased levels of education should be expected to affect the general levels of income inequality in a country.

More knowledge about what factors might influence income inequality is relevant for both academic and political purposes. Observing a relationship between these two factors would imply that policy makers could highlight education as a tool to reduce income inequality. However, if the findings show that education fails to reduce, or even increase income inequality, this will indicate that there is a need for more knowledge surrounding the possible driving and counteracting factors of income inequality today – and that the meritocratic ideal where "learning is earning" might be challenged.

Fifty years of research have not resolved this conundrum. Thus, it is highly relevant to investigate the empirical findings regarding educations possible effect on income inequality.

1.2 Existing empirical findings

Cross-country differences in income distribution have received a lot of attention in the literature since the middle of the 1950's (Psacharopoulos 1977, 383). Simon Kuznets' theory of an inverted U-shaped curve - where levels of inequality would first be increasing and then decrease due to the financial development of a country - became a popular starting point for discussing how income might be distributed in countries over time (Kuznet 1955, 1963; Barro 2000, 9; Ahluwalia 1976a, 308). With more detailed theories on human capital accumulation in the 1960's, education became a central variable for studies investigating income distribution (Psacharopoulos 1977, 383; Ahluwalia 1976b, 131). Aigner and Heins (1967, 180), Ahluwalia (1976a; 1976b), Adelman and Morris (1973) and Chenery and Syrquin (1975) are among the scientist in the 1960's and 70's that found empirical observations that suggest a clear negative relationship between the improvement of education and income inequality within their cross-country analysis, proposing that education has an equalizing effect. The opposite effect of education is found by Psacharopoulos (1977), Rogers (1983), and Ram (1984), as they observe that increase in educational levels is correlated with higher income inequality. Several other authors also find contradicting evidence *within* their studies, which can be used to argue that there is no clear one-way relationship between the two variables (Jha 1996; Winegarden 1979; Tsai 1995; Chiswick 1971).

When topics are heavily investigated by many researchers, one might expect some clearer findings on the subject to emerge over time - as the dialog between researchers might contribute to understanding why the existing results tend to differ. However, the divergence in findings on this specific relationship can also be observed in more recent research on the subject. Chu and Hoang (2020), Le et. Al (2020), Lee and Vu (2020), Calderon and Chong (2009) and Chong (2004) are among the researchers that identify negative effects of education on income inequality, in contrast to Shabadi et. Al (2018), Park and Mercado (2018) Andres and Ramlogan Dobson (2012) and Carter (2007), who find predominantly positive correlations between education and income inequality. Recent studies also include contradicting findings on the subject, suggesting a blurred picture (Chong 2004; Gruber and Kosack 2014; Dorrenberg and Peichl 2014).

As there have been divergent results on this matter for over 60 years, a common literature search regarding the observed relationship between education and income inequality is unsuitable for providing a clear answer to what effect we observe education to have on income inequality. Researchers on the subject have suggested that the wide variety of measurements, estimations, and statistical frameworks could provide a possible explanation for the contrasting findings on this topic (Chiswick 1971, 33; Williamson 1999, 4; Barro 2000, 7; Gruber and Kosack 2014, 254). This indicates that there is a knowledge gap related to the underlying explanations as to why the research on education and income inequality is so heterogenous.

1.3 Background for this study

Due to the variation of findings on this topic, Abdul Abdullah, Hristos Doucouliagos and Elizabeth Manning published the study "Does Education Reduce Income Inequality? A Meta-Regression Analysis" in the *Journal of Economic Surveys* in 2013. Based on data from 63 studies from 1970 to 2010, they found that education had a moderate negative effect on income inequality(Abdullah, Doucouliagos, and Manning 2013, 12).[1] More specifically, they found that education has a stronger effect on reducing the income share of the rich and increasing the income share of the poor. Further, they suggested that secondary education is more important in reducing inequality than primary schooling and higher education.

Since the publication of the meta-analysis by Abdullah et al. (2013), several scholars have drawn attention towards a rise in income inequality, largely driven by disparity in wages between high- and low-skilled sectors (Solga 2014, 270; Piketty 2014, 304). Some even argue that the findings on economic inequality in recent years have "transformed our economic discourse" (O'Neill 2017, 343; Krugman 2014, 71). At the same time, scientists propose that the world is more educated than ever before (Ortiz-Ospina 2016). Does this indicate that the equalizing powers of education has weakened?

[1] This conclusion is draw according to Doucouliagos (2011) guidelines for interpreting for interpreting effect sizes in meta-regression analysis, where partial correlations that are smaller than 0.07 can be considered as a small effect, 0.17 considered to be moderate effect and 0.33 as a large effect between the two variables. When they adopt Cohens guidelines for interpreting partial correlations, their average partial correlations is considered as showing a small to moderate effect of education on income inequality.

The average reported estimate about the effect of education on income inequality in the meta-analysis by Abdullah et al. (2013) is from 1982. The most recent study in their sample was published in 2010. I chose to utilize the findings of these researchers, and couple it with new research on the relationship between education and income inequality within social sciences between 2010 and 2022. Revisiting and developing an earlier meta-analysis is relevant, because there are observable trends within inequality and education that indicate that the effect of education on income inequality might have changed over time. With this backdrop of divergent findings and an earlier conducted meta-analysis, it is relevant to shed light on the observable effect of education on income inequality in political, economic, and sociological literature today.

1.4 The aim of this meta-analysis

The chosen approach within this thesis can be considered a *descriptive* meta-analysis. A descriptive meta-analysis seeks to describe the existing research population, focusing on how the findings vary due to choices in methods and research design (Stanley and Doucouliagos 2012, 98). A second view, is of meta-analysis as *explanatory,* aiming to make inferences to the population of existing research results, identify the underlying response and developing a "best estimate" to make predictions about the effect of the investigated subject. The first approach may be considered more "rigorous" in its conclusions, since it acknowledges this nature of constructed knowledge (Walsh and Downe 2005, 207), whereas the second has a stronger focus on creating generalized conclusions and predicting effects on the basis of their data. Some researchers argue that the divisions between these meta-analyses are too generic, and that one should strive for a more detailed description of the actual methodological perspective of the individual analysis (Walsh and Downe 2005, 207).

In the meta-analysis utilized in this study, the chosen studies will be treated as a sample of observations of education's effect on income inequality, adopting a focus on this sample as being only the observed effect of education on income inequality, and not the "true" effect between these phenomena. An average effect of education on income inequality weighted by study quality will be calculated, as standard within meta-analysis. There will not be conducted prediction models to identify the "best estimate" for predicting the effect of education on income inequality. Information about the studies will be collected and analysed, predominantly to focus on whether differences between these studies can explain their divergent results.

Having investigated the patterns in the different results on educations effect on income inequality, these findings will be used to discuss what they may imply about the actual phenomena of education and the observed effect on income inequality. For example, if studies using primary school enrolment as a measure for education report systematically higher effect sizes, this forces a discussion on whether we should expect a stronger relationship between primary schooling and income inequality, than with other types of education. In other words, analysing observed effect of a phenomena within a field of science, does open for discussions about the nature of this phenomena.

Further, it must be clarified that this study only seeks to test existing theories on education's effect on income inequality by reviewing existing research, not to establish a casual argument about the effect of education on income inequality. The relationship under investigation is a causal linear argument (Gerring 2012, 226), where increased levels of education is expected to affect levels of income inequality negatively, i.e., reduce income inequality. Findings of this relationship might therefore be interpreted to be either a "confirmation" or "denial" of the existence of such linear causal a relationship. This is not the case, as creating a foundation for casual arguments of even just a probable nature within social sciences requires: finding a credible causal mechanism that connects education to income inequality and ruling out the possibility of reverse causation, covariation and confounding variables, to mention some (Kellstedt and Whitten 2019, 60). This requires many different studies over time, with more than a bivariate focus. Therefore, a descriptive quantitative analysis of existing findings is insufficient to either confirm or discredit the existence of such a causal relationship. Nevertheless, the findings from this thesis can be used to describe the variation of observation on the effect of education on income inequality, how they relate to specific study characteristics, if these findings are in line with dominating theories in this field, and what the systematic differences between the findings may tell us about the nature of the relationship under investigation.

1.5 Research questions

Because meta-analysis differs from many other types of analysis, the nature of this approach must be considered when constructing research questions. As described, an advantage of meta-analysis is that it allows for *proposing an average reported effect of education on income* *inequality* and a closer investigation of *why researchers still debate this relationship*. However, when chosen research subjects are already existing reported effects, one must account for the fact that existing research might not serve as a "random" or sample of the evidence that might exist, but only the evidence that has been published on this topic. This is often referred to as the issue of *publication bias*. To provide a valid conclusion on what reported effect there is between education and income inequality, it is urgent to account for possible bias amongst the sample of research. Measures to handle this issue will be described in the later method section.

Thus, the following research questions are investigated in this thesis are:

(Q1) What is the mean reported effect of education on national levels of income inequality in research today?

(Q2) Is there publication bias in the literature regarding education's effect on income inequality?

(Q3) Are there systematic differences (heterogeneity) between the studies on education effect on income inequality when we control for possible publication bias?

1.6 Scope of the thesis

The chosen research approach utilizes suitable tools to investigate the aim of this thesis, which is to provide more knowledge on the effect of education on income inequality according to scholarly literature. Trying to acquire more knowledge on the proposed subject within the format of this analysis still involves several fundamental limitations that must be highlighted.

The meta-analysis builds on a positivistic methodological perspective. More precisely, it relies on Bayesian logic, where one allows for generalizing on the basis of empirically observable patterns (Knutsen 2019, 266). This is a recognized methodological perspective, especially within empirical and quantitative political science. However, it does leave out the possibility of investigating the role of contextual influences that might have been better investigated with a constructivist methodology (Knutsen 2019, 9).

Because the chosen research subject is existing empirical analysis, we only investigate the reported effect of education within *the universe of published research*. Including all research available on this topic was not possible due to the sheer amount of research that exist and time this would take within the frames of a master's thesis. A sample of published research sample was therefore drawn. Drawing such a sample also leaves out unpublished works, which could have been included to give an even more "holistic" description of knowledge on the phenomenon. However, unpublished data has not gone through the peer review process, and as such is prone to error. Therefore, I limit the scope of the subject to investigating empirical effects of education on income inequality within a sample of cross-country studies from political, sociological, and economic research dating back to 1970. The conclusions from this analysis are therefore limited to saying something about what we perceive to be the effect of education on income inequality in central political, economic, and sociological science over the past 50 years.

This study is also limited regarding the possibility of providing an actual replication of the earlier study by Abdullah et al. (2013). Due to a lack of clear guidelines for replication and efficiency, the scope of this thesis is narrower than the one adopted by these authors. I chose to focus on cross-country studies, leaving out case studies and studies conducted on a regional level. I strive to still use a selection process and method close to the one adopted in the earlier meta-analysis. Despite not being an exact replication of the earlier study, I argue that the data is compatible and that some of their conclusions are still comparable to the ones derived from this analysis.

This is only a brief presentation of the scope of the study. Specific limitations related to the different stages of the meta-analysis approach will be accounted for more closely in chapter 3.

1.7 Research contributions

The findings from this analysis contribute several new insights to the field of research. By compiling large amounts of research, it is detected that researchers report on average, a stronger negative correlation between education and income inequality in research now than 50 years ago, insinuating that the equalizing effect of education has not lost its power or relevance. This finding shed new light on the older development- and economic theories which anticipated a long-term equalizing effect of education on differences in income. Criticism of the traditional development- and economic theories is however also supported by the findings in this thesis, as there is detected a stronger equalizing effect associated with primary schooling than other types of education – meaning that the generalized argument of education "always" serving as an equalizer lacks nuance regarding the different effects that different types of education might have on society. These main findings are of great interest for policymakers engaging in the discussion of what factors might affect income inequality in different societies.

From a researcher's perspective, this analysis provides some central insights into the state of knowledge on this specific topic, and empirical studies in social science research in general. Theories about the effect of education on income inequality from the past 50 years are presented, and differences in reasoning regarding this relationship is highlighted in a unique and useful way for other researchers interested in the theoretical conflicts on this subject. Identifying no practical issues with publication bias in the literature on this topic and detecting heterogeneity between the divergent findings leads to a valid and important investigation of the possible sources of differences between the studies. By further demonstrating how the divergent findings are partly due to differences in estimation techniques, choices of measurement and data material, researchers are provided with useful information about sources of variation on this topic specifically, and sources of variation that might affect research within social sciences in general. Highlighting these systematic differences is an important first step for researchers wanting to design their research as less prone to being affected by systematic differences.

1.8 Outline of the thesis

This thesis is constructed in seven parts: introduction, theory, method, data, analysis, discussion and concluding remarks. In the following theory section, I will present how the research included in this meta-analysis conceptualizes and measure the two phenomena education and income inequality. Further, I'll present dominating theories on why education is hypothesized to effect income inequality and opposing theories. This chapter will illustrate where studies on this relationship differ theoretically and differ in terms of conceptualization and measurement. The opposing theories will be used as a framework for discussing the findings of this meta-analysis. Differences in conceptualization and measurement will be used to investigate in the section regarding heterogeneity amongst the studies. In the third section on method, I will present meta-regression analysis and justify this choice of method to answer the selected research questions. In this section I will also present the steps of meta-regression analysis and highlight the subjective choices that have been made by me during this meta-analysis. In the fourth chapter, I will present the data that has been used for the analysis. Descriptive statistics are used to illustrate the variation between findings over time. In the fifth chapter on analysis, I will conduct the statistical tests and models that are described in the method section. The results of the different tests and models will also be presented, answering if there has been detected any publication bias in the sample, the average effect of education on income inequality one can observe having accounted for study quality, and if some of the characteristics of different studies has caused systematic differences in the reported results. The results are discussed further considering the dominating theories in the field in the sixth chapter. Finally, the most critical findings from this analysis will be used to guide future research in the concluding remarks of this thesis.

2 Theory

The purpose of this section is to provide an overview of dominating theories within the research literature on the relationship between education and income inequality. As the purpose of this thesis is to answer what we know about the empirical effect of education on income inequality, and why there is still no agreement between researchers on this topic, it is essential to outline *what we perceive as definitions and measures of education and income inequality*, and *why researchers believe that these two phenomena affect each other*.

The purpose of this theory chapter is twofold. First, understanding what researchers perceive as measures of education and income inequality is important because social science phenomena are challenging to isolate and measure. The various operationalizations and measurements might however serve as sources for differences in findings about the same phenomena. An understanding of these differences this will be essential to disentangle some of the causes of variation in research on this topic.

Second, investigating the detailed logical reasoning for why these phenomena are thought to influence each other is important to provide a discussion regarding what the findings of this analysis might infer on this topic. As social phenomena are "complex and difficult to unravel" different conditions can be combine in a variety of ways to produce a given outcome (Ragin 1987, 24-25). The ability to say something about the relation between social phenomena depends therefore both on the documented co-variation between the phenomena, and on the strength of the argument about a possible linear relationship between these. Social sciences call for clear and logical reasoning for how social phenomena might interact – as these cannot be easily verified by replaying the same events (Gerring 2012, 1496; Rohlfing 2012). Hence, when investigating a possible relationship between education and income inequality we need to both investigate the empirical findings, together with a theoretical foundation for *why* such a relationship can be expected. The theories presented in this chapter will serve as background to discuss are the findings in line with traditional theories on the subject,

The following outline of *how we perceive and measure education and inequality* together with *existing theories about this relationship* is based on central literature on the subject, including the 66 articles within this meta-analysis. First, I outline different definitions and operationalizations of the two phenomena. Second, I will highlight proposed causal mechanisms between the two phenomena, to uncover the difference between a variety of explanations on how education and income inequality might interact. The proposed causal mechanisms are divided into two parts: the arguments for *why education should have a negative/equalizing effect on economic inequality* in section, and the arguments for *why education would have no or a positive effect on the levels of income inequality.* As an extension of presenting theories about how these phenomena might interact, I will highlight the limitations to investigating such casual relationships in social sciences.

2.1 Defining and measuring education and income inequality

It is essential to define the subjects of investigation, as any cumulation of knowledge depends upon reaching an understanding about what to call a thing and how to define it (Gerring 2012, 114). An empirical concept within the social sciences is usually divided into a term, attributes, indicators, and phenomena (the referents, extension or denotation of a concept) (Gerring 2012, 116). I will describe different views of the term, phenomena, chosen attributes and indicators of *education* and *income inequality* in the studies that are included in this meta-analysis in the following.

This section has been created based on the research that has been sampled for this meta-analysis, and by including additional central literature about income inequality and education. The process of creating this section can be viewed as part systematic and part unsystematic process, including both making inferences based on inductive and deductive logic. Defining the concepts of education and income inequality is important in meta-analysis for two reasons. First, one must identify what the included studies and additional literature mean by education and income inequality to understand the variation amongst existing studies when it comes to definitions and measurement, so that these differences can be used later in the analysis to control for heterogeneity. Second, an explicit determination of what is perceived as the scope of the phenomena *education* and *income inequality* within this specific thesis is necessary for other researchers seeking to understand why other research is excluded from this meta-analysis.

2.1.1 Education

A natural linguistic understanding of the term "education" refers to the systematic transfer of knowledge, in formal environments such as schools or universities. This is often seen in opposition to various nonformal and informal means of socialization (Nakosteen 2021). The phenomena education has existed in different forms through history and can be thought of as the "transmission of the values and accumulated knowledge of a society" (Nakosteen 2021). Education is seen as a method for improving the "skills" or "human capital" of an individual (Tselios 2008, 409; Carvajal and Geithman 1978, 924; Keller 2010, 72; Alderson 1995, 683; Ahluwalia 1976a, 322). This is because acquiring skills increases the individual's ability to solve tasks that they could not solve, or not solve as effectively, without these specific skills. Since the enlightenment, educating people has by many been considered the path to granting *individual freedom* through giving people the ability to function as critical thinkers, *to uphold democratic regimes,* depending on people devoted to citizenship, and to *financial self-development*, through work and the accumulation of skills requested on a within a market economy (Round 2004, 291; E. Glaeser and Shleifer 2007, 77; Dewey 1944, 328; Mill 1963-1991; 2002, 252-256).

There are several attributes to the phenomena of education which is highlighted in the literature: *teaching, learning*, and *testing*, skills like *reading, writing, calculating*, etc. on various levels and thorough various means. These attributes mirror the concept of a process for *acquiring skills*, which enables individuals to solve different tasks. Educational levels are usually divided into general literacy, primary-, secondary-, and higher education.

Within all the selected research in this meta-analysis, education is seen as a graded variable, with countries having a smaller or larger degree of education. Indicators used in the selected articles for measuring the attributes of this education are: literacy rate; initial schooling; primary- secondary- and high school enrolment; primary- secondary- and high school attainment,; median school years completed; mean years of primary-, secondary- and tertiary schooling; average years of schooling; percentage of population with higher education; variance of educational attainment; standard deviation of school years completed; educational inequality and distribution of years of schooling. These indicators are used to measure quantity of education, not quality, and some can be categorized as stock variables (measured at one point in time), other as flow variables (measured over a period). Other researchers use related proxies of education, such as public funding of education or financial support for students, under the assumption that these factors are so closely related to the increase of human capital that they serve as measures of education. Such proxies are not accepted as valid translations of the phenomena education here, as we strive to identify the observable effect of increasing actual levels of education in the population, and not measures taken to increase these levels. Other popular translations of education are indexes that are constructed to measure the level of education, with the purpose of capturing specific aspects, details, or a more holistic view of the given phenomena. For example, Psacharopolous (1977) measures educational inequality within a given country by using the coefficient of variation of enrolments by school level and Tsakloglou (2006) uses a weighted index of the educational level of a country, giving one third of its weight to the female primary education enrolment ratio and two thirds to the female secondary education enrolment ratio (Tsakloglou 2006, 519; Psacharopoulos 1977, 387). As the relevant indexed are not seen as proxies, but different and more complex ways to capture the changes in actual educational levels. Due the low level of indexes and their very specific design, these are not controlled for in the later analysis. This is because of issues with multicollinearity from including to many variables in the model, so that only a selection of the measures thought to create significantly systematic differences are included.

Based on the scope of the term education that has been laid out here, some studies have been left out based on their definition and measurement of education. The phenomena *education* as it is defined in this thesis can still be considered as broad, due to the purpose of the analysis: which is to investigate a large sample of research that is assumed to investigate the same phenomenon in different ways. As the different measurements in the collected research tap into different aspects of education, the described differences in measurement will be controlled for in the later analysis. This will allow for investigation into whether measurement variances cause research to generate systematically different results on the relationship between education and income inequality. All reviewed research has been classified by their measurement of the two variables. Because the most common use of measurement for education are primary- secondary- and tertiary education enrolment or attainment, these are types of measurement used to investigate possible correlations between findings and measurement in the analysis chapter 5. Before investigating these measurement differences more closely, the phenomena of income inequality must be properly identified, using the same inductive and deductive approach as in this section.

2.1.2 Income inequality

The term "inequality" refers to a deviation from an equal distribution of some sort, often used in relation to a resource or right. "Income" is commonly referred to as the "compensation of labour" (L'evy 2004, 2). It is often used synonymously with "earnings" which equals the wages times the amount of time an employee spends working (Partridge, Mark D. Partridge, and Rickman 2007, 282). Income can also be more broadly defined, as *the acquisition of financial resources that gives one the possibility to consume or invest* (Weinberg 2001, 2). Such financial resources can be a person's return from working, providing a service or capital investment. The scope of this term is usually draw in opposition to *capital ownership*, which referrers to the pure ownership of capital goods (L'evy 2004, 2). However, the term "income" can have different meanings depending on the context (Bojer 2021; Mechling, Miller, and Konecny 2017, 30). It is common that "income" is used to describe wages, income from savings and investments, private transfers, and public transfers, through rights and benefits – and in regards of financial resources before or after taxation. There is no clear standard translation of this term into financial measurement.

A central attribute to the term, is that it refers to the disposable resources for individuals within a market economy. This can be measured both by looking at expenditure (consumption) or the possession of such financial means (income). One might argue that by not including consumption, one provides a skewed picture of the utility function of individuals (Salverda and Smeeding 2009, 42; UNU-WIDER 2021, 4). However, when looking at the variation in access to control economic resources, income is a preferred measure. Additionally, there is also more available data on inequality and poverty in reference to income, compared to consumption (UNU-WIDER 2021, 4; Deaton and Zaidi 2002, 1).

Within econometrical research, income is commonly measured with the variable *household income*. Gross household income in a given year is the sum, across all household members, of labour market earnings from employment or self-employment, income from savings and investments, incoming private transfers such as receipts of gifts or alimony, and public transfers such as social insurance or social assistance benefits (Salverda and Smeeding 2009, 42). There is also no agreed basis of definition of "income" as in the case of national accounts data, but some steps have been taken towards developing international standards concerning this (UNU-WIDER 2021, 4). It is however important to note that current standards on measuring income inequality are relatively new, and that there is still large variation in the included data material from the 61 studies due to different definitions of income.

Regarding measurement, there is no best way to summarize income inequality graphically (Salverda and Smeeding 2009, 46; Singh 1984, 251). Common representations of income inequality are *Kernel density estimates* for illustrating the different fractions of a population with an income range, *Pen's parade quantile function* to highlight the presence of large incomes or the *Lorenz curve* to highlight the distance between a perfectly equal distribution of income and the actual distribution. To measure inequality by a single number, one might choose between the different indices constructed to summarize inequality (Salverda and Smeeding 2009). These indices build on different perspectives on how to best summarize distributions in a meaningful way (Singh 1984, 250). The different measures of income inequality within this meta-analysis are *Gini coefficient, percentile ratios, Atkinson index, Theil index, Mean log deviation of household, Lowest income share (growth of), Income variance* and *Income ratio.* I will explain the characteristics of the most common measures on income inequality included in this study in the following. Because most of the studies use Gini index or percentile ratios, these are the measures chosen to control for in the later modelling of differences between the studies.

Gini coefficient

The Gini coefficient is referred to as the most popular inequality index (Salverda and Smeeding 2009, 50; Holasut 2020, 2). Of the selected studies, 76 % employed at least one variation of the Gini coefficient as a measure of income inequality. The coefficient ranges from 0 (perfect equality) to 1 (perfect inequality), showing the ratio of the area enclosed by the Lorenz curve and the perfect equality line to the total area below that line (Salverda and Smeeding 2009, 50).

A central strength of the index is its ability to summarize inequality of the entire income distribution by using a single statistic that is easy to interpret, allowing for comparison between countries with different population sizes (Holasut 2020, 2). Further, it allows negative values for income and wealth, meaning that negative wealth (debt) could drive the Lorenz curve below the x-axis. However, the Gini coefficient also has limitations, as, it takes all the data from the Lorenz curve and converts it to a single number. Thus, two countries with very dissimilar income distributions can have the same Gini coefficient, meaning that a substantial amount of information on income distribution amongst the population is lost in the conversion to a graph (Holasut 2020, 3). As a means of controlling for some of this hidden variation, many researchers supply the Gini coefficient with percentile ratios (Shorrocks 2005, 1).

Percentile ratios

One can analyse income inequality across countries by using inter-decile ratios, comparing the income share held by different percentiles of the population. This measurement of income inequality is used in 25 % of the studies included in this sample. It is normal to compare the income share held by a smaller percentage at the top, with a larger percent at the bottom, for example the income share of the top 10% with the income share held by the bottom 60% (Ahluwalia 1976a; Alderson 1995; Breen and García-Penalosa 2005; Beck and Asli Demirgüç-Kunt 2007). The data used to calculate these inter-decile ratios are regularly updated and reported along with the Gini Index by international organizations such as the World Bank, the OECD and the Human Development report Office as measures of inequality (Holasut 2020, 3). However, while these inter-decile ratios seem easy to understand, they might be hard to interpret. Sitthiyot and Holasut argue that "from a mathematical and practical point of view percentile ratios are more difficult to interpret and compare among countries, since they have no upper bound relative to other inequality indices whose values are bounded" (Holasut 2020, 3). This is said to make the values of these indices more "tangible to human perception" and ignores the income of those in the middle percentiles of distribution (Holasut 2020, 1). Further, it is not always clear what ratio should be view as the "lowest share" of income, or the "top share of income", as this depends on the individual distribution within each country. To code the different percentile ratios in this analysis, the bottom share is counted as the bottom 40%, the next 40% as the middle, and any percentile within the top 20% as a measure of the richest part of the population. If studies investigate the share of the bottom 80% of the population, then this is considered a measure of the income share of bottom- and middle class.

Based on the substantial variations in definitions and measurements of the term "income", a wide definition of income inequality is adapted within this analysis. Essential literature on this phenomenon illustrates how widely this term is when used within social sciences, and it would therefore be harmful to the meta-analysis to adopt a narrower definition of this phenomena than what is common within the research field itself. All the used measurements from the selected studies are accepted as logical translations of the phenomena income inequality as it is investigated here. Measurement differences and database differences are instead coded to account for the existing variation amongst the studies in their operationalization of income inequality.

2.1.3 Using the different measurements to understand heterogeneity

For social scientists it is essential to be explicit and descriptive regarding the phenomena they are seeking to measure within quantitative research. Without the important steps of conceptualization, measurement and aggregation, there is no way to control the results and guarantee that they are comparable to each other (Munck 2002, 15-22). As illustrated in this section, well-known and much-researched phenomena might be measured in different ways, tapping into different aspects of the phenomena. In this analysis, central differences in a study's chosen measurement will be used to model heterogeneity. Based on the presented ways of measuring education, the variables included are *primary-, secondary-,* and *tertiary school attainment or enrolment*, *literacy rate* and *average years of schooling*. Based on the presented measures of income inequality I chose to control for study using *Gini coefficient*, *income ratios*, and *ratios for income share of bottom, middle,* and *top ratio,* adding an extra variable for studies that investigate the ratio share of both *bottom and middle class*.

More detailed coding could have been conducted, especially on the measurement differences of income inequality. A limited selection of coding is here preformed as the more detailed differences between the studies (whether they include data based on income after or before taxation, whether the data is produces under circumstances etc.) is hard to find for each study, as these details are not sufficiently transparent in many of the research articles. The choice of limiting the differentiation of ways to measure income inequality is also done to avoid issues with multicollinearity within the later models, and because only a smaller selection of variables is seen as theoretically essential to answer the main research questions.

Having described how the phenomena of education and income inequality is perceived within central research in economic, political, and sociological literature, it is essential to outline how researchers perceive that these phenomena interact.

2.2 On investigating a bi-variate one-way causal argument

In the following sections, theories about why education is expected to have an equalizing effect on income inequality is presented. The proposed relationship that is investigated is a one-way causal argument regarding the effect of one variable on another, which is arguments that are hard to verify within social sciences. As mentioned, establishing probable causal relationship within social sciences is challenging. Social sciences can be characterized as a scientific field that investigates phenomena that exists as products of aggregated human behaviour. Because there is an almost infinite variation of circumstances to consider when investigating these phenomena – making generalized statements become a complicated matter. The complexity of many of these social phenomena's means that they are hard to isolate, and causal arguments regarding how they interact is very challenging to test and reproduce. This leaves us with problems of reversed causation, collinearity, simultaneity, endogenous variables, to mention some.

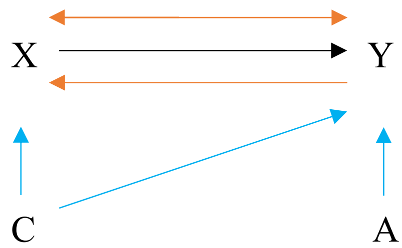


Figure 2.2 Examples of possible hidden relationships affecting one-way linear causal arguments.

The black arrow is the causal relationship under investigation, considered to move in a one-way direction between two variables. The red arrows pose as possible directions of influence between these factors. The blue arrows are examples of intervening and hidden influence from additional factor.

The issue of reversed causation stems from the fact that changes in education levels can be influenced by an increase in income inequality. Since controlled randomized experiments are almost impossible to conduct within social sciences, longitudinal panel data has become the preferred option for many researchers to control for this issue (Leszczensky 2019, 837). This type of study allows for investigating changes in specified units over time, reducing the risk of confounding variables and making it easier to observe the casual direction. Many of the studies included in this meta-analysis use panel-data.

Not all studies apply panel data to investigate the effect of education on income inequality, as some of them only investigates the effect of education on income inequality, without aiming to provide findings for a causal argument. The question of the probable effect of education on income inequality could be answered by investigating if countries with higher levels of education also has lower levels of income inequality, or if countries that experience an increase in education also experiences a following decrease in income inequality. Only the latter research design would be sufficient to identify an actual effect of education on income inequality in a one-way direction, but both research designs can be adopted to investigate the possible correlation between these two phenomena. A study of whether countries with higher levels of education also report lower levels of income inequality, build on the assumption that differences between countries might be interpreted as changes within countries, so that the findings of a given time can be generalized to another point in time. These studies are often criticized for being prone to selection bias (as the group of selected countries would highly influence the results) and because simultaneous studies also have trouble dealing with other simultaneous effects, known as the issue of multicollinearity (Lauridsen and Mur 2006, 318). Using meta-analysis, Broderstad (2015), investigates the effect on income on democracy, and highlights differences between simultaneous and studies that investigates effects over time (Broderstad 2015-06-02, 11). One way to separate the studies looking at temporal causal effects and those measuring simultaneous effects, is by controlling for the use of *lagged variables*. For instance, researchers wanting to uncover the effect of an increase in tertiary school enrolment on income inequality, would need to lag the income inequality variable with the expected years that individuals would have to spend finishing this education, and adopt a reasonable prediction for when the effect of this education can be expected to be observed within income statistics. Within this study, both research that investigates a simultaneous relationship and research that investigates a later effect of education and income inequality is included in the sample, as they both can shed light on the relationship between education and income inequality. By controlling for studies using *lagged variables* one can still investigate if there are systematic differences in the results produced by studies with simultaneous or lagged effects.

Similarly, omitted variable bias can be problematic in this field of research. This entails that there might be underlying factors that causes heterogeneity that are not included in the model. Estimation methods deals differently with the issue of endogenous variables. Ordinary least squares (OLS), Generalized methods of moments (GMM), Instrumental variables (IV) and Stage last square (SLS) are all estimation methods that deals with heterogeneity in varying ways, OLS being the most common within this sample. OLS is known for being unsuitable to distinguish if variation in findings are due to variation between groups or within group. This can however be controlled for by adopting fixed effects models. Controlling for studies that uses OLS with fixed effects is therefore relevant for the later investigation of the divergent results on the effect of education on income inequality.

As presented earlier, the subject of research is an empirical analysis of the relationship between education and income inequality. This quantitative research can only be used to test the hypothesized relationship between the two, to see if we can find variation that give reason to believe that income inequality varies in relation to education. Thus, this study is inappropriate as a foundation for a casual argument, but suitable for testing whether we can observe empirical tendencies that are in line with the existing causal arguments. It must therefore be kept in mind that what is studied in this thesis is not the actual effect of education on income inequality, but the *observed effect of education on income inequality in a sample of social science research over the past 50 years*. The causal arguments proposed by central researchers on this topic is presented in the following sections.

2.3 Arguments in favour of education influencing economic inequality

Why would an increase in the number of people getting different levels of education affect the levels of income inequality? The relationship between the distribution of income and the process of development is one of the oldest subjects of economic enquiry (Ahluwalia 1976a, 307). The nature of changes in economic inequality, economic growth and different types of development has been extensively discussed (Ahluwalia 1976a, 130; 1976b, 128). Still, researchers argue that there is nothing more than "introductory work" on the topic of the development process and reasons for distributional differences in income. This makes it harder to determine *what* regarding the process of capital accumulation in a society – often referred to as economic development – that can make the personal distribution of income become more equal (Heins 1967, 175; Chiswick 1968, 495). As presented, education is often assumed to play an important role in reducing income inequality (Rodríguez-Pose and Tselios 2009, 414-415; Sylwester 2002, 43). There are several different theories about just *how* education affects income inequality, and I have divided these into two main strands of literature: *theories on economic development* and *theories on human capital accumulation.*

2.3.1 Economic development and Kuznets' inverted U-shape curve

A general and much used starting point to assess the determinants of inequality involves some version of the Kuznets curve (Kuznet 1955, 1963; Barro 2000, 9; Ahluwalia 1976a, 308). Based on data from around the beginning of 1900- until 1950, Kuznets argued that the secular behaviour of inequality follows a U-shaped pattern, with inequality first increasing and then decreasing with development (Ahluwalia 1976b, 128; Kuznet 1955; Silva 2007, 122; Park 1996, 51). Kuznets focuses on "development" as the *movements of persons from the agricultural and rural sectors to the urbanized and industrial sectors*.

A central premise for this theory is that the agricultural and rural sector initially constitutes the bulk of the economy, featuring low per capita income and relatively limited inequality within this sector, whereas the creation of a small industrial and urban sector starts out small, with higher per capita income. Kuznets then predicts that the movement of people to the higher paid urban and industrial sector increases the economy's overall degree of inequality (Barro 2000, 8; Kuznet 1955, 1963). As the size of the agricultural sector then is suspected to decrease, the main effect on inequality from the continuing urbanization is that more of the agricultural workers are enabled to join the higher paid industrial sector. More supply of workers in the growing urban sectors is hypothesized to drive the level of wages down. The decreasing size of the agricultural labour force is expected to drive up relative wages in that sector. Combined, these forces are though to reduce overall inequality. Hence, at later stages of development, the relation between the level of per capita product and the extent of inequality tends to be negative. Using indices measuring economic inequality graphically, one would then expect to see a an inverted-U, which is the curve named after Kuznets. The U-curve can be interpreted as: inequality first rises and later falls as an economy develops (Barro 2000, 9).

Since Kuznets published his theory on the U-shaped curve of inequality, numerous authors have tried to check that inequality tends to be low in less poor countries, increases in middle income countries and decreases again in richer countries (Bourguingnon and Morrisson 1990, 1113). As presented, the evidence are still far from conclusive (Bourguingnon and Morrisson 1990, 1113). Aigner and Heins state that "there is no formal theory available which satisfactorily concerns itself with the relationship between development and the inequality of incomes" (Heins 1967, 1761). In his work, Kuznets indicates that we lack "a firm set of links between the observable changes in the production structure that constitute economic growth and the observable associated changes in the income distribution" (Kuznet 1963, 2; Chiswick 1968, 495). In Kuznets discussion of the relationship between the level and inequality of income there is no explicit consideration of the effect of training or schooling on the distribution of income (Chiswick 1968, 495; Kuznet 1955, 1963). Instead, several authors have implemented the role of education in different interpretations of the Kuznets curve.

Barro highlight how education enables the shift of people from sectors using old technologies, towards richer sectors using more novel and advanced technologies (Barro 2000, 9). Mobility between these sectors "requires a process of familiarization and re-education" (Barro 2000, 9).

Education is viewed as an essential "bridge" in the development that Kuznets predicts - enabling the movement of people from the low-paying agricultural and rural sectors to the higher paying urban and industrial sectors – as these sectors tend to require trained or skilled labour.

Nielsen and Alderson (1995) include Kuznets curve as a part of their theoretical model of three stages of development, where they also include assumptions about an expected population growth. According to their model, the shift of the labour force out of agriculture produces an inverted-U shaped trajectory of inequality in line with Kuznets theory, where simultaneously the population growth should increase and decrease again due to demographic transition – contributing to the inverted-U shaped trend of inequality, with the changes in the supply of unskilled labour. Finally the spread of education should produce a secular trend of declining inequality by increasing the supply of skilled (and higher paid) workers (Alderson 1995, 676).

Ahulwalia argues that, in addition to the driving forces of inequality between the two described sectors, there are two long-term forces operating to reduce inequality within the modern sector: the cumulative impact of *an extended education system* and *long-established modern sector* that creates a highly trained labour force with a more equal dispersion of skills, generating both an increase in the share of wage income as well as greater equality in its distribution. This tendency should further be strengthened by improvement in labour organization (Ahluwalia 1976b, 130-131). Ahulwalia creates a link between equal access to education and political involvement, adding the possible equalizing power of labour organizations. One could refer to this as a "political human capital argument", meaning that an increase of skills for the individual through education is expected to affect their ability to participate in the political arena, and influencing income distribution in several ways.

According to the different lines of argumentation presented here, education should have a negative effect on income inequality in different countries, due to the transmission of workforce moving from rural to urban sectors, followed by an equalizing effect by providing more general opportunities to participate in the urban and better paid sector, which in turn will drive the general level of wages within this sector down. This implies that improving literacy rate, increasing the amount of population with primary-, secondary-, and higher education will have a long-term equalizing effect, depending on what level of skills the urbanized and technologically developed sectors require.

Kuznets curve, and the presented interpretations of this theory, all focus on the long term relationship between development and inequality, generated by long term changes in economic structure (Ahluwalia 1976b, 130-131). There are, however, many authors focusing on the more short-term forces which might promote or prevent income inequality, independent of the long-term phenomenon of development. One general strand of theory focusing on how education reduces income inequality in a short-term perspective, is the theory of human capital.

2.3.2 Theory of human capital accumulation – The human capital hypothesis

Mill (1848) was one of the first to predict that the spread of education in a population would entail a decrease in inequality (Mill 1963-1991, 2002; Alderson 1995, 683). This belief became widely shared among scientists, and increased educational attainment in a population became an aspect of the general process of human capital accumulation (Alderson 1995, 683). Following the development of human capital theory in the early sixties, education became a popular independent variable in income distribution studies (Psacharopoulos 1977, 383). It is often referred to as an explanation for why educational attainment of individuals or groups should have a significant effect on income distribution (Ismail 2000, 20).

The theory of human capital is often referred to as a "neoclassical" economic argument (Silva 2007, 129; Tsakloglou 2006, 519). It builds on the assumption of people's ability to make rational choices for themselves and the forces of supply and demand in a job market. The core of the theory is that "increased availability of qualified personnel increases competition and produces a relative decline in the higher wages and salaries" (Lecallion and Felix Paukert 1984, 88; Alderson 1995, 683). In other words, the human capital hypothesis assumes that technologies in production over time will call for more skilled workers, leading to a fall of capital share and a rise in labours share of income (Piketty 2014, 21). This logic resembles the market trends that are assumed by Kuznets.

According human capital theory, earnings are considered to be largely a "flow of discounted returns to investments in human capital" (Nord 1980c, 196; 1980a, 866). This makes the concentration patterns in human skills a cause of income inequality in line with the concentration of physical assets (Ahluwalia 1974, 81). The level of education can be seen as a proxy for the "level and rate of improvement of human resources", and according to this theory, higher educational levels for the 'poor' increases their marginal productivity, therefore also their wage rates and income share (Tsakloglou 2006, 519). The human capital approach suggests that while inequality in concentration of education contributes to income inequality, expanding education "of the right type" to the lower income groups increases their productivity and thus wages, thereby improving income distribution, and the marginal product of labour remains high despite the increased supply of skilled labour (Ahluwalia 1974, 81-82; Keller 2010, 52). In formula, Becker (1964) and later Mincer (1974), proposes a model where (log) incomes and years of education should be linearly related (Checchi 2001, 24). The theories on human capital accumulation through education has resulted in many scientists making an empirical assessment of the impact of schooling on income inequality (Nord 1980c, 196; Ram 1981, 254; Motonishi 2004, 474; Ismail 2000, 20).

Ahluwalia elaborates on the technological assumptions about factor productivity, arguing that the human capital hypothesis assumes that there is a substantial scope for substituting skilled for unskilled labour in the production process, without a decline in the marginal productivity in the former (Ahluwalia 1976a, 321-322). He argues that the shift from low-paid unskilled employment to high paid skilled employment due to education, creates an increase in the share of total wages output, which is likely to be more equally distributed compared to other capital intensive patterns, because “there is a limit beyond which human capital can be accumulated in a single person” and that it cannot be inherited across generations in the same manner as physical capital (Ahluwalia 1976a, 322). This should ensure a more equal distribution of income across the population.

In the same manner as with Kuznets hypothesis, researchers have added a political argument to the argument on how accumulation of knowledge through education should have an equalizing effect in society over-all. Tsakloglou argues that one can expect that higher educational levels increases self-esteem and hence leads to an “increasing claim for equality” (Tsakloglou 2006, 519). This is also an interpretation of education as a fuel for political engagement and collective action through political parties or labour organizations.

The core of the human capital theory is that education will give individuals the possibility to apply for jobs which requires these skills, and that this consequently will affect income distribution (Ismail 2000, 19-20). While it resembles Kuznets premises, human capital theory is still more detailed on the specific markets reactions that one should expect to see with educational development. It is also less focused on making a generalized statement about the possible outcome for all countries over a large span of time, and more related to how decisions of individuals under specific circumstances might produce more equal outcomes due to specific economic factors in the labour market. The human capital approach has become popular within public policy, especially since promoting education is a widely accepted as an area for state involvement (Ahluwalia 1974, 81). Within cultures with a meritocratic ideal regarding distribution of wealth, the human capital approach is central for developing policies, where the possibility to invest in education should result in wages that reflect one's effort and talents. Adopting the human capital hypothesis, one should expect a negative correlation between education and income inequality, observing education as an equalizing factor.

2.4 Arguments in favour of education having no- or a positive effect on income inequality

There are many answers to why educations should have no or a positive effect on income inequality. Most of these theories are based on two different strands of argumentation: *criticism* of the earlier presented theories about why education should affect income inequality and *theories proposing other factors that should be more influential than education for determining income inequality*, hence reducing the value of education as an explanatory variable for understanding distribution of income within countries. This chapter consists of a selection of criticism based on the included research in this meta-analysis, and by screening other central literature on the subject. The section opens with common criticism towards the presented theories Kuznets curve and human capital hypothesis, followed by theories about competing or other influencing variables that should have stronger explanatory power than education when it comes to income distribution on a country level.

2.4.1 Criticism of the presented theories

<u>Too broad and too general</u>

A major part of the criticism towards the traditional development and human capital theories is the simplistic or all 'other things being constant' assumption, because the equilibrium of economic development might be more detailed and less obvious than these traditional theories put forward (Checchi 2001, 6). Several authors highlight the lack of focus on labour market imperfections as a central weakness within the development and human capital theories regarding the effect of education on income distribution.

In addition to being a widely used theory, Kuznets's curve can be considered a much disputed topic (Breen and García-Penalosa 2005, 381). Being a general and broad theory on the universal development of countries, central criticism addresses that it leaves out important circumstances. Bourguingnon argue that "even if evidence [of Kuznets curve] were not disputed, there would be no reason to believe a priori that a cross-sectional statistical relationship should necessarily transform into a longitudinal relationship [...] implying some kind of iron law relevant for any country, at any period and under any circumstances" (Bourguingnon and Morrisson 1990, 1113).

One of the major critiques towards using Kuznets curve for cross-sectional analysis is that it ignores each country's freedom in shaping their income distribution through redistribution policies, and their different distribution paths in the course of its development process, which can be very unlike the ones of already developed countries (Bourguingnon and Morrisson 1990, 1114; Husnain, Haider, and Khan 2021, 11477). Further, the Kuznets curve is said to be too complex to be measured using GDP per capita, ignoring labour market imperfections in the given assumption regarding rural-urban migration (Bourguingnon and Morrisson 1990, 1113) and leaving out other macroeconomic variables associated with the distribution of income, such as civil freedom, financial development and the distribution of land (Breen and García-Penalosa 2005, 381; Li, Squire, and Zou 1998, 42).

<u>Not accounting for return on human capital</u>

Both Kuznets curve and the human capital has been criticized more specifically for ignoring the fact that the effect of increased average schooling on income inequality may depend on *the rates of economic return to education* (De Gregorio and Lee 2002, 395; Chiswick 1968, 496). Return on education is the economic "payoff" individuals receive through education experience in the job market. In other words, the relationship between the number of years of schooling and income earned. Chiswick (1968) argues that one cannot adopt the hypothesis of a clear relationship between education and income distribution without accounting for the average rate of return from schooling and the inequality in the distribution of years of schooling attended (Chiswick 1968, 496). Lee and De Gregorio (2002) stresses that an expansion of education in an economy with unclear/unstable return on education may become more unequal in terms of income distribution, only benefitting some of the individuals who choose to invest in education (De Gregorio and Lee 2002, 397).[2] By including a model of return on education within the traditional models for levels of earnings according to traditional human capital models they illustrate that an increase in schooling can reduce income inequality, if the covariance between the return to education and the level of education is negative (De Gregorio and Lee 2002, 397). However, their model also shows that with other variables held constant, and with independent rates of return on education and schooling level, an increase in the level of schooling will lead unambiguously to a more unequal income distribution.

Still, measuring the return on education is a much-contested subject. Trostel (2005) argue that even though it has been estimated in large amounts of studies, the majority of this work is criticized for assuming that the marginal rate of return is constant over all levels of education (Trostel 2005, 191). This might be considered less logical, as many researchers stress the different consequences of demand and supply in the labour market relating to the various types of education dominating the workforce. It has also been criticized for being unable to account for the different types of education and educational systems (Bernasek 2005). This leads us to two other essential variables in the critique of traditional development and human capital theories, which is their lack of *differentiation between different types of education*, and not accounting for *unequal access to education*.

<u>Ignoring the differences between different types of education</u>

The problem with *not differentiating between different types of education*, is that the central assumption of these theories regarding supply and demand in the labour market might not be as straight forward as higher levels of education leading to a wage compression and lower rates of income inequality. Many researchers' favours increasing levels of primary and secondary education as a way to attain less skewed income distributions and postulates that increasing higher education might lead to *more* income inequality (Sylwester 2003, 250; Rodríguez-Pose and Tselios 2009, 360; Gruber and Kosack 2014, 253). Gruber and Kosack (2013) argue in their study on inequality in African countries that increased tertiary education amongst parts of the

[2] The term «investing» in education is here used to refer the investment of time and effort spent outside the job market to improve human capital through education. The term "investing" in education can therefore be applied within educational systems that are also free of charge.

population creates a “tertiary tilt” where rising enrolments is associated with higher inequality in the future (Gruber and Kosack 2014, 262). What kind of “composition” and “wage compression” effect different types of education has also depended on the demand and supply of the labour market, as described earlier. Using human capital- and development theories to justify for example public spending on the access to higher education, might not be justified by educations general “equalizing effect” – as it might only benefit a small part of the population.

Unequal access to education

Access to education might depend on the resources across the population. Researchers highlight that, educational choices are often shaped by whether education is a private investment or publicly founded. It has been shown that households with a higher-than-average level of education tend to invest money received from remittances in the accumulation of human capital and durable assets, increasing their capacity to generate higher incomes in the long term (Leon 2007, 136). Thus, income inequality may prevent access to education when education is too costly for the family: the more skewed the income distribution, the higher the population share excluded from schooling and the higher the inequality in educational achievements, creating what Checci (2001) calls “a self-perpetuating poverty trap that can only be avoided by easing access to education” (Checchi 2001, 5; Tselios 2008, 409). This implies that improving the access to education for the lower-income households will reduce income inequality (Checchi 2001, 5). To further elaborate on the difference between public- and private educational systems, Glomm and Ravikumar (1992) models the different scenarios of people choosing to invest in education in a public versus a private educational system and predict possible outcomes for inequality within the two systems. They argue that while it is relative if income inequality declines under a private education system, income inequality “unambiguously declines under a public education system” (Sylwester 2002, 43; Glomm and Ravikumar 1992, 819). This is supported by other studies, where expansion in public education is seen to lower the levels of income inequality over time (Verdier 1993, 400; Zhang 1996, 387; Sylwester 2002, 43). Sylwester argues that many researches have created models showing agents must have sufficient resources to even consider education as an opportunity, if income inequality should decrease (Sylwester 2002, 43).

Lack of focus on geographical differences

Another argument is that the differences in geographical access to education and the different jobs in the labour market will affect the shape of the relevant supply and demand curves (Tselios 2008, 405-406). Regarding the effect of geographical location on the job market, urban economists theorize that individuals' localization choices will depend on differences in productivity and amenities, meaning that areas with industries that are particularly good for low- or high-skill workers should disproportionately attract those workers (Edward Glaeser, Matt Resseger, and Tobio 2009, 632). This might create areas where the return on education varies, and that the supply and demand of educated labour might not be a linear relationship correlated with economic development, as proposed by Kuznets. Nord (1980) illustrates such local differences by arguing that education appear to produce an equalizing effect on the distribution of income in median sized cities by reducing the adverse effects of the schooling and experience variables, but not necessarily in smaller cities (Nord 1980b, 508). Regarding these local and regional differences in the labour market, these differences have certain spillover-effects that might affect income inequality, due to cultural, social, political, and institutional factors (Tselios 2008, 404). It is also argued that concentration of education within areas might create local spill-over effects, by individuals seeing that others investments in education has economic payoff, creating an incentive to make the same investment (Tselios 2008, 404). Some researchers argue that different access between urban and rural areas are so large when it comes to income inequality, that they should be studied as two separate phenomena, hence focusing on local inequality instead of national inequality (Edward Glaeser, Matt Resseger, and Tobio 2009, 631). The differences between inequality on national and regional levels are not controlled for in this analysis, as the focus is toward the effect of education on income inequality on national levels. However, it is worth mentioning that the effect of education is hypothesized by many to have different effects on regional level from national level, illustrating how the result from this analysis is limited to say something general about the expected effect of education within countries, as it ignores the regional and local level. Only focusing on a national level is a clear limitation to this thesis, and a limitation to Kuznets curve, as will be presented in the following.

<u>The national focus</u>

The last market imperfection that might affect the traditional view of education as an equalizer, in line with Kuznets theory on economic development or the theory of human capital, is the impact of a global job market. Brown and Tannock (2009) is among the researchers claiming that due to the neoliberal view of market competition, there is a rise of a "global war of talent", were the intricate global labour market challenges the anticipated linear relationship between "learning and earning" (Brown and Tannock 2009, 377). They argue that the competition on the international labour market might only benefit the few and talented (Brown and Tannock 2009, 388). Adapting the perspective of individuals as agents on an international job market instead of a national job market, this raises the questions about whether the expected "composition" and the "wage compression" effect of education that the traditional theories build on might not work the same way as within a national job market. One might theorize an international job market where the most educated people decide to move to countries with more developed and higher paying technological sectors, and therefore not creating the shift of labour from low-skilled to high skilled-sectors within their national economies. This challenges the idea of the Kuznets curve within all nations, and partly the supply and demand mechanisms within the human capital hypothesis.

2.4.2 Other forces influencing the distribution on income

In addition to market imperfections, a large part of the critique towards the view of education having an "equalizing" effect on income inequality is that it underestimates other factors that might have stronger or interfering effects on income distribution. Amongst the proposed factors influencing income distribution are *welfare regimes/distribution politics* (institutions), *population growth* and *economic growth*.

Institutional factors

The degree of welfare regime within a country is an institutional factor thought to affect supply and demand curves in the labour market, based on its terms of the inheritance of property income, the equality of opportunity, the distribution of abilities and the subsidies given to education (Rodríguez-Pose and Tselios 2009, 405; Tselios 2008, 403; Edward Glaeser, Matt Resseger, and Tobio 2009, 645; Ahluwalia 1976a, 309; Zou 2000, 154). Welfare regimes are often classified based on their extent of redistribution politics, from more to less extensive. It is argued that more extensive welfare regimes encourage women's participation, which, in turn, determines demand for labour (Tselios 2008, 405-406). One might argue that more extensive welfare states will be able to grant more equal access to financial opportunities, education being one of them. Due to this argument, distributional politics can be hypothesized to have a stronger, intervening or confounder effect on income inequality.

Demographic factors

Population growth is considered a relevant determinant regarding income inequality in a wide array of studies (Ram 1981, 257; Park 2010, 944; Ram 1984, 419; Singh 1984, 256; Ahluwalia 1976a, 314; Alderson 1995, 676; Tsai 1995, 470). According to interpretations of Kuznets curve, the rate of population growth "increases at first and later declines in the course of the demographic transition; and a negative effect on inequality of the spread of education, causing a long-term decline in inequality as education spreads" (Alderson 1995, 694). Also, in theories not building on Kuznets anticipation, a high rate of population growth is hypothesized as a dis-equalizer (Ram 1984, 419; Rodgers 1983, 444). This can be described by hypothesizing a poverty threshold where below the threshold it is economically preferable to have several children, and that over this threshold a smaller number of children would be preferred. If the number of people living under the threshold increases, this would increase the economy-wide fertility rate and possibly reduce economic growth (Round 2004, 291). Critics of Kuznets curve and human capital hypothesis on this point argue that the different premises and factors affecting population growth is essential, as the number of workers on the job market is predictive for the mechanisms of demand and supply which these theories build on. Higher rates of population growth might therefore be considered able to hamper the possible effect of education on income inequality.

Economic growth

Economic growth is another common variable that researchers use to predict income distribution within a country (Ahluwalia 1976a, 314; Alderson 1995, 676; Digdowiseiso 2009, 1; Park 2010, 943). As proposed by Kuznets, many researchers support the notion of growth first increasing income inequality rates, before declining it. In the wake of this Kuznets theory, different strands of development research has evolved on this subject; some trying to identify a mechanical relationship between the level of development and inequality, others to identify the causal factors influencing the evolution of growth and inequality independently (Squire 2003, 326; Park 2010, 943; De Gregorio and Lee 2002, 395). However, there is still uncertainty regarding the direction of causality of the inequality/growth relationship (Martino 2008, 375; Beck and Asli Demirgüç-Kunt 2007, 28; Digdowiseiso 2009, 1; Motonishi 2004, 12). Economic growth is presumed to be enhanced through human capital accumulation, and human capital accumulation is thought to enhance economic growth (Round 2004, 291; Squire 2003, 333). In other words, testing the two mechanisms and their impact on income distribution is a complicated matter.

Investigating the effect of economic growth on income distribution as a one way causal mechanism through existing theory provides conflicting predictions (Beck and Asli Demirgüç-Kunt 2007, 28). Some models imply that financial development enhances growth and reduces inequality (Digdowiseiso 2009, 1). Beck argues that growth will weaken credit constraints, which should benefit the poor because they might utilize more efficient ways to accumulate capital and could expect higher returns on their investments (Beck and Asli Demirgüç-Kunt 2007, 28). Beck and Levine argue that financial development disproportionately boosts incomes of the poorest quintile and reduces income inequality, because it is associated with an increase of people living in extreme poverty (Beck and Asli Demirgüç-Kunt 2007, 27). Others argue that that faster short-term growth is not systematically associated with higher inequality at a given level of development (Tsai 1995, 471). Contradicting models predict that financial development favours the rich, as improvements in the financial sector might not affect the economy of the poor, if this relies on informal, family connections for capital (Beck and Asli Demirgüç-Kunt 2007, 28). Or if growth is concentrated in a specific sector, lags in labour mobility could create factor market disequilibria, generating significant income differentials (Ahluwalia 1976b, 129). These theories have been tested by many researchers, also with divergent results (Sadoulet 2005, 267). This illustrates that economic growth might be a stronger, intervening or reinforcing effect on the relationship between education and income inequality. This is essential to keep in mind, as this analysis only investigates the reported findings on the one-way causal relationship between bivariate variables. In the later discussion of how to interpret or findings, these possible confounding or intervening variables are included.

2.5 Summary of theory chapter

As shown in this chapter, there are a multitude of theories regarding the relationship between education and income inequality, as well as a diversity of theories regarding the effect of other factors on income inequality. Development-theories, macro-economic, micro-economic and political theories have been presented. This chapter illustrate the complicated task of measuring and determining effects between large and multifaceted social phenomena. With detailed knowledge regarding the central characteristics of these studies, differences in measurement of education and inequality can be used to model heterogeneity between these findings in the later multiple MRA. A closer selection and description of chosen independent variables is given in section 4.2. These theories also serve as the backdrop for discussing the findings from the conducted MRA in chapter 6.

3 Method

This thesis is set out to answer three main research questions: (Q1): What is the mean observed effect of education on levels of income inequality in research from the past 50 years? (Q2): Is there publication bias in the literature regarding this relationship? (Q3): And are there systematic differences between the studies on this relationship, when controlling for publication bias? The chosen method to answer these questions multiple meta regression-analysis.

In this chapter I will answer three questions about multiple meta-regression analysis: what, why and how. I start by describing what multiple meta-regression analysis is, based on its place in the methodological landscape, general definitions, and its strengths and weaknesses as a method for analysis. Second, I explain why this method is chosen for the current analysis, based on its relevance for answering research questions and the specific advantages of adopting this method within the format of this thesis. Third, and lastly, I will present how I have conducted a meta-regression analysis, in six general steps.

3.1 What is a meta-regression analysis?

3.1.1 Meta-analysis – one way to conduct a systematic review

Within all fields of science, a major task is the development of theory. To do this, the researcher must have knowledge on available results of several previous studies on the subject of interest (Hunter 2015, 320). If research is to fulfill its purpose as a contribution to a certain field or topic, it is a general scientific standard that new research should be put in relation to existing knowledge (Grønmo 2004, 75; Card 2011, 3; Hunter 2015, 14). In order to determine what we already know about a certain phenomenon, scholars often conduct a some sort of *review* to summarize previous research, "bringing together and summery or synthesize previous published work" (Dacombe 2018, 149; Frey 2018). There are several ways to conduct such reviews; *traditional reviews, narrative reviews* and *systematic reviews* being the most common within social sciences. The terminology regarding various types of research synthesis in social sciences is still developing, and therefore also somewhat unambiguous (O'Connor and Sargeant 2015, 262). The description of *traditional reviews, narrative reviews* and *systematic reviews* presented here is therefore based only on a selection of sources.

<u>Traditional reviews, narrative reviews, and systematic reviews</u>

Collecting information on a topic without systematic guidelines for such collection is often referred to as a traditional review (Hattie and Hansford 1982, 71). This way of collecting data that is widely criticized as being unreliable, as many subjective choices made by the researcher during the research process are not made visible – increasing the chances of researchers giving a skewed picture of existing research on a specific topic, and eliminating the possibility for others to verify the reviewing process (Frey 2018, 71; Hattie and Hansford 1982). Narrative reviews contains a research question as a starting point for the literature search, a transparent research strategy – including clarity of inclusion and exclusion of literature - and conducting a process by which findings are analysed and pulled together (Frey 2018). The guidelines of narrative review are however very varied, and is by some researchers considered to be a "good", but not obligatory practice (O'Connor and Sargeant 2015, 262). Compared to traditional reviews, narrative reviews still aim to being systematic and transparent in the search for literature on a subject. Narrative reviews are however rarely used to answer a single research question, and the results of these reviews are often presented in the form of one or more "propositions", which may lead to new theories or research or summarize current practice, often presenting controversies and emerging issues that may not have presented themselves in individual works (Frey 2018). Although the narrative review follows a more systematic process than the traditional review, the systematic review lays an even greater emphasis on explicit, structured, and replicable processes, which are principally aimed at maximizing the criteria for reviews outlined earlier (Frey 2018). A systematic review differs from more conventional narrative reviews by its exhaustive searches, in an attempt to include all studies meeting explicitly stated criteria (Stanley and Doucouliagos 2012, 3). Systematic reviews can be used to answer more specific questions and relies solely on empirical studies to answer it (Frey 2018). In addition, conscious efforts are made to locate relevant studies, control for publication bias and other biases within existing research.

While narrative reviews are still the standard within most social sciences, these are criticized for increasing the risk of avoiding or dismissing studies or findings that do not fit into preconceived notions or theories (Stanley and Doucouliagos 2012, 2). If the aim of the researcher is to give a comprehensive illustration of existing research on a given topic, a systematic review is better suited for such investigations. By adapting a specified scientific framework for conducting review of existing research, one can provide a more reliable and valid process for "bringing together/summery/synthesize previous published work", which in turn will make it easier to put new research in relation to existing notions and theories within a certain scientific field. This demonstrates the benefits – and the importance – of systematic review.

Systematic reviews: literary review, research synthesis and meta-analysis

More specifically, systematic review is a methodology for "locating existing studies, selecting and evaluating contributions, analysing, and synthesizing data and report evidence in such a way that allows reasonably clear conclusions about what is known and what is not known" (Denyer 2009, 672). Using systematic review might also help researchers sharpen the focus and their methodological positions within their own work (Dacombe 2018, 154). Common types of systematic review are *literary review*, *research synthesis* and *meta-analysis*.

Literary review is often used to describe the beginning sequence within a study that aims to give the reader an introduction to central theories earlier findings from other researchers on a given subject. *Research synthesis,* is set out to collect empirical evidence to answer a research question and answer this question by summarizing already existing research on this question (Chalmers, Hedges, and Cooper 2002, 16-17). *Meta-analysis* involves a statistical and systematic review of relevant research on a specific topic, often with more rigid methods for testing for biases in the selected literature and resulting in a quantitative summary of the sampled observations (Stanley and Doucouliagos 2012, i; Chalmers, Hedges, and Cooper 2002).

Compared to meta-analysis, literary review and research synthesis rely more heavily on the researcher doing subjective selection of studies, leaving the samples in literature review and research synthesis more vulnerable to *selection bias*, where the researcher misses out on possibly important findings by doing a purely subjective evaluation of existing research, and *publication bias* in the literature, ignoring the fact that some research is more likely to be published than other research. This is partly why literature review and research synthesis are criticized for not meeting the standards of reproducibility, objectivity, and an unbiased case selection (Stanley 1989; Stanley 2001; Hunter & Schmidt 2004; Stanley 2005; Stanley & Doucouliagos 2012). Meta-analysis on the other hand, is set out to include an inclusion of existing research that is as comprehensive as possible. This type of analysis also includes the statistical methods to control for publication bias and contains stricter guidelines for transparent and verifiable case selection.

To sum up, within the toolbox of methods available for social scientists, systematic review is a process of summarizing previous research, suitable for researchers aiming to provide a reliable review of existing research on a given research topic. Meta-analysis is one way to conduct such a systematic review. This method provides the opportunity to systematically chose search for a more comprehensive case selection, followed by the possibility to control for publication bias in this sample.

3.1.2 Defining meta-analysis and multiple meta-regression analysis

Meta-analyses are known as "quantitative methods for combining information across different studies" (Tweedie 2001, 9717) or “statistical and systematic review of all relevant research” on a specific topic (Stanley and Doucouliagos 2012, i). It is an approach constructed to locate existing studies, selecting, and evaluating contributions, analysing, and synthesizing data and report evidence in such a way that allows reasonably clear conclusions about existing knowledge on a topic (Denyer and Tranfield 3009, 672).

Within this thesis, a meta-regression analysis is used. A meta-regression analysis can be considered a “regression analysis of regression analyses” (T.D Stanley and Jarell 1989, 299). Broadly speaking, meta-regression analysis (MRA) can be defined as a “statistical synthesis of data from separate but similar, i.e. comparable studies, leading to a quantitative summary of the pooled results” (Chalmers, Hedges, and Cooper 2002, 17). A common way to quantitatively summarize the results, is to identify an overall mean effect within the pooled observations. This is often done with weighing techniques, where the individual estimates are assigned different values, based on the quality of the study it stems from.

Because the research subject is existing literature, it is standard to use statistical techniques to check for publication bias in the sample. Publication bias refer to the issue that arises when researchers, reviewers and editors treat some results more favourably than other results (T. D. Stanley 2008, 279). This is often referred to as the “file drawer problem” and occurs when studies that contain certain findings, or only is in line with the accepted theories on the certain topic ends up being published. This might create a skewed picture of existing knowledge within a field, causing problems for researchers, policymakers or others who are trying to acquire knowledge on various subjects. Hence, testing for such bias is important within this analysis.

Like regular regression analysis, MRA can be used to study the correlation between two variables (X and Y). If the research question suggest that one needs to control for multiple other factors to be certain to investigate the actual relationship between X and Y, one can control for multiple Zs (other factors), this can be investigated through a multiple MRA (Card 2011, 156). The purpose of multivariate MRA is to use the techniques of meta-analysis to obtain average correlations from multiple studies, while controlling for possible sources of heterogeneity (Card 2011, 285). Using multiple MRA is both a standard, and recommended, approach to model systematic heterogeneity (Stanley and Doucouliagos 2012, 81). Part of this method involves identifying suspected sources of systematic differences between the findings and using these as independent variables to investigate if these variables are systematically correlating with differences in the pooled findings. More detailed description of these core elements of MRA is outlined in section 3.3. To answer the question of what a meta-regression analysis is; it can be defined as a systematic review where statistical methods are used to investigate the reported effect of something across several studies and controlling for publication bias and heterogeneity within these findings.

3.2 Why conduct a multiple meta regression-analysis?

When deciding on a research method, two aspects should be considered. Generally, one must consider the general strengths and weaknesses of the chosen method. More specifically, one needs to evaluate the advantages and disadvantages of using this method on chosen research question and consider how one might counteract the general weaknesses of the method within the specific project. Hence, I will answer this question by outlining general strengths and weaknesses for using MRA in social sciences, before explaining why meta regression-analysis is suitable to answer the given research questions and to account for possible weaknesses followed by this choice of method.

3.2.1 Meta regression analysis within social sciences – general strengths

There are several advantages to using meta-analysis. Six of them are highlighted here. First, when conducted properly, meta-regression analysis can serve as an objective and critical methodology to integrate conflicting research findings (Stanley and Doucouliagos 2012, 2). Conflicting research findings are usually considered as a driving force for creating new theories, but at the same time, conflicts between findings risk being portrayed as more consequential than they are. This is because research often receive more attention if it contains findings that clearly deviates from existing research (Stanley and Doucouliagos 2012, 2). This makes it harder to spot actual divergence between findings. Meta-analysis allows for integrating empirical findings across such studies and might be useful to illustrate the actual divergence in the findings in a systematic way, as well as revealing simpler patterns of relationships that underlie the research literature (Hunter 2015, 14). Second, meta-analysis provides the ability to summarize a larger area of empirical inquiry. Within the social sciences the production of quantitative research has increased sharply in recent years. By using MRA, one can review large amounts of theoretical work and summarize large amounts of findings in a systematic way. As the advancement of scientific knowledge is based on building on the foundation of prior research, the ability to summarize and describe existing knowledge is essential (Card 2011, 3). Third, in addition to combining and summarizing conflicting research, meta regression analysis allows a closer study of what might be the reason for systematic variation found in different research on a specific topic. As mentioned, a common finding among meta-analyses conducted within different social sciences is excessive heterogeneity (Stanley and Doucouliagos 2012, 80). Heterogeneity stems from the circumstance that an empirical estimation from a study is not only a reflection of what one want to measure, but dependent on numerous other factors, such as place, time, other variables included in the model, the measure of the dependent variable, chosen estimation technique, to mention some (Stanley and Doucouliagos 2012, 81). By using MRA, one can uncover the degree of conflict between findings, systematize these differences and study if and how the researchers' methodological choices might explain observed systematic differences. Fourth, MRA is as mentioned, suited to control for evident publication selection bias. Publication bias is a central topic within meta-analysis. If there is evidence indicating that one should expect publication bias in the chosen field of research, this means that the sample is not likely to include what is known about a subject, but only the knowledge that is more likely to get published on a subject. Fifth, meta-analysis allows for testing rival theories. As one can compare results from a large amount of research, it is possible to compare the effects documented within the different studies, and seek to explain these differences, which allows for a deeper discussion of the strength of rival theories. Sixth, general meta-analysis can be useful within the social sciences for studying phenomena that are subjects to policy. Theories within positivist social sciences are built on the observations of the real world and is often used as a foundation for evidence-based policy changes later. A meta-analysis offers a systematic and objective way to summarize and understand current research, which gives a foundation for policy discussions that is not built on subjective interpretation of existing research.

Based on the above, meta-analyses are important to political science because they contribute to reducing bias and perform a synthesis in a manner which is transparent enough to ensure valid replication (Dacombe 2018, 150). In political science, there is rarely consensus on theory, and important insights into the "validity of normative claims" can be given by systematizing empirical knowledge (Dacombe 2018, 153). Their primary value lie in the possibility for investigating reported effects on a phenomenon across large numbers of findings, in order to describe, investigate and model the observed results from an entire field of research (Dacombe 2018, 155). There are however legitimate objections to applying empirical standards developed in other sciences to political studies, and general concerns about the weaknesses of a meta-analysis approach, which will be outlined in the following.

3.2.2 Meta regression analysis within social sciences – general weaknesses

Yet, while they are common in medicine and psychology, MRA are still not common within social science (Cancela 2016, 265; Dacombe 2018, 148). Even though there are many areas where creating a synthesis of existing research is useful, there are legitimate objections to applying this type of research synthesis in political studies (Dacombe 2018, 156). Four central objections to using this approach in social sciences is outlined here.

One objection to using MRA in political science, is concerned with the ways in which this approach deal with qualitative research (Dacombe 2018, 152). More specifically how meta-analysists adopts the perspective that qualitative studies with smaller N are less reliable evidence, and how there is little focus on the intellectual basis of qualitative work that one might include (Wallace 2004, 465; Dacombe 2018, 152). This is partly due to the common practice of "weighting" observations from studies based on the sample size of the specific study. Further, this quantitative focus might increase the risk of research studies being abstracted from their social context, potentially resulting in findings that are far from reflecting the phenomena they were supposed to (Wallace 2004, 549-550; Hammersley 2001). There are several ways to contradict this weakness. Sandelowski and Barroso (2006) has developed an approach to synthesize qualitative research that can be used by social scientist conducting meta-analysis (Sandelowski 2007, 9-10; Ludvigsen 2015, 320). This framework does however not provide crystal clear prescriptions and are considered "a large undertaking for novice researchers and inexperienced students" (Sandelowski 2007, 9-10). Another option is to change the often rigid focus of meta-analysis on "what is confidently known about a subject" (Wallace 2004, 467), to what is "perceived as known" on this subject. This can be done by narrowing down a given research question and make transparent to the audience that one does not review all science on a given topic, but a systematic selection of such research, prone to favour quantitative research. This is the case of the meta-analysis conducted here, which only investigates the empirically observed effect of education on income inequality in a selected sample of political, economic, and sociological research over the past 50 years.

A second subject of criticism is that meta-analysis is "theoretically vacant", and not suitable for answering questions within the social sciences because of the theoretical complexity of these questions (Dacombe 2018, 153). Critics argue that meta-analysis are prioritizes statistical techniques over intellectual craft work as concepts and theories (Dacombe 2018, 153). One might argue however, that MRA does not seek to provide overreaching answers to complex issues. Instead, they aim to give unique insight in the theoretical differences between studies, by using their empirical differences as a starting point, and then investigating their theoretical foundations and research design possible explanations for their results. Within MRA, theory serves as the foundation for: creating research questions, constructing inclusion and exclusion criteria, and as possible explanations to the systematic differences we have detected between the different studies through statistical analysis. Several handbooks underline the great importance of theoretical insight to conduct a satisfying meta-analysis (Card 2011, 300; Stanley and Doucouliagos 2012, 29). The prevailing counterargument to this criticism is that meta-analyses are far from theoretically vacant, because that they demand high levels of theoretical insight on a certain field.

A third objection is related to researchers' choice of inclusion and exclusion of different studies, and their treatment of the studies included. Meta-analyses are supposed to be as comprehensive as possible when including data, and at the same time only include studies that can be statistically compared in a meaningful way. This requires clear-cut inclusion and exclusion criteria. When studies are included in the sample, the estimates are commonly given a presumed weight based on the quality of the study from which it is derived. For this reason, many criticize meta-analyses for constructing a "hierarchy of evidence" to identify the studies which are most reliable and therefore most suitable for inclusion. Both the creation of criteria for inclusion and weighting the selected estimates might lead to 'losing' important evidence through the exclusion or downgrading of methodologically imperfect studies (Dacombe 2018, 152) (Boaz et al., 2002; Victor, 2008). On the other hand, meta analysist must also take into account the criticism that highlights how including studies of low quality may taint the analysis (Stanley and Doucouliagos 2012, 17). These objections illustrate the diverse challenges of selection bias and different treatment of empirical observations, issues that are not unique to meta-analysis. Without a clear-cut answer to how one might avoid tainting the meta-analysis by including low-quality studies, this analysis adopts a transparent and justifiable exclusion and inclusion criteria, and a theoretically reasonable way to evaluate the quality of the studies included. Regarding the latter, Stanley & Doucouliagos propose "calculating precision as the inverse of the estimate's standard error, where more precise estimates are assigned a higher weight" to avoid judging a study's quality on beforehand, where more subjective perceptions can shape the inclusion and exclusion of studies (Stanley and Doucouliagos 2012, 34). Precision is used here as a measure of quality, leaving less chance of including research that might hamper the statistical analysis, as this choice relies on only objective statistical information (Stanley and Doucouliagos 2012, 34). At the same time, one must acknowledge that the choice of measure of study quality within this analysis favours large-N data and quantitative studies.

A final issue with multiple meta-regression analysis, is the problem of including studies with misspecification. Within econometrics, MRA was initially proposed to correct known misspecification biases (Stanley and Doucouliagos 2012, 3; T.D Stanley and Jarell 1989). Meta-regression analysis has the potential to correct the original econometric research for a variety of biases, including misspecification, omitted-variable bias and others (Stanley and Doucouliagos 2012, 21). For example, while all studies on a specific topic might be drawn from the World Bank Development Indicators, they may not include the same set of countries and time periods. Misspecification errors might occur through the researchers' idiosyncratic choices of exact model specifications and methods. MRA controls for the country composition of the samples, the time periods used and other potentially dependent dimensions of the research results (Stanley and Doucouliagos 2012, 36). There is, however, still a risk of misspecification bias, largely due to omitted-variable bias, which is when a statistical model leaves out one or more relevant variables for investigating a phenomenon. Because of both statistical and practical issues regarding data, researchers sometimes need to exclude important explanatory variables, or ignore these variables. This increases the chance of omitted variable bias, of mis-specified studies (Stanley and Doucouliagos 2012, 86). Both political, sociological, and economic researchers will have large problems with trying to include all relevant variables in their models. It has been confirmed by other studies that misspecification biases are frequently detected in empirical research, many even significant enough to have an effect on how we perceive a research subject (Stanley and Doucouliagos 2012, 3). There is no reliable way to know which model specification is correct. Within this meta-analysis, the presented theory is used to highlight possible variables that might affect the relationship between education and income inequality, that are usually not controlled for in all analyses. However, there will be no control of possible omitted variables, as this analysis is limited to investigate the systematic differences between common study characteristics and their results.

In sum, there are general weaknesses related to using multiple MRA within social science research. MRA is fairly under-exploited in the political science literature (Doucouliagos and Ulubaşoğlu 2008), but as illustrated here, the adoption of more systematic reviewing techniques can demonstrate its value. By highlighting one's methodological choices, clarifying the scope of one's research, being clear about the theoretical foundation of the project, and how it might need to be supplied by research with different ontological and methodological perspectives, the researcher using meta-analysis is better equipped to contribute to the "tower of knowledge" within social sciences in a meaningful way. It is also essential to highlight that many of these criticisms can be directed towards almost any quantitative research that seeks to collect and compare knowledge. There are legitimate objections to applying empirical standards developed in the applied sciences to political studies, but it is clear that there are many advantages of adopting this approach, as has been demonstrated in other social science disciplines (Dacombe 2018, 156).

3.2.3 Using meta regression analysis to answer chosen research question

As argued in this section, meta regression analysis can be a useful asset to many political researchers. Meta-analysis is in general considered a fruitful approach to answer research questions about *what we perceive to know about a certain topic*. Considering the presented advantages of meta-analysis – the possibility of comparing a large amount of knowledge in a transparent and systematic way with the ability to control for several biases –this approach can be considered an appropriate method to be able to fulfil the aim of this thesis. This is to answer: (Q1): What is the mean observed effect of education on levels of income inequality in research from the past 50 years? (Q2): Is there publication bias in the literature regarding this relationship? (Q3): And are there systematic differences between the studies on this relationship, when controlling for publication bias? One of the general measures taken to counteract the weaknesses of meta-analysis listed in this past section is providing a transparent and replicable model of the specific research process, presented in the following.

3.3 How to conduct multiple meta regression analysis

A classical meta-analysis can be used to investigate whether there are observed levels of income inequality correlating with increased levels of education. As described earlier, a classic meta-analysis consists of calculation the average effect of education on economic inequality across studies, using different weighting techniques – followed by controlling the results for publication bias and checking for systematic differences between these findings (heterogeneity).

The process of conducting a meta-analysis is here divided into six stages, inspired by existing meta-analyses and relevant literature (Stanley and Doucouliagos 2012, 148-150; Card 2011, 9; Broderstad 2015-06-02, 35). The stages are (**A) Search relevant literature and make decisions on study inclusion; (B) obtain the effect size; (C) use descriptive statistics to investigate trends (D) accommodate publication bias; (E) use meta-regression analysis to model heterogeneity; (F) guiding further research**. Within step (A) –(C) considers collecting, summarizing and interpreting existing research evidence. Step (D)–(F) includes modelling of the data, which is essential to check for biases that might give a skewed picture of the combined findings, and to explore the variation across studies and evaluate possible patterns in their reported results. These stages of conducting a meta-analysis constitutes a systematic and objective procedure, but within these stages the researcher must make several subjective methodological choices. Subjective choices that have been made through these stages of the research process will be highlighted and justified in the following – with the aim of ensuring a transparent and replicable study.[3]

(A) Search the relevant literature and selecting and coding of estimates

Meta-analysis starts with collecting all existing evidence on a prespecified research topic (Tong 2019, 5). As described earlier, researchers are faced with several subjective considerations on

---

[3] This section has sought to satisfy the Guidelines for Meta-Analysis of Economics Research Reporting (T.D. Stanley et al. 2013, 390; Havranek 2019).

what studies to include and exclude in the process of collecting estimates that are comparable both within and between studies (Stanley and Doucouliagos 2012, 13). Limiting this possibility of selection bias is essential within MRA, as the very purpose of this method is to discover systematic biases within the general literature on a subject (Stanley and Doucouliagos 2012, 13). At the same time, it would undermine the function of meta-analysis if the researcher cannot exclude studies that are not comparable or relevant for the certain topic. To account for these considerations, a) the search for literature must be as comprehensive as possible b) the guidelines for the systematically selection of research must be clear-cut, transparent and replicable (Stanley and Doucouliagos 2012; Tong 2019). In practice, the researcher must decide where and how to search for existing studies and construct inclusion criteria regarding what outcomes, explanatory variables, and populations are under study to create a pool of comparable studies (Tong 2019, 5). This way studies are included first on their basis of relevance for the research question, and secondly evaluated – and some excluded – dependent on whether they contain sufficient information to be used in a meta-analysis.

<u>Selection on beforehand - Searching for studies</u>

I set out to investigate existing empirical findings on how education might affect economic inequality on a country basis, within political, sociological, and economic literature. Intuitively, the scope of this research question requires searching for data in databases containing literature from these fields, for empirical studies with education as independent variable, and economic inequality as a dependent variable. As this thesis is built on already collected data from the meta-analysis conducted by Abdullah. et. al in 2010, I only searched for research published after this meta-analysis was conducted. Further, adopting a research strategy close to the one used by Abdullah et. al. was beneficial to the analysis. This raises several challenges, as the search process in meta-analysis is often only described in general terms and as the earlier meta-analysis was conducted over more time and with more accessible resources. Both efficiency and accuracy are considerations that have been considered.

Abdullah et. al. describe their search process as a use of 12 different keywords in various combinations within four databases, leaving them with over 200 articles. Using different combinations of the same keywords leaves me with over one million articles. This might be because I used more combinations of the search phrases, or because there has been published a large amount of research containing the different search words over the past 10 years.

Nevertheless, I am forced to limit the search for studies. On a general basis, limiting the search in different ways might be reasonable when a literature is known to be enormous (Stanley and Doucouliagos 2012, 13). More specifically, the search must be modelled to ensure inclusion of theoretically fitting articles. I therefore make three choices: reducing the number of databases, reducing the number of key words, and adapt advanced searching techniques within two of these databases.

Three databases are used in the search: Google Scholar, JSTOR and ProQuest. All three are considered more general databases and are commonly used by meta-analyst within the social sciences (Abdullah, Doucouliagos, and Manning 2013, 303; Ahmadov 2013, 1243; Nikos Benos 2014, 672). Studies that are published within other databases, or studies that have not been published on such online databases are excluded. I choose to exclude Econlit and RePec due to limited access to these databases and include ProQuest instead. This could potentially reduce the number of selected articles with an economic focus, but this is not considered to be the case, as 70% of the selected articles is published in economic journals.

The chosen search string was: “education” OR “schooling” OR “literacy” AND “income distribution” OR “income inequality” AND “regression” AND “cross-country analysis”. Five of these search phrases are the same as Abdullah et. al. To avoid the large number of articles that was not empirical analysis of the chosen relationship across countries, I chose to include the words “regression” and “cross country analysis”. The settings within the databases are set only to search for articles published between 2010 and 2022. Due to the vast amount of information within two of these databases, advanced search was conducted within ProQuest and Google Scholar. The narrower search phrase “education” AND “income distribution” AND “regression” AND “cross-country analysis” was adopted, and within ProQuest the search was conducted only within the Social Science Database, the Sociological Database, and the Political Science Database. This resulted in 31 articles from JSTOR, 108 articles from ProQuest and 59 articles from Google Scholar. This left a pool of 198 relevant articles.

<u>Selection after - inclusion criteria</u>

To ensure that the chosen articles are comparable, one must include a specific and explicit set of selection criteria, hat define the population of studies that are collected (Stanley and Doucouliagos 2012, 14). These criteria should be constructed based on the defined scope of one’s research question (Tong 2019, 5). The aim of this analysis is to investigate the empirical results of education on income inequality in cross-country studies. Based on this research aim, only studies that met the following criteria were included:

(a) *Income inequality must be the dependent variable and education the independent variable*: If the impact of education on income inequality is reported, this criterion is satisfied.

(b) *Reported econometric estimates:* Only empirical studies that provide regression results are included in the meta dataset (Stanley and Jarrell 1989; cited in Abdullah et. Al. 2015, 303). Qualitative studies and studies that do not use regression analysis for their empirical research therefore excluded.

(c) *Aggregate x and y the same way:* To make sure that the different measures are similar enough to be considered measuring the same phenomena. Within this analysis it is essential that both variables are coded as graded phenomena.

(d) *Contains necessary estimates:* Such as standard errors, sample size, and coefficients.

(e) Only cross-country studies: Studies conducted on different levels than the country level, and studies with only one N, are excluded.

(f) *English language:* To avoid the risk of misinterpretation by translation, studies from other languages are excluded. Within meta-analytical research, including only studies in English has not been proven a clear disadvantage for the over-all quality of the analysis (Stanley and Doucouliagos 2012, 15).

(g) *Possesses necessary quality:* The impact of study quality on findings is treated as a subsequent empirical evaluation, as discussed above. Hence the inclusions of studies are not based on a face-value quality judgement. However, to ensure quality and to answer the research question of why scientists within the scientific literature have not reached consensus on this topic, it is convenient only to include articles that are published within journals which are rated in the Web of Science. This excludes results from theses, books, research reports etc. This choice is justified in terms of efficiency and ensuring the inclusion of high-quality academic work. This limitation is not considered to harm the aim of the analysis, as the focus of the thesis is to investigate why *central research* on this subject show provide divergent results.

Based on these criteria, the articles then went through an initial screening, based on their title and abstract. This resulted in 73 articles (32 from Google Scholar, 31 from ProQuest and 10 from JSTOR). Then each article was controlled against the full list of inclusion criteria. The majority of the excluded studies did not include a cross-country analysis or did not have income inequality as the dependent variable and education as the explanatory variable. This left me with a pool of 10 articles.

Backwards searching

Despite being a systematic selection process, both the search choices and the stages of screening does include subjective choices, increasing the chances of missing out on possibly relevant studies for this meta-analysis. A way to counteract this problem, is by using extra strategies for searching for such studies. One of them is backward searching, sometimes referred to as "footnote chasing" – searching for relevant studies cited in the selected articles – to identify relevant studies omitted in the literature search (Card 2011, 50). Despite the potential biases of backward searches, many believe that they represent a valuable method of searching (Card 2011, 50). This approach is therefore adopted here, and all footnotes in the selected 10 articles was evaluated.[4]

In addition to the classical backward searching, I also made use of the "related works" function on Google Scholar. This function allowed for me to make use the selected works and find related work that is not necessarily cited within the actual articles. This type of search was also conducted for all the articles in the pool. After the classical footnote search, and the search for related works, I was left with a selection of 20 articles.

Search results and coding

After identifying a population of 20 studies, with 370 corresponding estimates, that could be meaningfully examined using the tools of meta-analysis, these studies were coded and included in the existing meta-data set. Of interest is the empirical evidence that these studies contain, such as "reported regression coefficients, sample size, standard errors and/or t-statistics" (Stanley and Doucouliagos 2012, 14). After the collection of research between 2010 and 2022, the dataset finally included 61 studies and 917 comparable estimates.

[4] To ensure a transparent case selection, a list of the studies included in the analysis can be found in appendix I.

(B) Generate standardized effect size and calculate average effect

Because the different articles in the sample contain different measures of the dependent and independent variable, it is essential to convert their findings into a common and comparable measure. Such a comparable measure is often referred to as the effect size. An effect size should allow us to measure the effect of a particular variable thought to be conditionally invariant (Stanley and Doucouliagos 2012, 22). In this thesis the effect economic effect of education on income inequality across studies is what one wants to measure.

A common measure for this metric is *partial correlation* (Stanley and Doucouliagos 2012, 29; Ahmadov 2013, 1245; Yesilyurt 2019, 356; Doucouliagos and Ulubaşoğlu 2008, 65). This is a unitless measure showing us the "strength of the association between the two chosen variables, holding all other factors constant"(Abdullah, Doucouliagos, and Manning 2013, 304; Broderstad 2015-06-02, 39). In contrast to regression coefficients, the partial correlation can be used to investigate heterogeneity produced by using scales of different variables (Abdullah, Doucouliagos, and Manning 2013, 304). Additionally, they can be used for a larger set of estimates and studies than many other effect size measures, and they also have the advantage that correlations are something most researchers are familiar with interpreting (Stanley and Doucouliagos 2012, 25). Partial correlation are computed as follows (Stanley and Doucouliagos 2012, 24-25):

$$r = \frac{t}{\sqrt{t^2 + df}}$$

Where the *t* is the given t-statistic of the exact regression coefficient and *df* refers to degrees of freedom of this t-statistic. By converting the different statistics from each article using the formula for partial correlations, one can meaningfully compare their found effect, in an efficient and straightforward manner. Partial correlation is therefore the chosen effect size measure in analysis. Within this thesis, the partial correlation results will be a generalized measure of the *reported impact of education on income inequality.*

A problem regarding the use of partial correlations is that its distribution "is not normal when its value is close to -1 or +1" (Stanley and Doucouliagos 2012, 25). This is however rarely a problem within economic or political research, because no or few partial correlations within are close to these limits (Stanley and Doucouliagos 2012, 25). Within the reported estimates, 35 out of the total 916 reported estimates are above 0,9 or below -0,9. As one can control for possible outliers here, this is not seen as an obstacle when partial correlations are used for obtaining effect sizes across studies.

To interpret partial correlation values, many meta-analyst adopt Cohens standard interpretation for partial correlations. According to this framework, ±0.10 is considered a small effect, ±0.30 as a medium effect and ±0.50 as a large effect (Ahmadov 2013, 1245; Cohen 1992, 98). Doucouliagos has presented alternative guidelines to interpreting the partial correlations, based on investigating the distribution of 22,000 partial correlations within the field of economics and compared them to Cohens guidelines for interpretations (2011, 10). He decreases the standards for interpreting partial correlations, proposing that a partial correlation that is less than ±0.07 should be considered small, ± 0,17 considered medium and greater than ±0.33 as a large coefficient (Doucouliagos 2011, 10; Abdul Abdullah 2015, 12). Both guidelines will be used to discuss the results.

Standard error of partial correlation

The partial effect includes information about the model from which the researcher has derived their results, and the effect that they find between the selected variables. As mentioned, one can also account for the "quality" of the individually reported estimate. To do this, we need to calculate the standard error of the partial correlation. This value is not the standard error of the specific estimate that is collected from the studies, but the standard error of the calculated partial correlation. The standard error of the partial correlation is calculated by:

$$\varepsilon = \frac{\sqrt{(1 - r^2)}}{df}$$

Where *r* is the standard error and *df* refers to degrees of freedom within the model in which the estimate is created (Ahmadov 2013, 1245).

Partial correlations and the "classical meta-analysis"

Aiming to understand more about the reported effect of education on income inequality across studies, detecting the mean partial correlation across studies is useful. Calculating a mean of all collected estimates would however not account for the fact that many estimates from a less reliable study could have a strong effect on the overall mean reported effect. As mentioned, several theoretical works on meta-analysis therefore suggests to weight the effect of the partial correlations based on their power (Doucouliagos and Ulubaşoğlu 2008, 65; Card 2011, 176). A regular estimation of such quality/power is sample size, as they provide more information for determining the true effect between different variables. Estimations from these studies can therefore reasonably weighted "heavier" than estimates from smaller samples (Doucouliagos and Ulubaşoğlu 2008; Card 2011, 176). Because this thesis investigated both cross-country and time-series data, the number of observations that each estimate is derived from is considered relevant to use as weights for determining quality of the studies. To calculate mean partial correlation across the studies, weighted for sample size, this formula from Ahmadov (2014, 1245) can be used:

$$\underset{\varepsilon}{-} = \frac{\Sigma[N_{ij} r_{ij}]}{\Sigma N_{ij}}$$

Where $\underset{\varepsilon}{-}$ is the mean observed effect of education on income inequality, ε is the standardized effect of from the ith estimate of the jth study and N is the number of observations, here deployed as weights. The results from such average partial correlations in meta-analysis are often considered the "most accurate" effect between the studies phenomena (Doucouliagos and Ulubaşoğlu 2008, 65; Ahmadov 2013, 1245). The weighted partial correlation can be interpreted with Cohen's (1992, 98) or Doucouliagos (2011, 10) guidelines for assessing effect size coefficients. To control for the average effect size across all sampled studies, weights are commonly adopted and used in different models. (Ahmadov 2013, 1245). In other models' different weights are used.

Within this thesis, four models are created and used with four different compositions of the sample. The different samples are the all-set; the average set; the all-set with studies that uses OLS FE; and the all-set with only studies that controls for endogeneity in another way than by using OLS. The average mean is derived from an unweighted model; a model where partial correlations are weighted by sample size; a weighted mean in a fixed effects model and a weighted mean in a random effects model. The results from these models are used to answer (Q1): what is the mean reported effect of education on national levels of income inequality in research today?

Because heterogeneity is considered a large issue in meta-analysis within social sciences, a Cochran's Q-test is used here to test explicitly for heterogeneity within the constructed models, to find the average partial correlation across studies. It is conducted by running a simple MRA with t-values on precision with no intercept, the sum of squared errors is the calculated Q-test, and distributed as s chi-squared with L-1 degrees of freedom (Stanley and Doucouliagos 2012, 49). It is worth mentioning that the Cochran Q-test is known to have little reliability, making the finding of no heterogeneity "only a reflection of the limitation of the test" according to Stanley and Doucouliagos (2012, 49). They suggest that one should assume heterogeneity in all cases of econometric research, and it is therefore relevant to investigate this further in the MRA. This test is still adopted within this meta-analysis, but heterogeneity is mainly investigated through the MRA. These results from this step in the meta-analysis is presented in the Results chapter, section 5.3. This includes mean partial correlations across different models using weights for study quality, and identification of possible heterogeneity in the sample.

(C) Use descriptive statistics to investigate trends

As in other empirical research, descriptive statistics should be included in meta-analysis (Stanley and Doucouliagos 2012, 49). In this analysis the descriptive statistics included is a time-series graph displaying the trends in *average reported effects of education on income inequality pr. year*, reported by partial correlations (see figure 5.1). A fitted line is used to describe the general trend of the findings from the included 50 years. The partial correlations her are weighted the same way as the mean partial correlations in the past section, by sample size. This is a common way to describe the trend of published adjusted partial correlations over time (Mandon and Cazals 2019, 281). The results from this section are in section 5.1.

(D) Accommodate publication bias

Publication bias; funnel plots and the PET-FAT tests

As educations effect on economic inequality is a subject of debate, researchers and editors might be equally interested in publishing studies both with and without significant findings, reducing publication bias. However, as made clear by earlier, publication bias occurs frequently making testing for this bias a necessity.

A common way to test for publication bias, is by using funnel plots (Stanley and Doucouliagos 2012, 53). A funnel plot is a graphical representation of the size of trials plotted against the effect size they report (W. Lee and Hotopf 2012, 140). The funnel plot illustrates the partial correlations of a given phenomenon and these estimates' precisions, i.e. the inverse of these estimates' standard errors. Interpreting the graphical funnel plot, one should check if the estimates are "randomly and systematically distributed around the true population parameter" (T. D. Stanley 2008, 60). In absence of bias, one should observe a symmetrical inverted funnel (hence the name funnel plot), where estimates from smaller studies scatter more widely at the bottom of the graph (Sterne and Egger 2001, 1046). If the pattern reveals that most estimates locate on one side of the positive or negative line, this will indicate that the spread is not random and might be caused by publication bias. However, because interpreting graphic plots is inevitably subjective (Terrin, Schmid, and Lau 2005, 894) statistical tests are used to account for the existence of publication bias in the created funnel plot.

According to central literature about meta-analysis, one of the best ways to test for empirical effect and control for publication bias is to use funnel asymmetry testing, FAT-PET MRA (Stanley and Doucouliagos 2012, 78; Alinaghi and Reed 2018, 285). The FAT-PET test contains two steps, funnel-asymmetry-test (FAT-test) investigating the symmetry of the created funnel, followed by the precision-effect-test (PET) to see whether the estimates indicate that the effect of interest is statistically different from zero (Stanley and Doucouliagos 2012, 62; Alinaghi and Reed 2018, 285). In the FAT-PET-MRA test, the standard error serves as a sort of "proxy" for their "likeliness to be published", meaning that one expects studies with smaller standard errors (and higher p-values) to be more likely to be published. Hence, a sample of publications with both low and high standard errors are less likely to be inflicted by publication bias (Rosenberger 2006, 7). The funnel-asymmetry-test and precision-effect-test meta-regression analysis can be expressed by using a weighted least squares-model (WLS), to accommodate heteroscedasticity, with the precision square $(1/SE)^2$ as weights:

$$r_i = \beta_1 + \beta_0 \left(\frac{1}{SE_i}\right) + v_i$$

where r_i is the reported estimates t-value, and *SE* its standard errors (Stanley and Doucouliagos 2012, 78). WLS is the best model for these tests, as they can contain substantial heteroskedasticity because standard error "is an estimate of the standard error of the elasticity measure that varies from observation to observation" (Rosenberger 2006, 7). Note that the weights are different than the one that has been used for calculating mean partial correlations and the descriptive statistics. For interpreting the results of the FAT test, Doucouliagos & Stanley (2013, 320-21) propose these guidelines:

1. If FAT is statistically insignificant or if $\beta_1 < 1$, then selectivity is "little to modest"

2. If FAT is statistically significant and if $1 \leq \beta_{!} \leq 2$ then selectivity is "substantial"
3. If FAT is statistically significant and if $\beta_{!} > 2$ then there is "severe" selectivity

The results from the precision-effect-test can be interpreted in accordance with Cohens or Doucouliagos guidelines for interpretation (Cohen 1992, 98; Doucouliagos 2011, 10).

If the results from these tests indicate the existence of publication bias in the sample, it is common to adopt a precision-effect estimate with standard error (PEESE) (Stanley and Doucouliagos 2012, 78). PEESE should however only be used if publication bias is found when modelling FAT-PET, or if the PET-term shows a genuine empirical effect. The PEESE test is considered to provide a better estimate of the "empirical effect corrected for publication bias" (Stanley and Doucouliagos 2012, 78). The PEESE-model takes the following form:

$$r = \beta_1 SE + \beta_0 \left(\frac{1}{SE_i}\right) + v_i$$

where the standard errors from the partial correlations are included.

These tests are common within meta-analysis, but not without flaws. Doucouliagos and Stanley (2012) point out that FAT tests are known for having low explanatory power. PET tests do not have the same issue as FAT tests but are more disposed to type I errors, where one detects an effect that is not necessarily there. This occurs if there is much unexplained heterogeneity in the meta-regression model (Stanley and Doucouliagos 2012, 64). This means that if one finds strong evidence (p<0,001) one should not rely on these tests exclusively. One alternative is to supply these tests with a multivariate MRA, as one can check for publication bias by controlling for the correlation between the partial correlation and standard error within these models as well.

Another central acknowledgement regarding testing for publication bias, is that no such test can allow for us to conclude with total confidence that there is no publication bias present in the literature. This is because statistical tests cannot control for unobservable phenomenon, such as publication bias. The statistical tests only allow for identifying variation amongst the reported results, reducing suspicion that the research field is dominated by systematically skewed findings. The results found at this stage is used to answer (Q2): *is there publication bias in the literature regarding educations effect on income inequality?*

(E) Use meta-regression analysis to model heterogeneity

Meta-analysts in economics and the social sciences routinely observe excess heterogeneity (T. D. Stanley and Doucouliagos 2017, 22). The issue of heterogeneity is as mentioned the influence of many factors on the reported estimates. These other factors are partly dependent on how the primary research is conducted (Kallager 2014, 36). Stanley and Doucouliagos stresses that all areas of research contain excess heterogeneity; that is, "greater variation than what would be expected by measured sampling error, SE_i, alone" (2012, 82). As highlighted in the beginning of this chapter, one of the main advantages with MRA is that is enables us to both control for publication bias, and to address and explain systematic variation in existing findings, hence, controlling for heterogeneity.

To conduct a multiple meta-regression analysis that will allow us to model heterogeneity, we rely on isolating the different study characteristics that we believe could influence the results (Stanley and Doucouliagos 2012, 81; Alinaghi and Reed 2018, 285). These study characteristics are variables related to the subjective choices made by the researcher designing the project. These variables differ from *essential variables*, which are the estimated effect size of each estimate, standard error, and measure of precision. To separate these two types of variables, variables that are potential sources of heterogeneity are coded as binary variables, and essential variables as continuous variables. Types of binary variables that are included in this analysis to control for heterogeneity are variables on the different study's estimation differences, measurement differences and data characteristics.

Further, heterogeneity can be modelled using a MRA with different statistical methods. It is expected find that WLS is the preferred model for the meta-analysis. The debate regarding strengths and weaknesses of using weighted least squares (WLS), fixed effects models (FE), or random effects models (RE) in MRA is presented in short here (Doucouliagos and Ulubaşoğlu 2008, 65-66).

WLS models are applied to summarize estimated reported regression coefficients and to explain heterogeneity in reported estimates, as it automatically allows for both heteroscedasticity and excess between-study heterogeneity (T. D. Stanley and Doucouliagos 2017, 20; Stanley and Doucouliagos 2012). Using a FE model, one assumes that the results differ only due to sampling error and systematic differences due to the research process. Using a RE model one assume that results differ both as a cause of sampling error and because of random differences between the studies (Doucouliagos and Ulubaşoğlu 2008, 66). Deploying fixed effects models in meta-analysis is not without flaws, as the assumption of this model is rarely met in analysis using meta-data, because it assumes that all estimates are “drawn from the same population with the same mean” (Broderstad 2015-06-02, 48). Using RE models within the MRA allows for between study variation, but it is also assumed that the effects are independent of all the explanatory variables. This premise is also rarely the case in meta-analysis. Accounting for this critique, neither the premises of fixed effect models nor random effects models are likely to fit the meta-regression perfectly.

Stanley and Doucouliagos test different models, focusing on their different ways of handling biases. They advised other social scientists anticipating routine and indirect heterogeneity to use WLS-MRA as the conventional approach for meta-regression of observational research. This is partly because WLS tends to have smaller bias than random effects models when there is publication bias. Because we cannot be sure about the absence of publication bias in our findings, it is preferable to use a model that is better for modelling data with such bias present. They state that FE models are better when there is no excess between-study heterogeneity whereas RE models should be applied when there is excessive between-study heterogeneity (T. D. Stanley and Doucouliagos 2017, 20).

The discussion of finding the best model for describing the data should be supplied with statistical tests of the presence of effects on study level. This can be done by running a Breusch-Pagan Lagrange Multiplier (LM) test to check for significant study level effects (Stanley and Doucouliagos 2012, 103). If the result of this test shows study level effects, a generic Hausman specification test should be applied, to test for endogenous predictor variables. This test estimate whether the observed effects are likely to exist due to random or fixed effects, answering whether one should prefer using a random or fixed effects model in the analysis. In this thesis, all three of these statistical models are conducted, making it possible to investigate and discuss the robustness of the findings across the different regression strategies. To control for within-study heterogeneity, a model with estimation clustered by studies is also constructed.

The post-estimation tests are used to discuss which model is most appropriate to use for interpreting the findings of the multiple MRA. In short, the path to choosing how to investigate heterogeneity can be described as: if there are multiple estimates per study, proceed to conduct a Breucsch-Pagan LM test to check for study level effects. If these are detected, run a cluster robust model with fixed effects and random effects, and use the Hausman test to identify the correct specification. Then, report all MRA model specifications and focus on the dimensions that have consistent findings through alternative model specifications (Stanley and Doucouliagos 2012, 105). As described, one expected find that WLS is the preferred model for the meta-analysis. The findings from these models are therefore considered more important to analyse. Stanley and Doucouliagos recommend to still including all models, even if FE and RE models are not considered suitable to evaluate the actual effect across studies. This is mostly to check for publication bias across models.

The WLS model that is adopted for the multivariate meta-analysis is given by:

$$r_i = \beta_1 + \sum \delta_j + K_{ji} + \frac{\beta_0}{SE_i} + \sum \frac{\beta_k Z_{ki}}{SE_i} + u_i$$

where r_i is the reported *t*-value of the *i*th reported effect, and SE_i is the standard error of this effect. The Z-variables are dummy variables for the study characteristics and K variables are related publication bias, reported through partial correlation and the standard error of the partial correlations. Following Ahmadov (2013) WLS models are weighted with precision (1/SE) and the FE and RE with precision squared as weights $(1/SE)^2$ (Ahmadov 2013, 1257).

To ensure studying variables that are not just hypothesized as relevant for the investigation on heterogeneity, but also shows actual relevance in relation to the findings of the meta-analysis, one can apply a general-to-specific modelling approach. This approach is recommended by Stanley and Doucouliagos, where one starts by including all potential contributing variables, then exclude the variables with the least effect one at the time, until only the statistically significant remain (Demena and Afesorgbor 2020, 5; Sebri and Dachraoui 2021, 6). General-to-Specific modelling (G-to-S) serves as a practical approach that allows for us to “minimize the potential of identifying spurious research dimensions through data mining” (Stanley and Doucouliagos 2012, 104). Although this approach involves leaving out information and variables, the limited degrees of freedom and the high multicollinearity forces us to simplify the MRA model. As Stanley and Doucalouliagos puts it, the G-to-S approach is the "least objectionable way to do so" (Doucouliagos, Stanley, and Viscusi 2014, 73). The results from this MRA are used to answer (Q3): *Are there systematic differences between the studies on educations effects on income inequality?*

(F) Guiding further research

Based on the results from the tests for publication bias, and models of heterogeneity between the sampled research, one ought to initiate a discussion on the state of the research. The findings should be evaluated and discussed based on the presented theoretical framework. After providing a thorough discussion of the findings from the analysis, relevant directions for further research on this topic are proposed.

3.4 Summary of method chapter

Despite favouring quantitative research and risking biases occurring at the different stages of meta-analysis, multiple meta-regression analysis is useful to answer questions about what we perceive to know about a subject within a field, and to discuss the possible systematic differences between the results on this subject. The steps of conducting a meta-analysis have been presented, requiring a systematic procedure for collecting studies to a sample, calculating a generalized measure of effect that can be used to compare the studies, use statistical methods to search for publication bias and at last, and to use the multiple MRA to investigate the possible heterogeneity within the findings. Before presenting the results of these plots and models, information about the sample of data is presented. The following chapter contains a presentation of the procedures of handling articles with missing data, an overview of the partial correlations belonging to each of the studies, and the independent variables that has been chosen to control for heterogeneity in the MRA.

4 Data

4.1 Dataset

The dataset consists of 61 studies, coded in 80 different variables, providing 917 individual estimates. This sample includes studies from over 50 years of research that empirically investigate the relationship between education and income inequality, from 1971 to 2020. It is built from a dataset published by Hristos Doucouliagos, which he, Abdul Abdullah and Elisabeth Manning used to produce the meta-analysis "Does education reduce income inequality? A meta regression analysis" published in *Journal of Economic Surveys* in 2013.

That dataset contained 66 studies and 82 variables. Among the 82 original variables, 30 variables were removed and replaced with 23 other variables that were more relevant to the aim and theoretical framework of this analysis. Further, 20 of the studies and their corresponding estimates were removed, because they did not fit my specification of only investigating studies with more than one N and on a country level. I proceeded to collect 20 more studies to include in the dataset. These studies were collected through the selection process described in the Methods chapter, part 3.3 (A).

Studies that lacked essential data such as standard error, t-statistics or degrees of freedom could still be included, and calculations of these estimates was conducted manually.[5] The following formulas were used to calculate t-statistic and standard error:

$$t = \frac{\beta}{SE} \quad \text{and} \quad SE = \frac{\beta}{t}$$

Degrees of freedom was calculated by:

$$v_i = n - 1$$

On average, each study contained 15 estimates. Four studies contained only 1 estimate, four reported over 40 estimates, and one reported more than 80 estimates. The partial correlation had -0,979 as minimum value and 0,981 as maximum value. The overall unweighted partial correlation was -0,043, which means that on average there was a reported negative effect of education on income inequality. A negative effect on inequality means that education is observed to reduce inequality. According to both Cohen's standardized interpretation of partial

[5] In one study (Andres and Ramlogan-Dobson, 2011) t-statistics and standard error was derived from p-values, using the statistical software R. The value was derived using the inverse cumulative density function of the distribution, given a certain random variable *x* and degrees of freedom *df.* The specific coding was *qt(1 – (p-value), samplesize).*

correlation and Doucouliagos' guidelines, this is considered less than a small negative association, as it is below 0,10 (1992, 89; 2011, 10). Publication bias and heterogeneity has however not been controlled for in this result, so this overall unweighted average partial correlation should therefore not be used to draw conclusions. The partial correlation of each reported estimate from the chosen studies serves as the as the dependent variable in the following MRA. In Table 1 the variability of the partial correlations in the dataset is described in detail.

Table 4.1 Reported effect size, sorted by studies included in meta-data

Authors	*No. of estimates*	*Min*	*Max*	*Mean*	*Std.dev.*
Ahluwalia (1976a)	49	-0,49	0,655	0,141	0,256
Ahluwalia (1976b)	8	-0,32	0,38	0,117	0,28
Barro (2000)	20	-0,28	0,29	-0,03	0,192
Beck et. al. (2007)	11	-0,13	0,19	0,059	0,091
Bourguignon and Morrison (1990)	16	-0,65	0,62	0,117	0,467
Breen and Penalosa (2005)	15	-0,32	0,26	-0,027	0,211
Gyimah-Brempong (2002)	4	-0,52	-0,19	-0,304	0,158
Chambers (2010)	9	-0,23	0,28	-0,0018	0,205
Chiswick (1971)	3	-0,83	0,65	0,10	0,813
Edwards (1997)	1	-0,29	-0,29	-0,29	0
Gregorio and Lee (2002)	7	-0,265	0,215	-0,030	0,172
Gupta and Singh (1984)	2	-0,25	-0,30	-0,28	0,030
Gupta, Davoodi and Terme (2002)	15	-0,255	0,33	0,081	0,152
Janvary and Sadoulet (2000)	1	-0,06	-0,06	-0,06	0
Jha (1996)	81	-0,49	0,53	-0,012	0,267
Keller (2010)	40	-0,59	0,39	0,016	0,285
Koechlin and Leon (2007)	10	-0,45	-0,10	-0,25	0,136
Lundberg and Squire (2003)	7	-0,26	-0,20	-0,24	0,248
Nielsen and Alderson (1995)	47	-0,30	-0,14	-0,228	0,039
Odedokun and Round (2005)	4	-0,22	0,26	-0,003	0,236
Papanek and Kyn (1985)	10	-0,19	0,17	-0,015	0,159
Park (2010)	6	-0,36	0,28	-0,04	0,342
Park (1996)	14	-0,43	0,37	-0,05	0,326
Perugini and Martino (2008)	20	-0,17	0,25	0,039	0,158
Psacharopoulus (1977)	4	0,33	0,48	0,373	0,072
Ram (1984)	8	0,06	0,40	0,235	0,129
Ram (1981)	3	-0,275	0,315	-0,048	0,318
Savvides (1998)	6	-0,331	-0,17	-0,26	0,058
Sylwester (2002)	13	-0,26	0,79	0,33	0,295
Sylwester (2005)	4	-0,083	0,19	0,102	0,125
Sylwester (2003)	10	-0,167	0,69	0,502	0,27
Sylwester (2003)	12	0,29	0,41	0,363	0,367
Tsai (1995)	19	-0,25	0,22	-0,09	0,134
Tsakloglou (1988)	2	0,47	0,55	0,51	0,055
Tselios (2009)	17	-0,12	0,22	0,02	0,110
Winegarden (1979)	2	-0,41	0,53	0,056	0,665
Hermes (2014)	6	-0,27	0,04	-0,17	0,132
Chu and Hoang (2020)	5	-0,37	-0,29	-0,328	0,031
Sanches and Perez Corral (2018)	4	-0,33	0,17	-0,051	0,225
Özdemir (2019)	26	-0,722	0,07	-0,306	0,320
Lee and Vu (2020)	15	-0,67	0,19	-0,15	0,244
Le, Nguyen, SU, Tran-Nam (2020)	55	-0,97	-0,73	-0,894	0,058

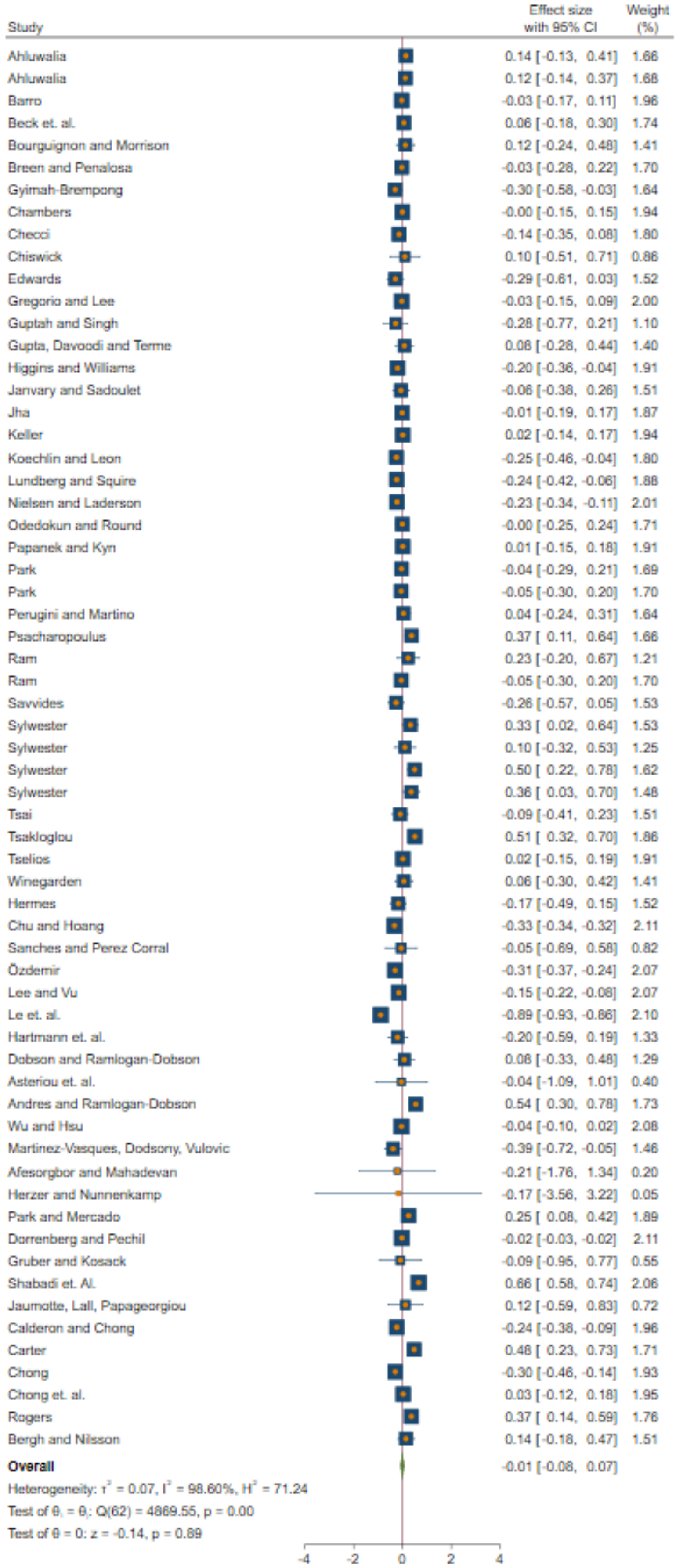
Hartmann et. al. (2017)	31	-0,49	0,31	-0,20	0,257
Dobson and Ramlogan-Dobson (2012)	11	-0,345	0,97	0,077	0,430
Asteriou et. Al (2014)	31	-0,89	0,98	-0,043	0,536
Andres and Ramlogan-Dobson (2011)	29	0,003	0,78	0,54	0,148
Jyun-Yi Wu, Chih-Chiang Hsu (2012)	16	-0,32	0,334	-0,04	0,161
Martinez-Vasquez, Dodsonv, Vulovic (2012)	17	-0,644	0,08	-0,385	0,242
Afesorgbor, Mahadevan (2016)	13	-0,64	0,47	0,211	0,312
Herzer, Nunnenkamp (2013)	4	-0,46	0,029	-0,17	0,220
Park and Mercado (2018)	4	0,026	0,544	0,25	0,257
Doerrenberg and Peichl (2014)	19	-0,63	0,378	-0,023	0,273
Gruber and Kosack (2014)	7	-0,633	0,419	-0,087	0,408
Shabadi et. Al. (2018)	1	0,663	0,663	0,663	0
Jaumotte, Lall, Papageorgiou (2013)	43	0,001	0,32	0,121	0,081
Calderon and Chong (2009)	3	-0,35	-0,15	-0,236	0,103
Carter (2007)	6	0,44	0,59	0,48	0,055
Chong (2004)	11	-0,46	-0,13	-0,3	0,107
Chong et. Al. (2009)	9	-0,05	0,08	0,029	0,043
Rodgers (1983)	1	0,367	0,367	0,367	0
Bergh and Nilsson (2010)	19	-0,07	0,31	0,144	0,120
Nr. of Studies: 61 *Total:*	916	-0,97	0,98	-0,043	0,377

As the dataset in this thesis contains all relevant estimates from the each of the included studies, it can be characterized as an *all-set* meta dataset. Other common meta-datasets are *average-set, independent-set* or *best-set* (Broderstad 2015-06-02, 57; Doucouliagos and Paldam 2008, 7). The average set consists of the mean of the effect size in the given studies; the independent-set contains results that are considered different from a conceptual point of view; *the best-set* is a selection of the estimates that the researcher considers as the "best" estimate from the exact study. All these datasets have advantages and disadvantages that must be considered when designing a meta-analysis.

Using an all-set, one can compare a larger set of estimates, allowing for a more detailed analysis. One will also avoid the possible selection bias that might occur when constructing a best-set, and it allows for comparing research that is conceptually similar, in contrast to the independent-set. A disadvantage is that the process of selecting "all relevant estimates" from each study also includes a subjective evaluation, which makes this type of dataset also open for selection bias to a certain degree. Another disadvantage is that the all-set increases the chance for interdependence between the estimations, which requires additional statistical testing. As such, testing will be carried out within this analysis and considering how all sets can be used to evaluate the heterogeneity between different studies at a more detailed level, the all-set analysis is considered the appropriate construction for a data-set in regards to the research question.

This analysis is also supplied with an average set that was manually created, with the average effect size from each of the studies. The average-dataset is only used to illustrate how the studies differ in effect size and precision (1/SE) through a forest plot (see figure 1). The size of the box illustrates the power, or meaningfulness, of the data, and the thin vertical line marks the study specific deviations from the overall effect size. The thin horizontal lines are the magnitude of the confidence interval, where longer lines indicate wider confidence intervals and therefore less reliable data. We can see in the plotting of the data that the overall unweighted effect size is close to zero.

Figure 4.1 Forest plot of average partial correlations



4.2 Variables

With information on effect sizes and other necessary data such as standard errors and sample size, statistical inferences about the reported effects of education on income inequality can be modelled. Variables such as partial correlations and precision of each individual estimate (standard error of the partial correlation) are referred to as *essential variables* in the meta-analysis because they allow us to compare the effects of education on income inequality, as well as the strength of these estimates (Stanley and Doucouliagos 2012, 20). These variables are included as graded variables in this MRA. However, to answer research question (Q3) more information is required in the MRA to assess why effect sizes may differ between these studies. Including other variables that separates the studies from each other based on possible sources of heterogeneity is therefore essential.

Card (2011) refers to the coding study characteristics as creating *interesting moderator variables.* Choosing moderator variables that might allow us to say something about why the studies on this relationship differ, requires a familiarity with the research field and more detailed knowledge about the studies that are included in the analysis. Investigating the variation in findings on the effect of education on income inequality, three main sources of variation within social sciences are considered: differences in *data, measurement and estimation methods.* The process of deciding more specifically what moderator variables to include is both an inductive and deductive process. Here, moderator variables were selected firstly by contracting common moderator variables described by works on meta-analysis, such as Stanley and Doucouliagos (2012) and Card (2010). Secondly, through the screening of each study, characteristics that werecommon but also varied between the studies were included in the dataset as moderator variables. Only the most theoretically interesting variables were still included to avoid issues with multicollinearity.

Estimation differences is considered a common source of heterogeneity in findings, and these are often controlled for in meta-analyses (Abdul Abdullah 2015, 6). Within the selected sample, most estimations are derived from using OLS, GMM, SLS or IV methods. A central difference among these estimation methods is that they control for endogeneity in regression models in very different ways. As OLS is the most commonly used estimation method in the sample, it is of interest to separate the studies that use OLS from those who do not. This way, we can see if controlling for endogeneity in other ways than through OLS has a statistically significant correlation with generating different results. The variable *Endogeneity* is generated, coding all studies that do not use OLS. Another possible source of variation due to estimation techniques is due to the authors choice of adopting a fixed effects model, or a random effects model. Because general OLS is unsuitable to identify if observed effect of one variable on another is caused by effects over time or between countries, many researchers adopt OLS with fixed effects to control for these differences. It is therefore theoretically interesting to see if this measurement leads to differences in findings. A variable controlling for studies that uses OLS with FE is included as a moderator.

Measurement differences, meaning how each research has measured their independent and dependent variable, is considered "often the most important explanatory variables" and essential to include in a meta-analysis (Stanley and Doucouliagos 2012, 86). Within this sample, all studies that measure the dependent variable income inequality use either the Gini coefficient or an income ratio to measure income inequality. The variable *Gini* is created to control for the differences between these two main groups of measurement. Further, because of the large variation within different types of income ratios, different types of income share ratios, such as share of the bottom-, middle- and upper class were coded. These differences allow for us to investigate if these different types of measurement results in systematic differences in findings and see if education has a higher reported effect on income inequality regarding the share of income for the wealthy, middle class or bottom share of the population. Regarding the measurement of education, most studies use the attainment or enrolment in primary, secondary and tertiary education, literacy rate or average years of schooling amongst the population. Only primary, secondary, and tertiary schooling is included as moderator variables here, due to possible multicollinearity. In addition to being prevalent within the chosen studies, differentiating between different types of education is appealing considering the previously presented theory on the subject, where authors highlight those different types of education report stronger or weaker effect of education on income inequality. Lastly, many of the studies have different expectations regarding what effect education might have on income inequality, especially related to when we should expect to see any effect. I therefore chose to code research that does lag their variables (to expect a future effect of the present education) and research without lagged variables.

Data characteristics. The time-period of study can be an essential source of differences in findings, due to changes in the world economy and different types of development. A binary variable that differentiates between studies that published before and in 1990, and studies that were completed after 1991. The cut-off between studies published until and after 1990, is set based on the many changes following rising access to education since 1990 (Lee and Lee 2016). In addition to time of study, many researchers register data based on regional differences. In the earlier dataset, Abdullah et. al. controlled for estimates regarding income inequality in Africa, Latin-America, and OECD. Because none of the included studies explicitly stated that any estimations were specifically from Africa or Latin-America, it will not be possible to include these regional variables, only to repeat the findings from their analysis. Only the regional variable OECD is included as a dummy. A dimension that is not in related to the specific research design or study characteristic is the professional affiliation of the author of the study. However, the academic discipline authors publish within has been found to be an important source of variation in research findings (Stanley and Doucouliagos 2012, 89) (Doucouliagos and Laroche, 2003; Doucouliagos and Paldam, 2008). Measure professional affiliation is challenging because departments are often divided based on different considerations at different universities, as some researchers work together in teams or on topics that are not directly related to their academic degrees. Therefore, academic disciplines were due to the type of journal that they were published within. Four categories dominate the sample: political journals, sociology journal, economic journals, and development journals. The latter category, development journals, is included because it is a very specific type of journal, where addressing the closer academic field is neither appropriate nor beneficial to the analysis. Finally, the two most used sources of data on income inequality and education amongst the studies were coded. The *Barro dataset* was included to control for data differences on education, and datasets from the *World Bank* as a source of data on income inequality. A presentation of the variables included in this MRA is as follows:

Table 4.2 Description of variables

Variable	**Description**	**Min**	**Max**	**Mean**	**Std. Dev.**
Essential variables					
Partial correlation	Partial correlation between education and income inequality	-0,97	0,98	-0,043	0,377
Standard error	Standard error of partial correlation	0,002	0,99	0,027	0,054

Estimation differences					
NonOLS	BV: 1 = Does not use OLS	0	1	0,342	0,474
OLS FE	BV: 1 = Uses OLS with fixed effects	0	1	0,657	0,474
Lagged variable	BV: 1 = uses lagged variable	0	1	0,530	0,499
Measurement differences					
Gini	BV: 1 = if Gini measure is measured	0	1	0,680	0,468
Rich	BV: 1 = measures income shares of top earners above 20% of population	0	1	0,775	0,267
Middle	BV: 1 = measures income shares of middle 40% of population	0	1	0,262	0,159
Bottom	BV: 1 = measures income shares of bottom 40% of population	0	1	0,107	0,309
Middle&Bottom	BV: 1 = measures income shares of middle 40% and of bottom 40% of population	0	1	0,033	0,180
Primary	BV: 1 = primary enrolment/attainment is measured	0	1	0,112	0,316
Secondary	BV: 1 = secondary enrolment/attainment is measured	0	1	0,337	0,473
Tertiary	BV: 1 = tertiary enrolment/attainment is measured	0	1	0,129	0,335
Literacy rate	BV: 1 = literacy rate is measured	0	1	0,058	0,233
Data characteristics					
End before 1990	BV: 1 = the study ends before or in 1990	0	1	0,284	0,451
New dataset	BV: 1 = if the study is published after 2010	0	1	0,402	0,499
OECD	BV: 1 = OECD countries are specified	0	1	0,575	0,494
Political	BV: 1 if published in political journal	0	1	0,054	0,227
Sociological	BV: 1 if published in sociology journal	0	1	0,052	0,222
Economic	BV: 1 if published in economic journal	0	1	0,557	0,499
Database World Bank	BV: 1 if data on income inequality is from WB	0	1	0,493	0,500
Database Barro	BV: 1 if data on education is from Barro dataset	0	1	0,360	0,481

BV: Binary variable. OLS: Ordinary least squares.

This table displays the variables, how they were measured, and to their frequency in the data. Regarding the estimation differences, we can note that 64,3% of the included studies used OLS, whereas 35,7% did not. We also observed that 27,9 estimations were created with models assuming that the variation was caused by fixed effects.

Gini is the most frequent measure of income inequality, and secondary attainment or enrolment is the most frequent measure of education in the sample. Only 53% of the studies used lagged variables.

Regarding the type of data, 28% of the studies ended before 1990. Another time difference is the one between the original dataset and the studies collected specifically for this thesis. The original dataset included studies produced between 1970 and 2010 and the additional dataset contained studies from 2010 until today, the old meta data making up 66% of the data and the new meta being 33%. Comparing the possible differences in effect sizes between these two variables entails comparing if there was a change in reported partial correlations on this topic over time. Forty percent of the studies included in the all-set is from the new dataset.

Regarding the type of data, we can see that 51% of the research was gathered from economic journals, 34% from development journals, 5% from exclusively sociological journals and 5% from exclusively political journals. The two most frequently used databases, the World Bank on income inequality and Barro on education, are also highly represented in the numbers, with almost 50% using data from the World Bank on income inequality and almost 40% using data from Barro's database for data on education.

5 Results

5.1 Descriptive statistics

The 61 articles in our sample were published within the period 1971-2020. The median article was published in 2005, and the most articles were published in 2014. The average reported partial correlation in a year varied both between high and low positive effects, as well as between positive and negative effects. The average reported effect size starts out positive in the beginning of the 1970s, but already in the 1980s the first negative overall reported effect size appeared. There was consistent variation during the 50 years analysed, such that the average partial correlation per year varied between positive and negative 15 times. Overall, there is a general decline in the average partial correlation by year, illustrated by the fitted line (Figure 5.1). There was a decrease from 0,10 in 1970 to -0,46 in 2022. This is a decrease in 0,01 units pr. year between 1970 and 2020. This can be interpreted as a general change in the field of research, where more negative effects are reported after 2005 than before. It must be mentioned that the effect sizes described here are not weighted by sample size or precision. This is for the sake of the later discussion on possible changes in the research findings on this topic over time, in section 6.5. By including a timeline showing unweighted average partial correlations, this compared to the results on the time variables in the multiple MRA, which are weighted by precision.

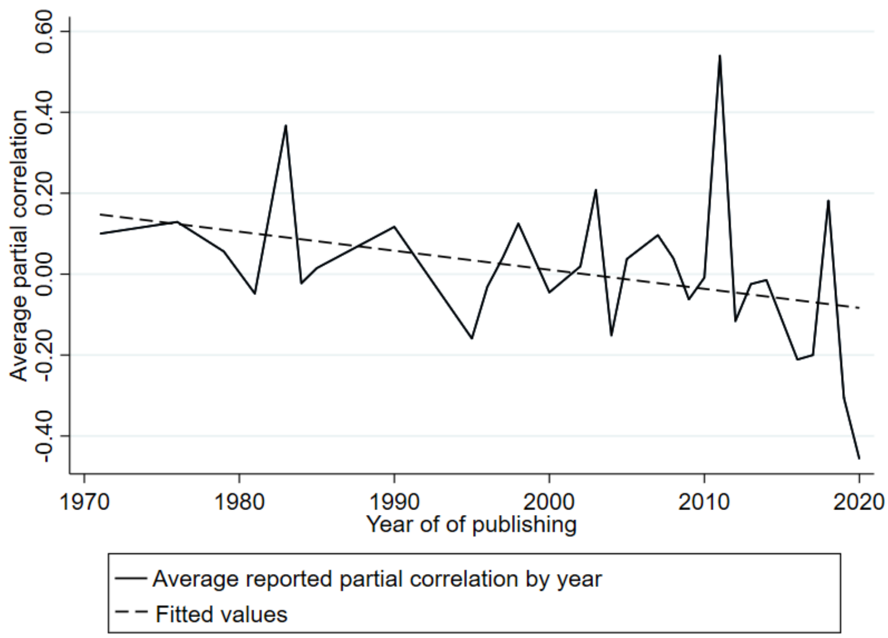


Figure 5.1 Time trends in average reported effects of education on income inequality pr. year reported by partial correlations

5.2 Publication Bias

To fulfil this stage of controlling for publication bias, a scatter plot of the effect estimates from individual studies against the measure of each study's size or precision is created, serving as a funnel plot (Figure 5.2) (Sterne et al. 2011, 1). The 917 reported empirical estimates from the studies (the partial correlations) are here plotted against the precision of the estimate (1/SE). At the bottom of the y-axis we find the estimates with the highest inverse standard errors, and the more precise estimates at the top of the y-axis. The effect size goes from -1 to 1. The funnel plot is created with Statas own meta-analysis toolbox. An inverted funnel shape is the shape that we should look for, as this indicates that there is no publication bias (Stanley and Doucouliagos 2012, 55).

The funnel plot can be considered quite symmetrical, with the shape of an inverse funnel where most studies have a low inverse standard error and therefore are closer to the bottom. The partial correlations appear to be equally spread out on both sides of 0, and therefore not indicating any bias or skewed reporting in findings. Most of the studies are clustered between 0-200 on the inverse standard error, with a wide spread of values on partial correlation.

High precision can be observed for estimates from four studies (Doerrenberg et al. 2014, Jaumotte 2013, Andres 2011, Hartmann 2017) with estimates containing a precision score over 250. 41 of these estimates are considered to have extremely high levels of precision. This high precision score comes from the fact that they have clustered their standard error at country level or by panels. This difference is an important explanation of their position as outliers in the funnel plot. Still, their estimates appear at both sides of the effect size scale, with both strong positive and strong negative effects. Estimates that are outliers due to their precision are harder to control for in the meta-analysis, as weighing the studies by sample size to a large degree controls for studies with little precision. Studies with such high precision are therefore harder to control for. Stanley and Doucouliagos highlight that estimates that are not outliers due to coding mistakes, should be perceived as important datapoints that should not be removed unless there is a valid reason to do so (Broderstad 2015-06-02, 64; Stanley and Doucouliagos 2012). As one is aware of the reasons for the high precision of these estimates, have checked that these outliers are not due to coding mistakes and the outliers are not affecting the expected symmetry of an inverse funnel, there is no apparent or valid reason to remove these estimates.

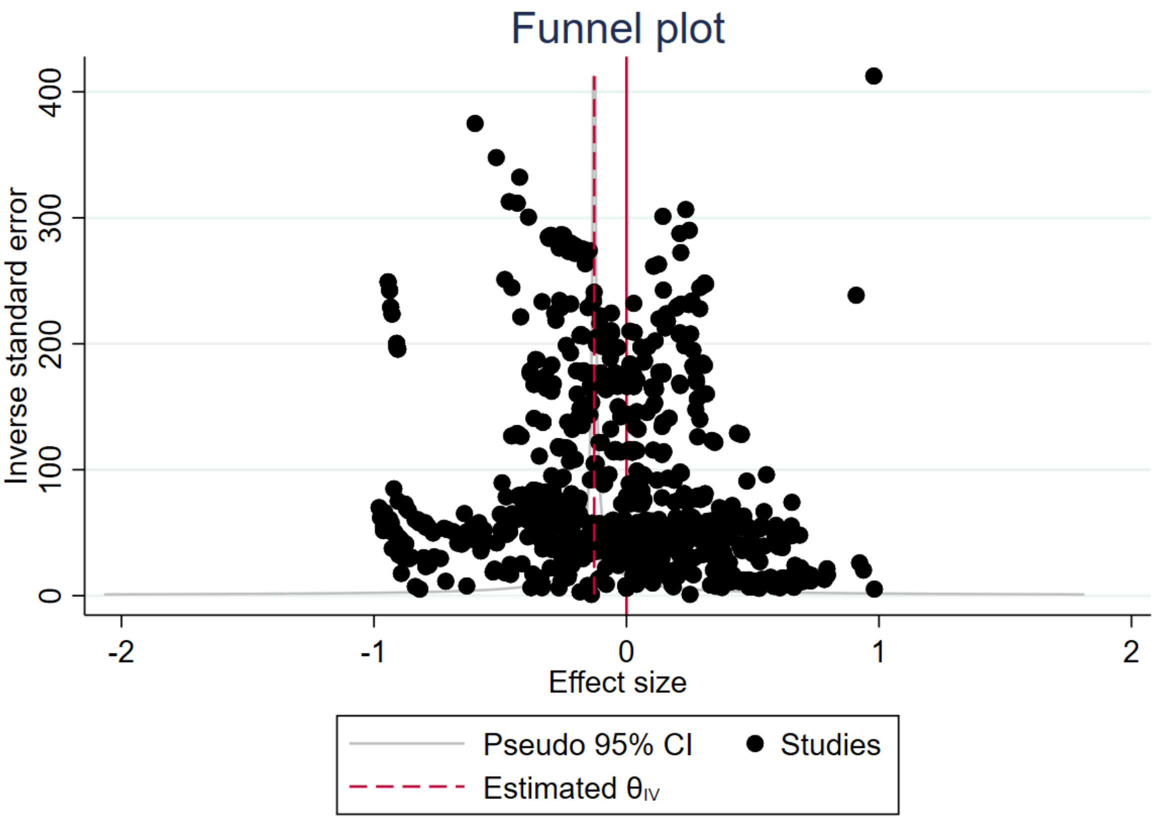


Figure 5.2 Funnel plot of partial correlations against inverse standard error (precision)

To thoroughly test for publication bias using the FAT and PET test, these are applied to four different models on the entire dataset (the all-set). The FAT-test itself is done by creating a regression analysis with the effect size and standard error as variables, weighed by precision squared ($(1/SE)^2$) and checked for robustness. Further, a PET-test is conducted, to test for an true underlying empirical effect beyond the potential distortion due to publication selection bias (Stanley and Doucouliagos 2012, 62). It involves testing the null hypothesis: $\beta_0 = 0$, where the β_0 is the coefficient on precision in equation (Stanley and Doucouliagos 2012, 63). The tests consist of a regression of partial correlation against precision of the standard error of the partial correlation, where the constant of the regression reveals the "meta-average" partial correlation, controlled for publication bias.

There is no statistical evidence indicating the presence of publication bias in the gathered literature on the relationship between education and income inequality. The level of publication bias in the sample was determined to be little or moderate based on the FAT test, as it is either below 1, or insignificant in all models. See section 3.3 (D) for interpreting guidelines.

The constant derived from the PET test in the WLS model is a meta-average of -0,025. According to Cohens guidelines $\pm$0.10 is considered a small effect, $\pm$0.30 as a medium effect and $\pm$0.50 as a large effect. The results are therefore considered to show a very small to no underlying empirical effect beyond potential distortion due to publication bias. investigating the other PET-results, no values exceed -0,043 and can also be considered show a small to no underlying effect, even when using Doucouliagos decreased standards for interpretation. The results from the PET test are also not significant at a 95% confidence level, with (p > 0,05) in any of the WLS or random effects models. In the fixed effects model, there is a negative significant effect of -0,043. Being a result from a fixed effect model, this should not be given much consideration, due to earlier presented reasons. The significant effect presented is also quite small, using the chosen guidelines for interpretation. No results for an underlying effect between the two subjects of investigation raises the question of what to do as a meta-analyst when the results from the PET-test is divergent from the test of the weighted mean partial correlation (classical meta-analysis).

With divergent results on the documented effect of education on income inequality, one should be cautions to rely on either the mean partial correlations detected, or the results from the PET-test exclusively to conclude on a real effect of education on income inequality. Stanley (2017) model use simulations from social/personality psychology research to test the reliability of the PET-test under different circumstances (T. D. Stanley 2017, 581). He identifies that the PET-test performs poorly in research areas where there is very high heterogeneity of results from study to study and identifies how “small effects can be doubled and small to medium effect sizes can be manufactured from nothing” using a PET-test. However, it is argued that this test is likely to be more reliable than conventional meta-analysis because classic meta-analysis face major problems with data that is affected by publication bias. As we know, one should always assume a slight presence of publication bias, even this cannot be detected statistically.

Investigating several meta-analyses, no common procedure dealing with insignificant PET-test that differ from the mean partial correlation results is identified. Authors vary in their practice of dealing with insignificant results on their PET-test, with many simply moving to conducting a multiple MRA without concluding on its importance for the analysis as a whole (Nikos Benos 2014, 677; Yesilyurt 2019, 358). Here, the results from the classical meta-analysis and the PET test are used to state that there is a reported negative mean effect of education on income inequality, but that this effect is hard to detect statistically when controlling for (non-identified) publication bias.

As there is no publication bias detected within the sample, no PEESE-test is carried out. Having controlled for publication bias, and not found the presence of problematic levels of homogeneity across the studies in the sample – answering (Q2): there is no indication of publication bias in the literature regarding educations effect on income inequality.

Table 5.2 Results from FAT-PET MRA testing

	WLS Robust (1)	WLS Cluster robust (2)	FE Clustered (3)	MEML Clustered (4)	Robust (5)
FAT	-0,001 (-0,02)	-0,001 (-0,03)	-0,04 (-4,85)***	0,0002 (0,01)	-0,06 (-4,35)***
PET	-0,025 (-1,92)	0,006 (0,19)	-0,043 (-4,38)***	0,006 (0,19)	0,006 (-0,19)
Nr. of obs.	916	916	916	916	916
Nr of studies	61	61	61	61	61

Note: t-statistics in brackets. Model 1 and 2 contains weighted least squares (WLS) regression with precision squared as weights with robust standard errors. Model 2 reports cluster-robust standard errors, clustered by studies. Model 3 uses fixed effects panel with study dummies (FE) also clustered by studies. Model 4 is a mixed effect multilevel model with random effects. Model 5 is a standard regression with robust standard error. *** means significant at 95% level.

5.3 Model Results

5.3.1 Model results not accounting for publication bias

In the models for estimating unweighted and weighted means, all results indicate that there is a negative effect between education and income inequality, even though this effect is not very substantial in some models. It varies from the highest value of -0,007 in the unweighted average set, to the lowest -0,44 in the all-set with OLS and fixed effects and weighed by sample size. The studies with different ways of controlling for endogeneity find smaller effects than those using OLS with fixed effects. Using Cohens standard for interpreting partial correlations, where 0,10 is a small association, and 0,25 a medium association and over 0.40 a large association, most associations in this case span between a small and medium effect. The weighted average of -0,213 can be considered a small to medium effect. Using Doucouliagos (2011) guidelines for interpretation of partial correlations in economic research, -0,213 is considered a moderate to large effect, as it is above 0,17 which is considered a moderate effect, and smaller than 0,33, which is considered as a large effect. In the set with estimated calculated from OLS with fixed effects assumption and weighted by sample size, we find an effect of -0,46, which qualifies as a large effect. One should however be reluctant to use the results for fixed effects models here, as the assumption for this model is rarely met in meta-analysis. The findings further illustrate that the choice of measurement, whether one chooses an OLS with fixed effects or a different model with random effects, results in quite different associations – a relationship that must be further investigated in the later control for heterogeneity due to such characteristics. The Cochran's test show as expected that there is heterogeneity amongst the findings. The weighted mean in the all-set is the preferred mean for discussing the overall reported effect of education on income inequality in scholarly literature today. The result from this section is used to answer (Q1), identifying a small to moderate negative reported effect of education on national levels of income inequality in research today.

Table 5.1 Education effects on democracy in different sets, with different weights

	All-set	Average set	All-set, *OLS FE*	All-set, Endogeneity
Unweighted mean	-0,046	-0,043	-0,29	-0,09
Weighted mean (N)	-0,213	-0,213	-0,44	-0,15
Weighted mean (FE)	-0,128	-0,209	-0,28	-0,097
Weighted mean (RE)	-0,046	-0,195	-0,29	-0,097
95% CI (N)	-0,24, -0,18	-0,23, -0,18	-0,52, -0,36	-0,19, -0,11
95% CI (FE)	-0,24, -0,14	-0,25, -0,16	-0,36, -0,21	-0,16, -0,02
95% CI (RE)	-0,23, -0,12	-0,23, -0,15	-0,02, -0,02	-0,17, -0,02
Cochran's Q test	0,000	0,000	0,000	0,000
Nr. of studies	61	61	25	36
Observations	915	30	146	313

*N = sample size, FE studies with fixed effects assumptions, RE uses random effects assumptions to estimate confidence intervals. The row for Nr. of studies, studies with both OLS and other models for heterogeneity are placed in the group "endogeneity" here, while studies with only estimates created by OLS and fixed effects are in grouped as OLS FE.

5.3.2 Model results after accounting for publication bias

Eight models are created to control for heterogeneity between the findings. Four WLS models, two fixed effects and two random effects models. To control for within- and between study dependency a WLS with clustering by study, and the random effect model use multilevel modelling is included. 23 moderator variables are chosen to control for across the models. Because using many variables can increase the problem with multicollinearity, a VIF test is conducted. The mean VIF in the main WLS model of the all-set with all variables is 9 and multicollinearity should in general not be considered to pose a threat to our sample.[6]

Model 1 contains a simple WLS regression model of all the chosen variables, with precision squared as weights and robust standard errors. Model 2 is a another WLS model, but with cluster-robust standard errors, to avoided author dependency, which might occur when you have many estimates reported from the same study.[7] As the two first models are very similar, their beta coefficients are corresponding, but their standard errors will vary due to the clustering of the estimates in model 2. Models 3 and 4 are a fixed effects regression clustered by study (FE) and a mixed (random) effects (MEML) multilevel-models. In the fixed effect study dummy variables for the authors are included, to check for author dependency. In MEML the partial correlations are assumed to be nested within studies, and hence within-study dependency is controlled for. This MEML model uses RE assumptions, and is modelled with two levels.

In the last four models, only a selection of the variables is included in the regression. These models are constructed based on a G-to-S procedure, using WLS clustered by robust standard error. G-to-S involves, as mentioned, removing one variable at the time, based on statistical significance. Only the one specific statistical significance remains. When using a two tailed test and a significance level of 95% results in a set of 7 statistically significant moderator variables are the results. The significance level of 95% is considered appropriate as the dataset contains almost 1000 estimates. A lower level of significance is considered to harm the reliability of the results. The WLS models are weighted with precision (1/SE) and the FE and RE with precision squared as weights $(1/SE)^2$ (Ahmadov 2013, 1257).

When using statistical tests to evaluate the created models, heteroskedasticity is detected between the findings. A Breusch-Pagan/Cook Weisberg test for heteroscedasticity suggests that the model contains heteroscedasticity, with p=0,000. As the model is not homoscedastic, this confirms our theoretical assumption about preferring a WLS approach instead of using OLS (Stanley and Doucouliagos 2012, 61). Recalling the path described earlier, The Breusch-Pagan Lagrange multiplier (BPLM) is used to check for significant study-level effects. Our results of

---

[6] Mean VIF under 10 is usually interpreted to not indicate problems with multicollinearity in our model (Micheal and Abiodun 2014, 410).

[7] By clustering the robust standard errors by studies, I account for dependency within the specific study, but not dependency between different studies produces by the same author.

chi 2819 < p=0,000 indicates that there are study level/within study effects in or data. According to Stanley and Doucouliagos one should then use a Hausman test to identify the correct model specification. Our results (21,38; p > 0,092) indicate that differences in our coefficients are not systematic, meaning that a random effects model should be preferred over the fixed effect model.

Based on these post-estimation tests, the second model, WLS clustered by study ID and with robust standard errors, ends up being the preferred model for this analysis. The result from this model is therefore used to conduct the G-to-S approach. The results from all the models will still be discussed, mainly to control for publication bias, as suggested by Stanley and Doucouliagos (2012, 105). However, as the assumptions of the fixed effects and random effects models are not the preferred models for the given data, the findings from these models will not be discussed at length.

The findings were different among the eight models, and only one variable was significant at 95% level in all models. The results from random effects models differ clearly from the results of the WLS models. In this comparison, one can observe the coefficients changing from positive to negative correlations and see large changes in their t-statistics. Overall, the mixed effect multilevel model reports slightly smaller effects than the other models. Four of the chosen variables were not significant at any level of significance, in any model.

Due to the results from the model evaluation in section 3.4.1, Models 1 and 2 are preferred due to the existence of heterogeneity, the presence of study level effects, and the low chances of these effects stemming only from random effects between the studies of fixed effects assumptions. Model 2 is preferred to Model 1, as it controls for within-study effects by clustering the estimations. The result from this model was therefore used to conduct the G-to-S approach. When accounting for the clustered robust standard errors in Model 2, four of the variables went from being significant to not significant in the second model. Subsequent models did not represent significant improvement to the results, including efforts at model reduction in model 5 and Model 6 using the G-to-S approach. In Model 2, seven variables are significant at a 95% level of confidence. These are OLS FE, Endogeneity, Gini, Income share of Middle and Bottom, Primary education, New dataset and database Barro.

Interpretation

Interpreting a meta-analysis resembles interpreting regular regression models, as it has a dependent variable that is predicted according to values of one or more independent variables. In meta-analysis, the dependent variable is the effect size, and the independent variables are study characteristics that is hypothesized to influence the effect size. Therefore, a coefficient from a meta-regression analysis describes how the dependent variables (the intervention effect) changes with units in the dependent variable (the potential effect modifier). The statistical significance of the regression coefficient is used to test if there is a linear relationship between intervention effect and the explanatory variable, where the effect size is weighted by study quality.

If a variable is significant, this tells us that the specific variable can help explain the variation of the different partial correlations. If a coefficient is significant and negative, this is interpreted to mean that studies with this specific characteristic, on average report larger negative (or smaller positive) correlations between education and income inequality. In the same way, a significant positive correlation is interpreted to indicate that studies with this characteristic are more likely to report a stronger positive (or smaller negative) effect of education on inequality, compared to studies without this characteristic. For efficiency, the further explanation is only focused on the variables that are significant in the models.

Estimation differences

Regarding estimation differences, OLS FE and endogeneity have moderate to high negative effects. This means that studies with OLS fixed effects and other models than OLS, reported a larger negative effect of education on income inequality, or a smaller positive effect than other studies. Studies that use OLS without fixed effects, or other models than OLS would therefore be expected to result in a stronger positive relationship between education and income inequality.

Measurement differences

Investigating the results on measurement differences, the variable *Gini* is significant in model 1,2 and 4, but loses its significance in the WLS G-to-S model with studies clustered. These results are not considered robust across the preferred models. As controlling for within-study effects in the second G-to-S model against other significant variables results in no relationship between Gini measurement and reported effect size, this is not a difference that is further discussed.

In contrast the measurement of income inequality share ratio of income of middle and bottom share of the population is positive and statistically significant at 95% level across all models. This effect is in our preferred WLS models between 0,46 and 0,49, which is also a relatively strong correlation. This variable captures the studies that use bottom 41% up to 80% of the population, a size of ratio that is only used in 3% of the reported studies. The coefficient of the Income Share Bottom and middle-class variable indicates that, compared to studies that use coefficient or other income share ratios, studies that this measurement report a stronger positive relationship between education and income inequality. This finding is regardless only based on an extremely small part of the sample.

Regarding the measurement of education, primary schooling is statistically significant with a negative coefficient in the ordinary WLS and the WLS with G-to-S approach. This indicates that compared with secondary schooling and tertiary schooling, primary schooling explains more of our partial correlations. The significant results from the WLS models varies between -0,276 to -0,289. Studies using primary education as a measurement for education are according to these findings, more likely to report a negative relationship between education and income inequality.

<u>Data characteristics</u>

The variable New dataset is significant in all WLS models, and only insignificant in the first multilevel model. This indicates, to a certain extent, that there is a difference between the partial correlations that are reported in the new dataset compared to the old dataset. This could entail coding mistakes, or a real change in the results on this research field – meaning that there is a reported stronger partial correlation in newer studies. The coefficient changes from negative to positive moving from WLS to fixed effects models, and it is therefore hard to say whether one should estimate that studies published since 2010 report larger positive (or smaller negative) effects, or the other way around. Combined with our earlier fitted values line for average effect size published by year, this seems to suggest that studies published after 2010 has a higher chance of reporting a stronger negative effect of education on income inequality.

The variable Barro database is also significant in all models, except from the multilevel models. Compared to the explanatory power of the World Bank database, the Barro database can be considered to have a significant explanatory power on parts of the partial correlations. Interpreted with caution, and due to the not suitable premises for the multilevel effects model, the Barro coefficient indicates that studies using data from this database tends to report a more positive effect of education on income inequality.

Table 5.3.3 Multiple meta regression analysis of education-income inequality research

Heterogeneity (Moderator variables)	**WLS R Full (1)**	**WLS R-CL Full (2)**	**FE Full (3)**	**MEML Full (4)**	**WLS R G-to-S (5)**	**WLS CR G-to-S (6)**	**FE G-to-S (7)**	**MEML G-to-S (8)**
Intercept ($\boldsymbol{\beta_1}$)								
Estimation differences								
OLS (FE)	-0,573***	-0,573**	-0,027	-0,068	-0,418***	-0,418**	-0,043	-0,09*
	(-2,37)	(-2,09)	(-0,34)	(-1,40)	(-6,08)	(-2,42)	(-0,48)	(-1,77)
NonOLS (Endogeneity)	-0,218***	-0,218**	0,066*	0,047	-0,241***	-0,241**	0,068*	0,032
	(-3,42)	(-2,62)	(1,79)	(1,15)	(-6,05)	(-2,50)	(1,86)	(1,01)
Measurement differences								
Gini	-0,203***	-0,203**	0,057	0,058	-0,122***	-0,122	-0,053	-0,053
	(-2,89)	(-2,42)	(0,69)	(1,15)	(-0,25)	(-1,07)	(-1,16)	(-1,57)
Rich	-0,027	-0,027	-0,067	-0,068				
	(-0,18)	(-0,54)	(-0,91)	(-1,36)				
Middle	0,034	0,034	0,317***	0,307***				
	(0,26)	(0,20)	(3,58)	(4,71)				
Bottom	0,069	0,069	0,27***	0,263***				
	(0,56)	(0,40)	(4,37)	(5,55)				
Middle&Bottom	0,491***	0,491**	0,484***	0,437***	0,460***	0,460*	0,325***	0,330***
	(2,65)	(2,19)	(4,37)	(6,95)	(6,78)	(1,92)	(2,72)	(5,85)
Primary	-0,289***	-0,289**	-0,188	-0,05	-0,276***	-0,276**	0,078	0,088***
	(-1,69)	(-2,11)	(-0,73)	(-1,06)	(-8,80)	(-1,84)	(0,99)	(2,59)
Secondary	-0,024	-0,024	-0,202	-0,147***				
	(-0,28)	(-0,34)	(-1,40)	(-3,85)				
Tertiary	0,171	0,171	-0,013	0,083*				
	(1,50)	(1,41)	(-0,09)	(1,67)				
Literacy rate	-0,028	-0,028	-0,192	-0,13**				
	(-0,30)	(-0,33)	(-1,34)	(-2,16)				
Lagged variable	-0,167	-0,167*	-0,223***	-0,159***				
	(-1,57)	(-1,76)	(-9,72)	(-2,66)				
Data characteristics								
End before 1990	0,047	0,047	-0,033	0,06				
	(0,83)	(0,54)	(-0,25)	(0,70)				
New dataset	0,329***	0,329**	-0,202*	-0,036	0,208***	0,208**	0,272**	-0,056

	(2,99)	(2,16)	(-1,88)	(-0,38)	(4,42)	(2,08)	(-2,56)	(-0,75)
OECD	0,106	0,106	0,004	0,01				
	(1,82)	(1,59)	(0,30)	(0,27)				
Political	0,196*	0,196	0,224*	0,016				
	(1,33)	(1,24)	(-1,90)	(0,12)				
Sociological	-0,05	-0,05	-0,139**	-0,242				
	(-0,48)	(-0,31)	(-2,47)	(-0,98)				
Economic	0	-0,0003	-0,561***	-0,11				
	(-0,01)	(-0,01)	(-4,02)	(-1,40)				
Database World Bank	-0,131*	-0,131	-0,449***	-0,021				
	(-1,83)	(-1,45)	(-13,06)	(0,30)				
Database Barro	0,377***	0,377**	-0,301***	0,113	0,241***	0,241**	-0,357***	0,052
	(3,22)	(2,43)	(-17,36)	(1,38)	(5,22)	(2,53)	(-7,43)	(0,73)
Publication Bias								
SE of partial correlation	-0,498	-0,498	0,075*	0,119	-0,111	-0,111	0,071	0,125
	(-0,43)	(-0,53)	(1,77)	(0,70)	(-0,25)	(-0,16)	(1,53)	(0,68)
R^2	0,259	0,259	0,686	-	0,212	0,212	0,636	-
MLR² level 1	-	-	-	0,169	-	-	-	0,08
MLR² level 2	-	-	-	0,150	-	-	-	0,10
F-test	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
No. obs. (n)	915	915	915	915	915	915	915	915
Nr. Studies (*k*)	61	61	61	61	61	61	61	61

***p<0,01, ** p<0,05, *p<0,1. Partial correlation is the dependent variable. Numbers in parathesis are t-statistics. MLR² for the MEML models are calculated using Stata, following the formula and procedure proposed in Snijders & Bosker (2012) for calculating R^2 for MEML models (Bosker 2012, 112-113).

5.4 Evaluation of MRA

R^2

Model fit R^2 varied from 0,08 to 0,0686. The average R^2 of the eight models that account for R^2, was 0,27. The models with FE have, as expected, the highest explained variation. The constructed models are assumed to explain only around 25% of our observed heterogeneity in the literature on educations effect on income inequality. However, with very low explanatory power in the multilevel models, and knowing that the results from the statistical model evaluation states that the model with fixed effects is not very fitting for the given data, this average explanatory power is to a certain extent misleading. Investigating only the explanatory power of the WLS models, the average R^2 is 23,5. This is lower than expected, as Abdullah et al. in the earlier meta-analysis explain around 35% of the variation in their findings with their models. This MRA is still different from Abdullah et. al.'s analysis, and central variables such as a regional variable for African countries, a variable for political stability and democratic regime have been excluded. A reduced explanatory power of the models presented here, compared to the models presented by Abdullah et al., is not alone evidence that states that variables on regions and regime-characteristics have a strong influence on the findings of educations effect on income inequality. This variation in explanatory power might however indicate that these variables should be considered by other researchers seeking to identify this effect.

Publication bias in the MRA

Stanley and Doucouliagos state that the models with fixed effects and random effects assumptions in the MRA should be included, because of the opportunity to check for publication bias and the existence of a genuine empirical effect by investigating the results across models (Stanley and Doucouliagos 2012, 105). To check for publication bias in a MRA, the coefficient of the standard error should not be significant, as this indicates that only studies with a certain level of standard errors get published. The standard error within the presented eight models not significant at a 95% level in any of the models, indicating that there is no publication bias, also when we include our independent variables in the model. The lack of correlation between the partial correlation and its standard error does not in itself control for publication bias, but together with the earlier findings of our FAT-PET MRA this gives reasonable evidence to conclude that there is no significant publication bias that is necessary to account for in our sample.

6 Discussion

The results from the multiple regression analysis can be used to discuss what the systematic differences in previous research may imply about the research done on this relationship, and the nature of the subject itself. A significant relationship between chosen estimation methods and sources of data implies that researchers should be aware of these sources of heterogeneity when designing and reviewing research on this topic. A significant de-equalizing effect of education on the income share rate of bottom- and middle class must be reviewed, as these findings could have policy implications. Lastly, detecting a stronger equalizing effect of primary education, and a stronger effect of education in more recent studies, are both findings both political and academic interest. These results must further be discussed considering their robustness and their relevance for existing theories on the subject. This is an important step in the search of gaining more knowledge on the reported effects of education on income inequality.

6.1 Heterogeneity and estimation difference

How one control for endogeneity matters, as there was a significant effect of estimation differences between studies that use OLS with fixed *effect* and those who did not, and between studies that uses different models than OLS to control for endogeneity. As we can detect that heterogeneity between studies can be explained partly by these differences, researchers should exercise caution when modelling the effect of education on income inequality. By controlling for endogeneity in different ways within the same study, with models such as GMM, VII, SLS and so on, one might provide more robust results. 23 of the 61 studies in the dataset combine different estimation methods for modelling endogeneity. Among the 20 studies published after 2010, half of them combine models that control for endogeneity in different ways. More recent studies within this sample can therefore be considered to account better for robustness in their findings, by using different estimation methods.

The results of a stronger negative (or smaller positive) effect of education on inequality in studies using OLS with fixed effect, invites to a discussion of the possibility of the causal direction of the effect between education and income inequality. Because WLS models with fixed effects are better suited for eliminating temporal differences, our results indicate that many researchers document an effect of education on income inequality, in the direction proposed by traditional theories. However, these findings do not rule out the possibility of a reverse causation between the two phenomena. Identifying that heterogeneity between the findings on the subject under investigation is partly explained by researchers' choice of using OLS with FE, researchers are advised to consider this source of variation between the findings on educations effect on income inequality, both at the stage of research design, and when evaluating earlier research on this specific topic.

6.2 How should we handle heterogeneity due to database differences?

In other meta-analyses within the social sciences, researchers' databases selection is a known variable for explaining heterogeneity in findings. Within this sample, almost 40% of the articles use Barro as a source for data on education, and almost 50% use data gathered by the World Bank on income inequality. This illustrates that information from a few larger sources influence central research on this topic. The variable World Bank was expected to explain some of the variation in the findings, but this was not the case. A possible explanation to this, is that the World Bank provides many different, and large, sets of data, and that a relationship is therefore less likely to appear. The dataset is also larger and more varied than the Barro data set, which could in part explain why using data from the Barro data set is more likely to affect the results of a study, by comparison. As the Barro dataset is a source of heterogeneity should stress that researchers should strive to vary or combine data on education, to help prevent their field from research being too heavily dominated by results from one source of data. Still, there is no easy solution to this, as some might claim that the Barro database is the most comprehensive, valid and transparent source of data on education – qualities that should not be sacrificed in the process of getting less divergent findings. This consideration must therefore be made by the individual researcher.

6.3 Education and unequal distribution of income within the bottom and middle class

The results of studies using *share of bottom and middle-class* was interpreted as more likely to report a stronger positive, or smaller negative, effect of education on income inequality. These results contrast with the earlier results from Abdullah et. al., who identified education as having the strongest equalizing effect on the top- and bottom tail of income distributions. A plausible explanation for why the income shares of middle- and bottom earners is seen to have a stronger explanatory power in this analysis, is the specific coding of this variable. This variable is quite different from the three variables that it is compared to, involving 80% of the segment of the population. It is therefore logical to observe education creating larger differences in 80% of the population, compared to within the studies that measure the effect on segments of 20% and 40% of the population. The observed differences here are therefore thought to stem from measurement differences and cannot be interpreted to say much about the effect of education on income inequality.

6.4 Does primary schooling have a stronger effect than other educational levels?

The results from this analysis indicate that primary schooling has a more equalizing effect than secondary and tertiary schooling. In the theory chapter, critics of the human capital theory and the Kuznets curve argued that these general theories ignore differentiation between the possible different effects from different types of education, on income inequality (Sylwester 2003, 250; Rodríguez-Pose and Tselios 2009, 360; Gruber and Kosack 2014, 253). It was especially highlighted that one should expect stronger equalizing effects from primary education, compared to tertiary education, which was hypothesized to have de-equalizing effects (Sylwester 2003, 250; Rodríguez-Pose and Tselios 2009, 360; Gruber and Kosack 2014, 253). The findings presented here are in line with this criticism, illustrating that not all types of education can be expected to have the same equalizing effects, and that primary education is the type of education researchers observe have the strongest effect on income inequality. There is no strong and robust positive correlation between secondary and tertiary education and income inequality, leaving the relationship between these types of education and income inequality unclear.

However, these findings cannot be considered to contradict the theories proposed by researchers relying on the Kuznets curve or the human capital hypothesis, as these theories do not oppose the importance of "basic" schooling and its impact on general income distribution. In accordance with both Kuznets and the human capital hypothesis, the importance of primary schooling might be due to the movement of people from rural to urban sectors, made possible with primary education access – and consequently a wage compression effect occurs when the supply of skilled labour increases. In that case, the missing effect of secondary- and tertiary education across studies could imply that the wage compression effect has not been enabled in sectors demanding this expertise, and that a further shift of people to sectors demanding higher educational levels would create this effect further down the road. Or, considering the researchers stressing that one must account for the expected return on education, there might be societies where there are low returns on the financial investment of secondary and tertiary schooling, and higher expected return on completing primary education. The finding of a stronger negative relationship between primary schooling and income inequality is therefore not opposing to many theories on the subject, but highlights the need for nuances between different types of education, when making generalized arguments on this relationship.

In contrast to earlier meta-analyses on this subject, this analysis did not identify an equalizing effect of secondary schooling on income inequality across studies. Abdullah et. al. highlights the effect of this particular type of education in their results. There are various explanations for why this effect is not apparent in the results here. One is that the effect of secondary schooling has decreased over time, as this analysis contains newer studies. For such an increase to follow in light of the reviewed theory, there should also be a decrease in the return on secondary education (changes within the national job market), or a decrease in the access to this type of education (Checchi 2001, 5; Tselios 2008, 409). Other intervening factors, such as institutional factors, population growth or changes in financial growth might help explain this change in observed effects. However, none of these factors explain why the specific effect of secondary education is unobservable in this meta-analysis, and none of these factors are controlled for here. A more plausible explanation for the differences in findings here compared to the earlier meta-analysis, is the differences in study design. As this meta-analysis avoids including studies on a regional level and studies with only one case, this might be the reason for finding different results regarding the effect of secondary education. As described in the theory section, many researchers highlight the possible different effects of education within different parts of a country, due to geographical differences (Edward Glaeser, Matt Resseger, and Tobio 2009, 632; Tselios 2008, 405-406). Some even argue that differences in access within more urban and more rural areas are of such a magnitude, that they should be studied as two separate phenomena (Edward Glaeser, Matt Resseger, and Tobio 2009, 631). Therefore, it might be suspected that secondary education has a stronger effect on education within some areas or regions that are not observable on the national level. No conclusions can be drawn based on the different findings between this meta-analysis and the findings of Abdullah et. al. The divergent results on the effect of secondary schooling in relation to regional differences could be of interest for further research.

6.5 Have we seen a change in the observed effect of education on income inequality in countries over the past 50 years?

There is no significant effect of studies being conducted before 1990 compared to after 1990. 70% of the findings in our sample end after 1990, meaning that we might find a larger variance in findings later in time. Recent studies included in the new meta-set (published after 2010) have a stronger explanatory power and are more likely to provide results about a more negative relationship between education and income inequality than studies published before this. This development is further illustrated by the time series model of the average partial correlation published by year as declining over a period of 50 years. A change in the observed effect of education on income inequality can therefore be theorized.

This change in observed effect of education on income inequality over time could have several explanations. One is that the effect of education on income inequality actually has increased in recent years. If one accepts the premise that many countries might have developed financially in the period from 2010 to 2022, then these findings are in line with both Kuznets curve and the human capital theory, where education increases its importance in determining income distribution within countries over time. A general trend of the labour force moving from more agricultural to more urban sectors over the past 50 years is a plausible explanation. The problem with such broad and general economic and development theories, is however that the mechanisms they describe are so general that more specific findings are unsuitable to test the explanation proposed. To be able to falsify the mechanisms of Kuznets predicted U-curve, one would have to also include data on economic development and the movement of labour force between sectors. Hence, the findings from this analysis do not contradict the traditional theories about educations effect on income inequality on a country level over time, but they cannot be used to verify the causal mechanisms that are hypothesised in these theories. An increase in the effect of education on income inequality can also be interpreted in line with other theories on the subject, for example if there is a simultaneous increase in the return on education, more equal access to education or economic growth. These are all trends than one can expect have developed in different directions in different parts of the world over the past 50 years. Further research on the suspected simultaneous or intervening effects, might shed more light on the possible relationship between education and circumstances such as population growth, institutional factors, economic development and so on.

Another explanation for why there is an observed increasing effect of education on income inequality within this analysis, is that this is more of a regional than a global trend. In the meta-analysis by Abdullah et. al, there is a significant relationship between the effect of education and estimations gathered from African countries. Due to the unclear information about regional differences in many of the recent published studies, regional coding was left out of this analysis. Since the studies are not coded for regional differences, a random sample selection of studies that include estimates largely from African countries for example, could skew the overall picture of the increasing effect of education in a global perspective. This could imply that there is an increase in the effect of education, but only within certain regions, which in this analysis, is not controlled for here.

There are also strictly methodological explanations for why there is an increase in the observed effect of education on income inequality in research over time. One explanation could be that more recent studies contain study characteristics that are more likely to generate negative or smaller positive effect between education and income inequality. As illustrated in this analysis, certain study characteristics correlate with stronger negative effects of education on income inequality. Out of the 20 studies sampled from after 2010, 3 use data from the Barro database, 11 include more than one estimate generated with OLS FE, 17 out of the 20 studies include a different way to deal with endogeneity, either instead of or in accordance with OLS models. The increased use of non-OLS models would then stand out as the most likely study characteristic to have affected the study findings, according to our results.

Another methodological explanation to why there is a statistically significant relationship between the post 2010-dummy and the reported effect of education on income inequality, is a general "decline effect" in the observations due to a general increase of quality in empirical political, sociological, and economic research over the past 10 years. Researchers and meta-analysists within psychology and biology have described the "decline effect" as a less acknowledged time bias where researchers struggle to reproduce past effect sizes (Pietschnig et al. 2019, 2; Clements et al. 2022). These researchers claim that the reproducibility of past studies has been poor, and that researchers attempting to reinvestigate earlier effect sizes systematically face the problem of observing the same, often sensationally high, effect. Contrary to the trend observed in biological and psychological research, a stronger effect size over time is found in this analysis. Suspicion towards earlier research as less reliable and more sensational can therefore not be seen as very relevant here. However, the decline effect bias might be important consider, because of the very nature of meta-analysis. As it is a standard procedure to weigh effect sizes in a multiple MRA, the analysis is constructed to emphasize studies with higher precision. An increase in the precision of empirical research can be expected to have increased over time, as new technology and modern data collection has enabled researchers to conduct larger studies with higher precision, than before. Therefore, when investigating a possible change in the field of research on a specific topic, meta-analysis with precision weighting is inclined to favour estimates from newer studies, compared to older studies. This can in part explain why a statistically significant stronger correlation between education and income inequality is observed in studies published after 2010: The effect sizes from these studies are all weighted heavier than the studies published before this point in time.

While this illustrates one explanation to why there is a statistical difference between the effect sizes in studies published before and after 2010, it fails to explain why one observes a stronger negative effect size over time. The time series plot in this analysis shows a decline in the average reported effect of education on income inequality, and this model is not weighted by study precision or sample size. In line with meta-analysis practice, unweighted results should not inform discussions about the effect between phenomena, but they do illustrate a change in the reported effect sizes over time. Hence, even though weighing procedures might cause portraying a difference between effect sizes reported over time, they do not cause a shift from a smaller positive correlation to a stronger negative correlation. On average, there is a stronger negative effect of education on income inequality in the models, and it is unclear whether this trend is due to actual factors that have been expected in theoretical work, or if this change is due to other circumstances.

7 Concluding remarks and guidance for further research

With a sample of 61 studies published between 1971 and 2022, a mean partial correlation weighted by study size of -0,21 has been identified. This effect is considered moderate, using Doucouliagos' guidelines for interpreting partial correlations in economic studies, and small to medium effect using Cohen's guidelines for interpretation. Nonetheless, this finding is in line with dominating theories on the subject, that claim a negative linear relationship between education and income inequality. This result is used to answer (Q1): the mean reported effect of education on national levels of income inequality today is a moderate to small negative effect.

Using statistical tests and a multiple MRA to investigate the variation of the precision of studies and their reported effect size, there is not found any evidence that indicates that there is significant publication bias within the literature on the relationship between education and income inequality. One should never assume the total absence of publication bias, still the result from this testing is used to answer (Q2): that there is no indication of publication bias present within the literature on education's effect on income inequality.

When investigating the relationship between the characteristics of studies and their reported effect size, it is found that studies published after 2010 report a stronger and negative effect of education on income inequality than studies published before 2010. Other study characteristics that researchers should be aware might influence the observed results on this relationship, is the use of different estimation methods; large income share ratios; measuring education as primary education; and using data from the Barro database. The presented findings answer (Q3): there are systematic differences (heterogeneity) between the studies of educations effect on income inequality. By identifying significant differences in results due to study characteristics, these differences might help explain why researchers have produced divergent findings on this complex relationship over the past 50 years.

The small to moderate effect of education on income inequality observed might still challenge the view of education as "the great equalizer" within policy. The finding of a stronger equalizing effect of primary schooling compared to other types of education, should by of specific interest for policy makers seeking to counteract income inequality with education.

In light of the results derived from this analysis, further research should be concerned with other possible sources of variation amongst findings on this topic. This thesis only describes part of the detected heterogeneity between the findings, which makes it hard to draw clear conclusions about why there is still divergent findings on this subject after over 50 years of empirical research. By strengthening the regional focus and investigate whether there is a stronger relationship between education and income inequality in countries that experience an increase in economic development, one might contribute with more knowledge on the fundamental mechanisms that theory proposes as reasons for why education should matter in determining the distribution of income.

8 Literature

Abdul Abdullah, Hristos Doucouliagos and Eloisabeth Manning. 2015. "Does Education Reduce Income Inequality? A Meta-Regression Analysis." *Journal of Economic Surveys* 29 (2): 301-316. https://doi.org/10.1111/joes.12056.

Abdullah, Abdul, Hristos Doucouliagos, and Elizabeth Manning. 2013. "Does Education Reduce Income Inequality? A Meta-Regression Analysis." *Journal of Economic Surveys* 29 (2): 301-316. https://doi.org/https://doi.org/10.1111/joes.12056. https://onlinelibrary.wiley.com/doi/abs/10.1111/joes.12056.

Ahluwalia, Montek S. 1974. "The Scope of Policy intervention." In *Redistribution with Growth*, edited by Montek S. Ahluwalia Hollis Chenery, C.L.G Bell, john H. Duloy and RIchard Jolly, 73-89. New York: Oxford University Press.

---. 1976a. "Inequality, Poverty and Development." *Journal of Development Economics* 3: 307-342.

---. 1976b. "Income Distribution and Development: Some Stylized Facts." *The American Economic REview* 66 (2): 128-135. https://www.jstor.org/stable/1817209.

Ahmadov, Anar, K. 2013. "Oil, Democracy and context: A Meta-Analysis." *Comparative Political Studies* 47 (9): 1238-1267. https://doi.org/10.1177/001041403495358.

Alderson, Fancois Nielsen and Arthur S. 1995. "Income Inequality, Development, and Dualism: Results from an Unbalanced Cross-National Panel." *American Sociological Review* 60 (5): 674-701. https://doi.org/https://doi.org/10.2307/2096317.

Alinaghi, Nazila, and W. Robert Reed. 2018. "Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work?" *Research Synthesis Methods* 9 (2): 285-311. https://doi.org/https://doi.org/10.1002/jrsm.1298. https://doi.org/10.1002/jrsm.1298.

Barro, Robert J. 2000. "Inequality and Growth in a Panel of Countries." *Journal of Economic Growth* 5: 5-32. https://www.jstor.org/stable/40216021.

Beck, Thorsten , and Ross Levine Asli Demirgüç-Kunt. 2007. "Finance, inequality and the poor." *J Econ Growth* 12: 27-49. https://doi.org/DOI 10.1007/s10887-007-9010-6.

Bernasek, Anna. 2005. "What's the Return on Education?" *The New York Times*, 2005, 3. https://www.nytimes.com/2005/12/11/business/yourmoney/whats-the-return-on-education.html.

Bojer, Hilde. 2021. "Inntekt." Makroøkonomi. Store Norske Leksikon. Accessed 03.02.2022. https://snl.no/inntekt#:~:text=Inntekten%20kan%20defineres%20i%20forhold,lengde%20p%C3%A5%20sin%20yrkesaktive%20tid.

Bosker, Tom A B Snijders and Roel J. 2012. *Multilevel Analysis*. London: Sage Publications Inc.

Bourguingnon, F. , and C. Morrisson. 1990. "Income Distribution, Development and Foreign Trade." *European Economic Review* 34: 1113-1132.

Breen, Richard, and Cecilia García-Penalosa. 2005. "Income Inequality and Macroeconomic Volatility: An Empirical Investigation." *Review of Development Economics* 9 (3): 380-398. https://doi.org/https://doi.org/10.1111/j.1467-9361.2005.00283.x.

Broderstad, Troy. 2015-06-02. "A meta-analysis of income and democracy. Why we still have not reached consensus after 60 years of research." MA Master thesis, Department of Comparative Politics, University of Bergen.

Brown, Phillip, and and Stuart Tannock. 2009. "Education, meritocracy and the global war of talent." *Journal of Education Policy* 24 (4): 377-392. https://doi.org/https://doi.org/10.1080/02680930802669938.

Cancela, Joao and Benny Geys. 2016. "Explaining voter turnout: A meta-analysis of national and subnational elections." *Electoral Studies* 42: 264-275. https://doi.org/https://doi.org/10.1016/j.electstud.2016.03.005.

Card, Noel A. . 2011. *Applied Meta-Analysis for Social Science Research*. Edited by Todd D. Little.*Methodology in the Social Sciences*. New York: The Guilford Press.

Carvajal, M.J , and David T. Geithman. 1978. "Income Distribution adn Economic Development: Some Intra-Country Evidence." *Southern Economic Journal* 44 (4): 922-928. https://www.jstor.org/stable/1057738.

Chalmers, Iain, Larry V. Hedges, and Harris Cooper. 2002. "A Brief History of Research Synthesis." *Evaluation & the Health Professions* 25 (1): 12-37. https://doi.org/10.1177/0163278702025001003. https://doi.org/10.1177/0163278702025001003.

Checchi, Daniele. 2001. Education, Inequality and Income Inequality. In *Centres for Economics and Related Disciplines, London School of Economics*, edited by Distributional Analysis Research Programme. London.

Chiswick, Barry R. 1968. "The Average LEvel of Schooling and the Intra-Regional Inequality of Income: A Clarification." *The American Economic REview* 58, no. 1 (3): 495-500. http://www.jstor.org/stable/1813775.

---. 1971. "Earnings inequality and Economic Development." *The Quarterly Journal of Economics* 85 (1): 21-39. https://www.jstor.org/stable/1881838.

Clements, Jeff C., Josefin Sundin, Timothy D. Clark, and Fredrik Jutfelt. 2022. "Meta-analysis reveals an extreme “decline effect” in the impacts of ocean acidification on fish behavior." *PLOS Biology* 20 (2): e3001511. https://doi.org/10.1371/journal.pbio.3001511. https://doi.org/10.1371/journal.pbio.3001511.

Cohen, Jacob. 1992. "Statistical power analysis." *Current directions in psychological science* 1 (3): 98-101.

Cowell, D. G. Champernowne and F. A. 1998. *Economic inequality and income distribution*. Cambridge: Cambridge University Press.

Croix, John Owens and Tania de St. 2020. "Engines of Social Mobility? Navigating Meritocratic Education Discourse in an Unequal Society." *British Journal of Educational Studies* 68 (4): 403-424. https://doi.org/https://doi.org/10.1080/00071005.2019.1708863.

Dacombe, Rod. 2018. "Systematic Reviews in Political Science: What Can the Approach Contribute to Political Research?" *Political Studies Review* 16 (2): 148-157. https://doi.org/10.1177/1478929916680641. https://journals.sagepub.com/doi/abs/10.1177/1478929916680641.

De Gregorio, José , and and Jong-Wha Lee. 2002. "Education and Income Inequality: New Evidence From Cross-Country Data." *Review of Income and Wealth* 48 (3): 395-416. https://doi.org/https://doi.org/10.1111/1475-4991.00060.

Deaton, Angus, and & Salman Zaidi. 2002. Guidelines for Constructing Consumption Aggregates Welfare Analysis. LSMS Working Paper 135. The World Bank, Washington DC.

Demena, Binyam Afewerk, and Sylvanus Kwaku Afesorgbor. 2020. "The effect of FDI on environmental emissions: Evidence from a meta-analysis." *Energy Policy* 138: 111192. https://doi.org/https://doi.org/10.1016/j.enpol.2019.111192. https://www.sciencedirect.com/science/article/pii/S0301421519307773.

Denyer, D. & Tranfield, D. 2009. "Producing a systematic review." In *The Sage handbook of organizational research methods*, edited by D.A. Buchanan & A. Bryman, 671-689. London: Sage Publications Ltd. .

Dewey, John. 1944. *Democracy and education*. Edited by Collier-Macmillan. New York: The Free Press.

Digdowiseiso, Kumba. 2009. "Education inequality, economic growth and income inequality: Evidence from Indonesia, 1996-2005." *Munich Personal RePEc Archive*. http://mpra.ub.uni-muenchen.de/17792/.

Doucouliagos, Hristos. 2011. How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics. Economic Series. Australia: Faculty of Business and Law, School of

Accounting, Economics and Finance, Deakin University.

Doucouliagos, Hristos, and Martin Paldam. 2008. "Aid effectiveness on growth: A meta study." *European Journal of Political Economy* 24 (1): 1-24. https://doi.org/https://doi.org/10.1016/j.ejpoleco.2007.06.002. https://www.sciencedirect.com/science/article/pii/S0176268007000614.

Doucouliagos, Hristos, T. D. Stanley, and W. Kip Viscusi. 2014. "Publication selection and the income elasticity of the value of a statistical life." *Journal of Health Economics* 33: 67-75. https://doi.org/https://doi.org/10.1016/j.jhealeco.2013.10.010. https://www.sciencedirect.com/science/article/pii/S0167629613001422.

Doucouliagos, Hristos, and Mehmet Ali Ulubaşoğlu. 2008. Democracy and Economic Growth: A Meta-Analysis. Detroit, MI :.

Frey, Bruce. B. . 2018. Literature Review. In *The Sage Enclopedia of Educational Research, Measurement, and Evaluation*, edited by SAGE Research Methods.

Gerring, John. . 2012. *Social Science Methodology - A Unified Framework*. Second ed. Cambridge: Cambridge University Press.

Glaeser, E., and Giacomo A. M. Ponzetto and Andrei Shleifer. 2007. "Why does democracy need education?" *J Econ Growth*, no. 12 (77): 77-99. https://doi.org/10.1007/s10887-007-9015-1.

Glaeser, Edward, Matt Resseger, and and Kristina Tobio. 2009. "Inequality in Cities." *Journal of Regional Studies* 49 (4): 617-646. https://doi.org/https://doi.org/10.1111/j.1467-9787.2009.00627.x.

Glomm, Gerhard, and B. Ravikumar. 1992. "Public versus Private Investment in Human Capital: Endogenous Growth and Income Inequality." *Journal of Political Economy* 100 (4): 818-834. http://www.jstor.org/stable/2138689.

Gruber, Lloyd, and Stephen Kosack. 2014. "The Tertiary Tilt: Education and Inequality in the Developing World." *World Development* 54: 253-272. https://doi.org/https://doi.org/10.1016/j.worlddev.2013.08.002. https://www.sciencedirect.com/science/article/pii/S0305750X13001848.

Grønmo, Sigmund. . 2004. *Samfunnsvitenskapelige metoder*. 2. ed. Bergen: Fagbokforlaget Vigmostad & Bjørke AS.

Hammersley, Martyn. 2001. "On «Systematic» Reviews of Research Literatures: A «narrative» response to Evans & Benefield." *British Educational Research Journal* 27 (5): 543-554. https://doi.org/10.1080/01411920120095726.

Hattie, J. A., and B. C. Hansford. 1982. "Self Measures and Achievement: Comparing a Traditional Review of Literature with a Meta-Analysis." *Australian Journal of Education* 26 (1): 71-75. https://doi.org/10.1177/000494418202600105. https://doi.org/10.1177/000494418202600105.

Havranek, Tomas. 2019. "Revision of Reporting Guidelines." Meta-Analysis in Economics Research. Accessed 13.05.2022. https://www.maer-net.org/post/revision-of-reporting-guidelines.

Heins, D.J. Aigner and A.J. 1967. "On the determinants of Income Inequality." *The American Economic Review* 57 (1): 175-184. https://www.jstor.org/stable/1815614.

Holasut, Thitithep Sitthiyot and Kanyarat. 2020. "A simple method for measuring inequality." *Nature* 6 (112): 1-9. https://doi.org/https://doi.org/10.1057/s41599-020-0484-6.

Hunter, Frank L. Schmidt & John E. 2015. *Intergrating Research Findings across Studies*. Edited by SAGE Publications.*Methods of Meta-Analysis: Correcting Error and bias in Research Findings*.

Husnain, Muhammad Iftikhar Ul, Azad Haider, and Muhammad Aamir Khan. 2021. "Does the environmental Kuznets curve reliably explain a developmental issue?" *Environmental Science and Pollution Research* 28 (9): 11469-11485. https://doi.org/10.1007/s11356-020-11402-x. https://doi.org/10.1007/s11356-020-11402-x.

Ismail, Rahmah. 2000. "Educational and Income Ineuqality in Malaysia." *Humanomics* 16 (2): 19-40. https://doi.org/https://doi.org/10.1108/eb018851.

Jha, Sailesh K. 1996. "The Kuznets Curve: A Reassessment." *World Development* 24 (4): 773-780. https://doi.org/https://doi.org/10.1016/0305-750X(95)00165-9.

Kallager, Per Kristian Roko. 2014. "Globalisation and the Welfare State - A Meta-regression Analysis." MA Thesis, Department of Comparing Politics, University of Bergen. https://bora.uib.no/bora-xmlui/bitstream/handle/1956/8270/121808400.pdf?sequence=1&isAllowed=y.

Keller, Katarina R. I. 2010. "How Can Education Policy Improve Income Distribution? An Empirical Analysis of Education Stages and Measures on Income Inequality." *The Journal of Developing Areas* 43 (2): 51-77. https://www.jstor.org/stable/40376250.

Kellstedt, Paul M., and and Guy Whitten. 2019. *The Fundamentals of Political Science Research*. 3 ed. Cambridge: Cambridge University Press.
Knutsen, Jonathon W. Moses and Torbjørn L. 2019. *Ways of Knowing*. London: Red Globe Press.
Krugman, Paul. 2014. *Why we're in a new gilded age*. Edited by B. DeLong and M. Steinbaum.*After Piketty: The Agenda for Economics and Inequality*. Cambridge: Cambridge University Press.
Kuznet, Simon. 1955. "Growth and Income Inequality." *The American Economic Review* 45 (1): 1-28. https://www.jstor.org/stable/1811581.
---. 1963. "Quantitative Aspects of Economic Growth: VII. Distribution of Income by Size." *Economic Development and Cultural Change* 11, no. 2 (2): 1-80. https://doi.org/https://doi.org/10.1086/450006.
L'evy, G'erard Dumenil and Dominique. 2004. Trends in Capital Ownership and Income. Pension Fund Capitalism and the Crisis of Old-Age Security in the United States». http://www.cepremap.ens.fr/levy/, New School for Social Research, New Left Review.
Lauridsen, Jørgen, and Jesùs Mur. 2006. "Multicollinearity in cross-sectional regressions." *Journal of Geographical Systems* 8 (4): 317-333. https://doi.org/10.1007/s10109-006-0031-z. https://doi.org/10.1007/s10109-006-0031-z.
Lecallion, Jacques, and Christian Morrisson and Dimitri Germidis Felix Paukert. 1984. *Income Distribution and Economic Development: An Analystical Survey*. Geneva, Switzerland: International Labour Office.
Lee, and & Lee. 2016. The enrollment ratios for the different education levels
by gender from 1820 and 2010 for individual countries. edited by Barro and Lee Dataset. http://www.barrolee.com/.
Lee, William, and Matthew Hotopf. 2012. "10 - Critical appraisal: Reviewing scientific evidence and reading academic papers." In *Core Psychiatry (Third Edition)*, edited by Pádraig Wright, Julian Stern and Michael Phelan, 131-142. Oxford: W.B. Saunders.
Leon, Valerie Koechlin & Gienmarco. 2007. "International Remittances and Income Inequality: An Empirical Investigation." *Journal of Economic Policy Reform* 10 (2): 123-141. https://doi.org/10.1080/17487870701346514.
Leszczensky, Lars. 2019. "How to Deal With Reverse Causality Using Panel Data? Recommendations from Researchers Based on a Simulation Study." *Sociological Methods & Research* 51 (2): 837-865. https://doi.org/https://doi.org/10.1177/0049124119882473.
Li, Hongyi, Lyn Squire, and Heng-fu Zou. 1998. "Explaining International and Intertemporal Variations in Income Inequality." *The Economic Journal* 108 (446): 26-43. https://doi.org/https://doi.org/10.1111/1468-0297.00271. https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0297.00271.
Ludvigsen, Mette S. and Elisabeth O.C Hall, Gabriele Meyer, Liv Fegran, Hanne Aagaard and Lisbeth Uhrenfeldt. 2015. "Using Sandelowski and Barroso's Meta-Synthesis Method in Advancing Qualitative Evidence." *Qualitative Health Research* 26 (3): 320-329. https://doi.org/10.1177/1049732315576493.

Mandon, Pierre, and Antonie Cazals. 2019. "Political Budget Cycles: Manipulation by Leaders Versus Manipulation by Researchers? Evidence from a Meta-Regression Analysis." *Journal of Economic Surveys* 33 (1): 274-308.

Martino, Cristiano Perugini and Gaetano. 2008. "Income Inequality within European Regions: Determinants and Effects on Growth." *Review of Income and Wealth* 54 (3): 373-406. https://doi.org/https://doi.org/10.1111/j.1475-4991.2008.00280.x.

Mechling, George , Stephen Miller, and and Ron Konecny. 2017. "Do Piketty and Saez Misstate Income Inequality? Critiquing the Critiques." *Review of Political Economy* 29 (1): 30-46. https://doi.org/https://doi.org/10.1080/09538259.2017.1255439.

Micheal, Vincent Abiodun, and Alfred Adewole Abiodun. 2014. "Estimation of regression coefficients in the presence of multicollinearity." *Social and Basic Sciences Research Review* 2 (10): 404-415.

Mill, John Stuart. 1963-1991. *Collected works of John Stuart Mill*. Edited by John M. Robson. Vol. 33. Toronto, London: University of Toronto Press, Routledge.

---. 2002. *Considerations on Representative Government*. Edited by John Gray.*On Liberty and Other Essays*. New York: Oxford University Press.

Motonishi, Taizo. 2004. "Why has income ineuqliaty in Thailand increased? An analysis using surveys from 1975 to 1998." *Japan and the World Economy* 18 (4): 464-487. https://doi.org/https://doi.org/10.1016/j.japwor.2004.11.003.

Nakosteen, Mehdi K. 2021. "Education." Lifestyles and Social Issues. Encyclopedia Britannica. Accessed 12. january. https://www.britannica.com/topic/education.

Neckerman, Kathryn M., and Florencia Torche. 2007. "Inequality: Causes and Consequences." *Annual Review of Sociology* 33 (1): 335-357. https://doi.org/10.1146/annurev.soc.33.040406.131755. https://doi.org/10.1146/annurev.soc.33.040406.131755.

Nikos Benos, Stefania Zotou. 2014. "Education and Economic Growth: A Meta Regression Analysis." *World development* 64: 669-689. https://doi.org/http://dx.doi.org/10.1016/j.worlddev.2014.06.034.

Nord, Stephen. 1980a. "Intre-Metropolitan Income Inequality." *Eastern Economic Journal* 6 (3/4): 195-202. https://www.jstor.org/stable/40324756.

---. 1980b. "Income Inequality and City Size: An Examination of Alternative Hypotheses for Large and Small Cities." *The Review of Economic and Statistics* 62 (4): 502-508. https://www.jstor.org/stable/1924774.

---. 1980c. "An Empirical Analysis of Income Inequality." *Southern Economic Journal* 46 (3): 863-872. https://www.jstor.org/stable/1057154.

O'Connor, Annette, and Jan Sargeant. 2015. "Research synthesis in veterinary science: Narrative reviews, systematic reviews and meta-analysis." *The Veterinary Journal* 206 (3): 261-267. https://doi.org/https://doi.org/10.1016/j.tvjl.2015.08.025. https://www.sciencedirect.com/science/article/pii/S1090023315003500.

O'Neill, Martin. 2017. "Survey Article: Philosophy and Public Policy after Piketty." *Journal of Political Philosophy* 25 (3): 343-375. https://doi.org/https://doi.org/10.1111/jopp.12129. https://doi.org/10.1111/jopp.12129.

Ortiz-Ospina, Max Roser and Esteban. 2016. "The world is more educated than ever before." Global Education. Our World in Data. Accessed 02.06.2022. https://ourworldindata.org/global-education#citation.
Park, Kang H. 1996. "Educational Expansion and Educational Inequality on Income Distribution." *Economics of Education Review* 15 (1): 51-58. https://doi.org/https://doi.org/10.1016/0272-7757(95)00000-3.
---. 2010. "Distribution and growth: cross-country evidence." *Applied Economics* 30 (7): 943-949. https://doi.org/https://doi.org/10.1080/000368498325372.
Partridge, James S., Mark D. Partridge, and Dan S. Rickman. 2007. "State Patterns In Family Income Inequality." *Contemporary Economic Policy* XVI: 277-295. https://doi.org/https://doi.org/10.1111/j.1465-7287.1998.tb00519.x.
Pietschnig, Jakob, Magdalena Siegel, Junia Sophia Nur Eder, and Georg Gittler. 2019. "Effect Declines Are Systematic, Strong, and Ubiquitous: A Meta-Meta-Analysis of the Decline Effect in Intelligence Research." *Frontiers in Psychology* 10. https://doi.org/10.3389/fpsyg.2019.02874. https://www.frontiersin.org/article/10.3389/fpsyg.2019.02874.
Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge: The Belknap Press of Harvard University Press.
Psacharopoulos, George. 1977. "Unequal Access to Education and Income Distribution - An International Comparison." *De Economist* 125 (3): 383-392. https://doi.org/https://doi.org/10.1007/BF01479118.
Ragin, Charles. 1987. *The Comparative MEthod: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California.
Ram, Rati. 1981. "Inequalities in Income and Schooling: A Different Point of View." *De Economist* 129 (2): 253-261. https://doi.org/https://doi.org/10.1007/BF01705811.
---. 1984. "Population increase, Economic Growth, Educational Inequality, and Income Distribution." *Journal of Development Economics* 14: 419-428. https://doi.org/https://doi.org/10.1016/0304-3878(84)90069-5.
Rodgers, Gerry. 1983. "Population growth, inequality and poverty." *International Labour Review* 122 (4): 443-460.
Rodríguez-Pose, Andrés , and and Vassilis Tselios. 2009. "Education and Income Inequality in the Regions of the European Union." *Journal of Regional Science* 49 (3): 411-437. https://doi.org/https://doi.org/10.1111/j.1467-9787.2008.00602.x.
Rohlfing, Ingo. 2012. "Varieties of process tracing and ways to answer why-questions»." *European Political Science* 12: 31-39.
Rosenberger, Randall and T.D Stanley. 2006.
Round, M.O. Odedokun and Jeffery I. 2004. "Determinants of Income Inequality and its Effects on Economic Growth: Evidence from African Countries." *African Development Review* 16 (2). https://doi.org/https://doi.org/10.1111/j.1017-6772.2004.00093.x.
Sadoulet, Alain de Janvry and Elisabeth. 2005. "Growth, Poverty and Inequality in Latin-America: A Causal Analysis." *Review of INcome and Wealth* 46 (3): 267-287. https://doi.org/https://doi.org/10.1111/j.1475-4991.2000.tb00843.x.

Salverda, Wiemer , and Brian Nolan and Timothy M. Smeeding, eds. 2009. *The Oxford Handbook of Economic Inequality*. Oxford: Oxford University Press.
Sandelowski, Margarete and Julie Barroso. 2007. *Handbook for Synthesizing Qualitative Research*. New York: Springer Publishing Company Inc.
Sebri, Maamar, and Hajer Dachraoui. 2021. "Natural resources and income inequality: A meta-analytic review." *Resources Policy* 74: 102315. https://doi.org/https://doi.org/10.1016/j.resourpol.2021.102315. https://www.sciencedirect.com/science/article/pii/S0301420721003251.
Shorrocks, Anthony. 2005. Inequality valoues and unequal shares. Jubilee COnference: WIDER thinking ahead: The future of development economics. UNU-WIDER.
Silva, Julie A. 2007. "Trade and Income Inequality in a Less Developed Country: The Case of Mozambique." *Economic Geography* 83 (2): 111-136. https://doi.org/https://doi.org/10.1111/j.1944-8287.2007.tb00339.x.
Singh, Gir S. Gupta and Ram D. 1984. "Income Inequality across Nations over Time: How Much and Why." *Southern Economic Journal* 51 (1): 250-257. https://doi.org/https://doi.org/10.2307/1058336.
Solga, Heike. 2014. "Education, economic inequality and the promises of the social investment state." *Socio-Economic Review* 12 (2): 269-297. https://doi.org/10.1093/ser/mwu014. https://doi.org/10.1093/ser/mwu014.
Squire, Mattias Lundberg and Lyn. 2003. "The Simultaneous Evolution of Growth and Inequality." *The Economic Journal* 113 (487): 326-344. https://www.jstor.org/stable/3590323.
Stanley, and Hristos Doucouliagos. 2012. *Meta-Regression Analysis in Economics and Business*. 1 ed. Vol. 5.*Routledge advances in research methods*. Florence: Florence: Routledge.
Stanley, T. D. 2008. "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection*." *Oxford Bulletin of Economics and Statistics* 70 (1): 103-127. https://doi.org/https://doi.org/10.1111/j.1468-0084.2007.00487.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0084.2007.00487.x.
---. 2017. "Limitations of PET-PEESE and Other Meta-Analysis Methods." *Social Psychological and Personality Science* 8 (5): 581-591. https://doi.org/10.1177/1948550617693062. https://doi.org/10.1177/1948550617693062.
Stanley, T. D., and Hristos Doucouliagos. 2017. "Neither fixed nor random: weighted least squares meta-regression." *Research Synthesis Methods* 8 (1): 19-42. https://doi.org/https://doi.org/10.1002/jrsm.1211. https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1211.
Stanley, T.D , and and Stephen B. Jarell. 1989. "Meta-Regression Analyses: A Quantitative Method of Literature Surveys " *Journal of Economic Surveys* 19 (3): 299-308. https://doi.org/https://doi.org/10.1111/j.0950-0804.2005.00249.x.
Stanley, T.D., Hristos Doucouliagos, Margaret Giles, Jost H. Heckemeyer, Robert J. Johnston, Patrice Laroche, Jon P. Nelson, Martin Paldam, Jacques Poot, Geoff Pugh, Randall S.

Rosenberger, and Katja Rost. 2013. "META-ANALYSIS OF ECONOMICS RESEARCH REPORTING GUIDELINES." *Journal of Economic Surveys* 27 (2): 390-394. https://doi.org/https://doi.org/10.1111/joes.12008. https://onlinelibrary.wiley.com/doi/abs/10.1111/joes.12008.

Sterne, Jonathan A. C., and Matthias Egger. 2001. "Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis." *Journal of Clinical Epidemiology* 54 (10): 1046-1055. https://doi.org/https://doi.org/10.1016/S0895-4356(01)00377-8. https://www.sciencedirect.com/science/article/pii/S0895435601003778.

Sterne, Jonathan A. C., Alex J. Sutton, John P. A. Ioannidis, Norma Terrin, David R. Jones, Joseph Lau, James Carpenter, Gerta Rücker, Roger M. Harbord, Christopher H. Schmid, Jennifer Tetzlaff, Jonathan J. Deeks, Jaime Peters, Petra Macaskill, Guido Schwarzer, Sue Duval, Douglas G. Altman, David Moher, and Julian P. T. Higgins. 2011. "Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials." *BMJ* 343: d4002. https://doi.org/10.1136/bmj.d4002. http://www.bmj.com/content/343/bmj.d4002.abstract.

Sylwester, Kevin. 2002. "Can education expenditures reduce income inequality." *Economics of Education Review* 21: 43-52. https://doi.org/https://doi.org/10.1016/S0272-7757(00)00038-8.

---. 2003. "Enrolment in Higher Education and Changes in Income Inequality." *Bulletin of Economic Research* 55 (3): 0307-3378. https://doi.org/https://doi.org/10.1111/1467-8586.00173.

Terrin, Norma, Christopher H. Schmid, and Joseph Lau. 2005. "In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias." *Journal of Clinical Epidemiology* 58 (9): 894-901. https://doi.org/https://doi.org/10.1016/j.jclinepi.2005.01.006. https://www.sciencedirect.com/science/article/pii/S089543560500082X.

Tong, Guangyu and Guang Guo. 2019. "Meta-analysis in Sociological Research: Power and Heterogenity." *Sociological Methods & Research* 11 (20): 1-39.

Trostel, Philip A. 2005. "Nonlinearity in the Return to Education." *Journal of Applied Economics* VIII (1): 191-202. https://doi.org/https://doi.org/10.1080/15140326.2005.12040624.

Tsai, Pan-Long. 1995. "Foreign Direct Investment and Income Inequality: Further Evidence." *World Development* 23 (3): 469-483. https://doi.org/https://doi.org/10.1016/0305-750X(95)00136-Z.

Tsakloglou, Panos. 2006. "Development and inequality revisited." *Applied Economics* 20 (4): 509-531. https://doi.org/https://doi.org/10.1080/00036848800000061.

Tselios, Vassilis. 2008. "Income and educational inequalities in the regions of the European Union: Gegraphical spillovers under welfare state restrictions." *Regional Science* 87 (3): 403-430. https://doi.org/https://doi.org/10.1111/j.1435-5957.2008.00191.x.

---. 2009. "Growth and Convergence in Income Per Capitaand Income Inequality in the Regions of the EU." *Spatial Economic Analysis* 4 (3): 343-370. https://doi.org/10.1080/17421770903114711.

Tweedie, R.L. 2001. "Meta-analysis: Overview " In *International Encyclopedia of the Social and Behavioral Sciences*, edited by Neil J. Smelser, and Paul B. Baltes., 9717-9724. Amsterdam: Elisevier.

UNU-WIDER. 2021. World Income Inequality Dataset - Users guide and data sources. 1-50.

Verdier, Gilles Saint-Paul and Thierry. 1993. "Education, Democracy and Growth." *Journal of Development Economics*: 399-407.

Wallace, Alice, Karen Croucher, Deborah Quilgars and Sally Baldwin. 2004. "Meeting the challenge: developing systemativ reviewing in social policy." *The Policy Press* 32 (4): 455-470. https://doi.org/10.1332/0305573042009444.

Walsh, Denis, and Soo Downe. 2005. "Meta-synthesis method for qualitative research: a literature review." *Journal of Advanced Nursing* 50 (2): 204-211. https://doi.org/https://doi.org/10.1111/j.1365-2648.2005.03380.x. https://doi.org/10.1111/j.1365-2648.2005.03380.x.

Weinberg, Timothy M. Smeeding and Daniel H. 2001. "Toward a Uniform Definition of Household Income." *Review of Income and Wealth* 47 (1): 1-24. https://doi.org/ https://doi.org/10.1111/1475-4991.00001.

Williamson, Matthew Higgins and Jeffery G. 1999. "Explaining Inequality the World Round: Cohort Size, Kuznets Curves, and Openness." *FRB if New York Staff Report* 79: 1-56. https://doi.org/http://dx.doi.org/10.2139/ssrn.173408.

Winegarden, C. R. 1979. "Schooling and Income Distribution: Evidence from International Data." *Economica* 46: 83-87. https://doi.org/https://doi.org/10.2307/2553099.

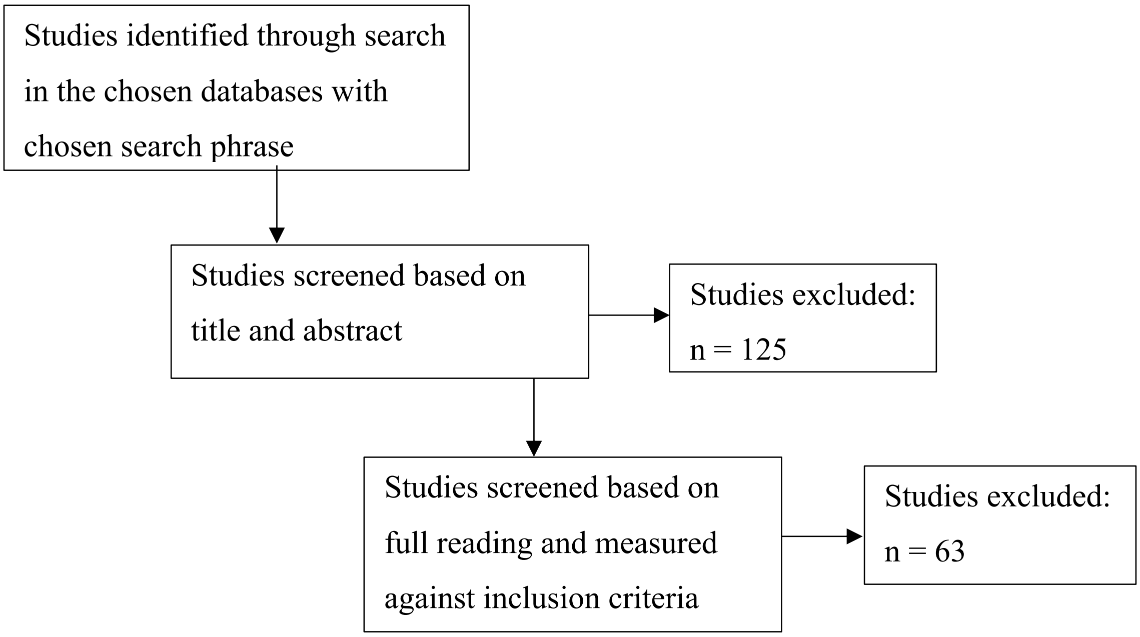
Yesilyurt, Filiz Yesilyurt and M Ensar. 2019. "Meta-Analysis, military expenditures and growth." *Peace Research* 56 (3): 352-363. https://doi.org/https://doi.org/10.1177/0022343318808841.

Zhang, Jie. 1996. "Optimal Public Investments in Education and Endogenous Growth." *The Scandinavian Journal of Economics* 98 (3): 387-404. https://doi.org/10.2307/3440733. http://www.jstor.org/stable/3440733.

Zou, Lixin Colin Xu and Heng-fu. 2000. "Explaining the changes of income distribution in CHina." *China Economic Review* 11: 149-170. https://doi.org/https://doi.org/10.1016/S1043-951X(00)00015-8.

Appendix I.

8.1 Flow chart of selection process



8.2 List of included studies using this selection process

The Impact of Economic Sanctions on Income Inequality of Target States

S. K. Afesorgbor and R. Mahadevan (2016)

Do liberalization and globalization increase income inequality?

T. N. Andreas Bergh (2010)

Is Corruption Really Bad for Inequality? Evidence from Latin America

D. C. R.-D. Antonio R. Andres (2011)

Globalization and income inequality: A panel data econometric approach for the EU27 countries

S. D. Dimitrios Asteriou, Argiro Moudatsou (2014)

The impact of redistributive policies on inequality in OECD countries

P. Doerrenberg and A. Peichl (2014)

The Tertiary Tilt: Education and Inequality in the Developing World

L. Gruber and S. Kosack (2014)

Linking Economic Complexity, Institutions, and Income Inequality

D. Hartmann, M. R. Guevara, C. Jara-Figueroa, M. Aristarán and C. A. Hidalgo (2017)

Does microfinance affect income inequality?

N. Hermes (2014)

Inward and outward FDI and income inequality: evidence from Europe

D. Herzer and P. Nunnenkamp (2013)
Rising Income Inequality: Technology, or Trade and Financial Globalization?
F. Jaumotte, S. Lall and C. Papageorgiou (2013)
The Impact of Tax and Expenditure Policies on Income Distribution: Evidence from a Large Panel of Countries
V. V. Jorge Martínez-Vázquez, Blanca Moreno-Dodson (2012)
How does economic complexity influence income inequality? New evidence from international data
D.P. H. Lan Khanh Chu (2020)
The Kuznets curve for export diversification and income inequality: Evidence from a global sample
T.-H. Le, C. P. Nguyen, T. D. Su and B. Tran-Nam (2020)
Economic complexity, human capital and income inequality: a cross-country analysis
K.-K. Lee and T. V. Vu (2020)
Financialization and the Labor Share of Income
O. Özdemir (2019)
Financial Inclusion, Poverty and Income Inequality
C.-Y. Park, Rogelio Mercado (2018)
Government Social Expenditure and Income Inequalities in the European Union
Á. S. a. A. L. Pérez-Corral (2018)
The Effect of Knowledge Economy Factors on Income Inequality in the Selected Islamic Countries
A. Shahabadi, M. Nemati and S. E. Hosseinidoust (2017)
Why is Corruption Less Harmful to Income Incequality in Latin America?
C. R.-D. Stephen Dobson (2012)
Foreign direct investment and income inequality: Does the relationship vary with absorptive capacity?
J.-Y. Wu and C.-C. Hsu (2012)

Appendix II

9.1 FAT-PET MRA Full models

9.1.1 FAT Robust

Linear regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpa r~l	.459	1.274	0.36	.719	-2.041	2.959	
Constant	-.001	.062	-0.02	.985	-.122	.12	

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.001	Number of obs	916
F-test	0.130	Prob > F	0.719
Akaike crit. (AIC)	-132.802	Bayesian crit. (BIC)	-123.162

**** p<.01, ** p<.05, * p<.1*

9.1.2 FAT-Cluster robust

Linear regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpa r~l	.459	.664	0.69	.492	-.869	1.788	
Constant	-.001	.035	-0.03	.974	-.071	.068	

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.001	Number of obs	916
F-test	0.478	Prob > F	0.492
Akaike crit. (AIC)	-132.802	Bayesian crit. (BIC)	-123.162

**** p<.01, ** p<.05, * p<.1*

9.1.3 PET-Robust

Robust regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Precision_se	0	0	-0.86	.387	0	0	
Constant	-.025	.013	-1.92	.055	-.051	.001	*

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.001	Number of obs	916
F-test	0.748	Prob > F	0.387

**** p<.01, ** p<.05, * p<.1*

9.1.4 PET Cluster robust

Regression results

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Precision_se	0	0	-0.37	.709	0	0	
Constant	.006	.033	0.19	.852	-.059	.071	

Mean dependent var	-0.044	SD dependent var	0.378
Overall r-squared	0.001	Number of obs.	916
Chi-square	0.140	Prob > chi2	0.709
R-squared within	0.000	R-squared between	0.004

**** p<.01, ** p<.05, * p<.1*

9.1.5 FAT Fixed effects clustered

Regression results

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpar~l	.104	.192	0.54	.588	-.273	.481	
Constant	-.047	.01	-4.85	0	-.066	-.028	***

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.000	Number of obs	916
F-test	0.293	Prob > F	1.000
Akaike crit. (AIC)	-36.391	Bayesian crit. (BIC)	-26.750

**** p<.01, ** p<.05, * p<.1*

9.1.6 PET Fixed effects clustered

Regression results

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Precision_se	0	0	-0.25	.8	0	0	
Constant	-.043	.01	-4.38	0	-.062	-.024	***

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.000	Number of obs	916
F-test	0.065	Prob > F	1.000
Akaike crit. (AIC)	-36.145	Bayesian crit. (BIC)	-26.505

**** p<.01, ** p<.05, * p<.1*

9.1.7 FAT MEML clustered

Mixed-effects ML regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpar~l	.15	.189	0.79	.428	-.22	.52	
Constant	0	.034	0.01	.995	-.066	.066	
Constant	-1.421	.103	.b	.b	.	.	
Constant	-1.407	.024	.b	.b	.	.	

Mean dependent var	-0.044	SD dependent var	0.378
Number of obs	916	Chi-square	0.630
Prob > chi2	0.428	Akaike crit. (AIC)	174.220

**** p<.01, ** p<.05, * p<.1*

9.1.8 PET MEML clustered

Mixed-effects REML regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Precision_se	0	0	-0.41	.68	0	0	
Constant	.006	.034	0.18	.856	-.06	.073	
Constant	-1.408	.104	.b	.b	.	.	
Constant	-1.406	.024	.b	.b	.	.	

Mean dependent var	-0.044	SD dependent var	0.378
Number of obs	916	Chi-square	0.170
Prob > chi2	0.680	Akaike crit. (AIC)	196.426

*** *p<.01*, ** *p<.05*, * *p<.1*

9.1.9 FAT Robust regression

Linear regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpar~l	.797	.377	2.12	.035	.058	1.536	**
Constant	-.066	.015	-4.35	0	-.095	-.036	***

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.013	Number of obs	916
F-test	4.476	Prob > F	0.035
Akaike crit. (AIC)	807.981	Bayesian crit. (BIC)	817.621

*** *p<.01*, ** *p<.05*, * *p<.1*

9.1.10 PET robust regression

Regression results

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Precision_se	0	0	-0.37	.709	0	0	
Constant	.006	.033	0.19	.852	-.059	.071	

Mean dependent var	-0.044	SD dependent var	0.378
Overall r-squared	0.001	Number of obs	916
Chi-square	0.140	Prob > chi2	0.709
R-squared within	0.000	R-squared between	0.004

*** *p<.01*, ** *p<.05*, * *p<.1*

9.2 Multiple MRA full models and VIF-test

9.2.1 WLS Robust full

Linear regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpar~l	-.498	.539	-0.92	.356	-1.556	.56	
olsfe	-.573	.085	-6.77	0	-.739	-.407	***
endog	-.218	.045	-4.81	0	-.306	-.129	***
Ginimeasure1yes	-.203	.048	-4.26	0	-.297	-.11	***

Incomeshareofric h~20	-.027	.052	-0.52	.607	-.129	.075	
Shareofmiddleclas ~40	.034	.056	0.61	.54	-.076	.145	
Shareofthebottom b~40	.069	.057	1.22	.224	-.042	.18	
Shareofbottomand mi~s	.491	.084	5.85	0	.326	.655	***
Primaryschooling 1yes	-.289	.033	-8.64	0	-.355	-.223	***
Secondaryschooli ng~s	-.024	.036	-0.66	.51	-.096	.048	
Tertiaryschooling 1~s	.171	.078	2.19	.029	.018	.325	**
Literacyrate	-.028	.083	-0.34	.734	-.192	.135	
Laggeddependent var~e	-.167	.032	-5.20	0	-.23	-.104	***
Endofbefore1990s	.047	.054	0.88	.38	-.059	.154	
Newmeta	.329	.089	3.69	0	.154	.504	***
oecd	.106	.077	1.38	.168	-.045	.257	
Politicaljournal	.196	.099	1.99	.047	.002	.39	**
Sociologyjournal	-.05	.172	-0.29	.772	-.386	.287	
Economyjournal	0	.038	-0.01	.992	-.074	.074	
WorldBank1yes	-.131	.051	-2.56	.011	-.232	-.031	**
Barro1yes	.377	.065	5.83	0	.25	.504	***
Constant	.08	.103	0.78	.434	-.121	.282	

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.259	Number of obs	915
F-test	14.865	Prob > F	0.000
Akaike crit. (AIC)	-365.720	Bayesian crit. (BIC)	-259.704

**** p<.01, ** p<.05, * p<.1*

9.2.2 VIF-test

Variance inflation factor

	VIF	1/VIF
olsfe	42.425	.024
Barro1yes	25.025	.04
Incomeshareofrich~20	15.996	.063
WorldBank1yes	15.633	.064
Newmeta	15.129	.066
endog	12.055	.083
oecd	10.173	.098
Ginimeasure1yes	7.549	.132
Economyjournal	6.715	.149
Secondaryschooling~s	6.32	.158
Laggeddependentvar~e	6.086	.164
Shareofmiddleclas~40	5.394	.185
Shareofthebottomb~40	4.231	.236
Endofbefore1990s	3.352	.298
Politicaljournal	2.362	.423
Standarderrorofpar~l	2.271	.44
Tertiaryschooling1~s	2.174	.46
Shareofbottomandmi~s	1.89	.529

Literacyrate	1.693	.591
Primaryschooling1yes	1.285	.778
Sociologyjournal	1.204	.831
Mean VIF	8.998	.

9.2.3 WLS robust clustered

Linear regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpa r~l	-.498	.936	-0.53	.597	-2.371	1.375	
olsfe	-.573	.274	-2.09	.041	-1.12	-.025	**
endog	-.218	.083	-2.62	.011	-.384	-.051	**
Ginimeasure1yes	-.203	.084	-2.42	.019	-.372	-.035	**
Incomeshareofric h~20	-.027	.132	-0.20	.841	-.291	.238	
Shareofmiddleclas ~40	.034	.176	0.20	.846	-.318	.387	
Shareofthebottom b~40	.069	.174	0.40	.693	-.278	.416	
Shareofbottomand mi~s	.491	.224	2.19	.032	.043	.938	**
Primaryschooling 1yes	-.289	.137	-2.11	.039	-.563	-.014	**
Secondaryschooli ng~s	-.024	.071	-0.34	.736	-.166	.118	
Tertiaryschooling 1~s	.171	.121	1.41	.163	-.071	.414	
Literacyrate	-.028	.086	-0.33	.744	-.201	.144	
Laggeddependent var~e	-.167	.095	-1.76	.084	-.356	.023	*
Endofbefore1990s	.047	.087	0.54	.589	-.127	.222	
Newmeta	.329	.153	2.16	.035	.024	.634	**
oecd	.106	.067	1.59	.117	-.027	.24	
Politicaljournal	.196	.158	1.24	.218	-.119	.511	
Sociologyjournal	-.05	.158	-0.31	.755	-.367	.267	
Economyjournal	0	.059	-0.01	.995	-.118	.117	
WorldBank1yes	-.131	.09	-1.45	.151	-.312	.049	
Barro1yes	.377	.155	2.43	.018	.066	.688	**
Constant	.08	.177	0.45	.651	-.273	.434	

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.259	Number of obs	915
F-test	.	Prob > F	.
Akaike crit. (AIC)	-367.720	Bayesian crit. (BIC)	-266.523

**** p<.01, ** p<.05, * p<.1*

9.2.4 Fixed effects full

Linear regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpa r~l	.075	.042	1.77	.083	-.01	.16	*
olsfe	-.027	.081	-0.34	.739	-.189	.135	
endog	.066	.037	1.79	.078	-.008	.141	*
Ginimeasure1yes	.057	.083	0.69	.494	-.109	.224	
Incomeshareofric	-.067	.074	-0.91	.367	-.215	.081	

h~20							
Shareofmiddleclas ~40	.317	.088	3.58	.001	.14	.494	***
Shareofthebottom b~40	.27	.074	3.65	.001	.122	.417	***
Shareofbottomand mi~s	.484	.111	4.37	0	.263	.705	***
Primaryschooling 1yes	-.118	.161	-0.73	.467	-.44	.205	
Secondaryschooli ng~s	-.202	.144	-1.40	.167	-.49	.087	
Tertiaryschooling 1~s	.027	.163	0.17	.869	-.299	.353	
Literacyrate	-.192	.143	-1.34	.186	-.479	.095	
Laggeddependent var~e	-.223	.023	-9.72	0	-.269	-.177	***
Endofbefore1990s	-.033	.134	-0.25	.804	-.301	.234	
Newmeta	-.202	.108	-1.88	.065	-.418	.013	*
oecd	.004	.013	0.30	.769	-.022	.03	
Politicaljournal	.224	.118	1.90	.063	-.012	.461	*
Sociologyjournal	-.139	.056	-2.47	.016	-.252	-.027	**
Economyjournal	.561	.14	4.02	0	.282	.84	***
WorldBank1yes	-.449	.034	-13.06	0	-.518	-.38	***
Barro1yes	-.301	.017	-17.36	0	-.336	-.267	***
Study no. : base 1	0	.	.	.	.	.	
2	-.562	.14	-4.00	0	-.843	-.281	***
3	-.157	.196	-0.80	.425	-.55	.235	
4	-.698	.13	-5.36	0	-.959	-.437	***
5	-.78	.166	-4.70	0	-1.112	-.448	***
6	.497	.053	9.31	0	.391	.604	***
7	-.543	.116	-4.69	0	-.774	-.311	***
8	-.611	.151	-4.05	0	-.913	-.309	***
9	-1.028	.152	-6.76	0	-1.332	-.724	***
10	-.776	.163	-4.75	0	-1.103	-.45	***
11	-.254	.155	-1.64	.107	-.565	.057	
12	-.716	.161	-4.44	0	-1.039	-.393	***
13	-.268	.089	-3.02	.004	-.445	-.09	***
14	-.803	.132	-6.07	0	-1.067	-.538	***
15	.469	.108	4.33	0	.252	.686	***
16	-.266	.137	-1.94	.057	-.539	.008	*
17	-.73	.125	-5.85	0	-.98	-.48	***
18	-.503	.118	-4.26	0	-.738	-.267	***
19o	0	.	.	.	.	.	
20	.17	.123	1.38	.174	-.077	.416	
21	-.691	.132	-5.23	0	-.956	-.427	***
22	-1.202	.142	-8.47	0	-1.485	-.918	***
23	-1.245	.142	-8.78	0	-1.529	-.962	***
24	-.905	.198	-4.56	0	-1.302	-.508	***
25	-.759	.138	-5.51	0	-1.034	-.484	***
26	-.251	.177	-1.41	.163	-.605	.104	
27	-.509	.147	-3.46	.001	-.802	-.215	***
28	-.769	.126	-6.09	0	-1.022	-.517	***
29	.21	.145	1.45	.152	-.08	.5	
30	-.092	.15	-0.61	.544	-.391	.208	
31	.116	.101	1.15	.254	-.085	.317	
32	.199	.146	1.36	.18	-.094	.492	
33	-.242	.062	-3.87	0	-.366	-.117	***
34	-.566	.171	-3.31	.002	-.908	-.223	***
35	-.884	.158	-5.61	0	-1.199	-.569	***
36	-1.289	.14	-9.23	0	-1.568	-1.01	***

37	-.247	.131	-1.90	.063	-.508	.014	*
38	-.709	.15	-4.72	0	-1.009	-.408	***
39	-1.081	.087	-12.47	0	-1.255	-.908	***
40	-1.157	.127	-9.08	0	-1.412	-.902	***
41	-.556	.038	-14.66	0	-.632	-.48	***
42	-1.85	.074	-24.92	0	-1.999	-1.702	***
43	-.66	.155	-4.26	0	-.969	-.35	***
44	-.181	.066	-2.75	.008	-.313	-.049	***
45	-.529	.118	-4.49	0	-.765	-.294	***
46	.291	.065	4.45	0	.16	.422	***
47	-.601	.093	-6.44	0	-.788	-.414	***
48	-1.188	.101	-11.82	0	-1.389	-.987	***
49	-.114	.174	-0.65	.516	-.462	.234	
50	-.715	.123	-5.82	0	-.961	-.47	***
51	-.21	.136	-1.54	.128	-.482	.062	
52	-.772	.123	-6.27	0	-1.019	-.526	***
53o	0	.	.	.	.	.	
54	.081	.094	0.86	.393	-.107	.27	
55o	0	.	.	.	.	.	
56	.045	.012	3.62	.001	.02	.07	***
57	.128	.154	0.83	.41	-.18	.436	
58o	0	.	.	.	.	.	
59o	0	.	.	.	.	.	
60o	0	.	.	.	.	.	
61o	0	.	.	.	.	.	
Constant	.542	.224	2.43	.018	.095	.989	**

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.686	Number of obs	915
F-test	.	Prob > F	.
Akaike crit. (AIC)	-217.179	Bayesian crit. (BIC)	-154.533

**** p<.01, ** p<.05, * p<.1*

9.2.5 MEML full

Mixed-effects ML regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpar~l	.119	.17	0.70	.485	-.215	.452	
olsfe	-.068	.048	-1.40	.162	-.162	.027	
endog	.047	.03	1.55	.12	-.012	.105	
Ginimeasure1yes	.058	.051	1.15	.249	-.041	.158	
Incomeshareofrich~20	-.068	.05	-1.36	.175	-.165	.03	
Shareofmiddleclas~40	.307	.064	4.78	0	.181	.433	***
Shareofthebottomb~40	.263	.047	5.55	0	.17	.356	***
Shareofbottomandmi~s	.473	.068	6.95	0	.34	.607	***
Primaryschooling1yes	-.05	.047	-1.06	.289	-.143	.043	
Secondaryschooling~s	-.147	.038	-3.85	0	-.222	-.072	***
Tertiaryschooling1~s	.083	.049	1.67	.095	-.014	.179	*
Literacyrate	-.13	.06	-2.16	.031	-.248	-.012	**
Laggeddependentva	-.159	.06	-2.66	.008	-.276	-.042	***

r~e						
Endofbefore1990s	.06	.086	0.70	.482	-.108	.229
Newmeta	-.036	.093	-0.38	.703	-.218	.147
oecd	.01	.036	0.27	.786	-.06	.08
Politicaljournal	.016	.138	0.12	.908	-.254	.286
Sociologyjournal	-.242	.246	-0.98	.326	-.725	.241
Economyjournal	-.11	.078	-1.40	.16	-.263	.043
WorldBank1yes	-.021	.069	-0.30	.762	-.155	.114
Barro1yes	.113	.082	1.38	.167	-.047	.273
Constant	.072	.117	0.61	.542	-.158	.301
Constant	-1.495	.107	.b	.b	.	.
Constant	-1.516	.024	.b	.b	.	.

Mean dependent var	-0.044	SD dependent var	0.378
Number of obs	915	Chi-square	220.281
Prob > chi2	0.000	Akaike crit. (AIC)	18.063

**** p<.01, ** p<.05, * p<.1*

9.2.6 G-to-S

Mean dependent var	-0.044	SD dependent var	0.378
Number of obs	915	Chi-square	220.281
Prob > chi2	0.000	Akaike crit. (AIC)	18.063

**** p<.01, ** p<.05, * p<.1*

Wald test, begin with full model:
p = 0.9965 >= 0.0500, removing Economyjournal
p = 0.8575 >= 0.0500, removing Incomeshareofrichtop20
p = 0.9680 >= 0.0500, removing Secondaryschooling1yes
p = 0.8642 >= 0.0500, removing Literacyrate
p = 0.7851 >= 0.0500, removing Standarderrorofpartialcorrel
p = 0.5089 >= 0.0500, removing Sociologyjournal
p = 0.4691 >= 0.0500, removing Endofbefore1990s
p = 0.2637 >= 0.0500, removing Shareofmiddleclassmiddle40
p = 0.2755 >= 0.0500, removing Shareofthebottombottom40
p = 0.1718 >= 0.0500, removing oecd
p = 0.1224 >= 0.0500, removing Ginimeasure1yes
p = 0.4258 >= 0.0500, removing Laggeddependentvariable
p = 0.4184 >= 0.0500, removing Politicaljournal
p = 0.1930 >= 0.0500, removing WorldBank1yes
Linear regression Number of obs = 915
F(7, 907) = 5.71
Prob > F = 0.0000

R-squared	=	0.1829
Root MSE	=	.20402

Partialcorr	Coefficient	Robust std.	err.	t	P>t	[95%
Primaryschoolin g1yes	-0.277	0.124	-2.240	0.026	-0.520	-0.034
olsfe	-0.274	0.096	-2.840	0.005	-0.463	-0.085
endog	-0.176	0.080	-2.200	0.028	-0.333	-0.019
Tertiaryschoolin g1yes	0.139	0.030	4.610	0.000	0.080	0.198
Shareofbottoman dmiddleclass	0.500	0.153	3.270	0.001	0.200	0.800
Newmeta	0.165	0.077	2.160	0.031	0.015	0.316
Barro1yes	0.182	0.048	3.790	0.000	0.088	0.276
_cons	-0.020	0.043	-0.460	0.646	-0.105	0.065

9.2.7 WLS Robust G-to-S

Linear regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpa r~l	.174	.452	0.38	.701	-.714	1.061	
olsfe	-.265	.065	-4.10	0	-.392	-.138	***
endog	-.169	.039	-4.39	0	-.245	-.094	***
Barro1yes	.182	.046	3.97	0	.092	.272	***
Primaryschooling 1yes	-.275	.032	-8.62	0	-.338	-.213	***
Tertiaryschooling 1~s	.14	.056	2.48	.013	.029	.251	**
Shareofbottomand mi~s	.498	.069	7.26	0	.364	.633	***
Newmeta	.159	.047	3.37	.001	.067	.252	***
Constant	-.026	.037	-0.71	.475	-.099	.046	

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.183	Number of obs	915
F-test	25.370	Prob > F	0.000
Akaike crit. (AIC)	-302.375	Bayesian crit. (BIC)	-259.005

**** p<.01, ** p<.05, * p<.1*

9.2.8 WLS Robust Clustered G-to-S

Linear regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpa	.174	.735	0.24	.814	-1.297	1.644	

r~l							
olsfe	-.265	.11	-2.42	.019	-.485	-.046	**
endog	-.169	.061	-2.76	.008	-.292	-.047	***
Barro1yes	.182	.071	2.57	.013	.041	.324	**
Primaryschooling 1yes	-.275	.132	-2.09	.041	-.539	-.011	**
Tertiaryschooling 1~s	.14	.032	4.43	0	.077	.203	***
Shareofbottomand mi~s	.498	.227	2.20	.032	.045	.952	**
Newmeta	.159	.079	2.02	.048	.001	.317	**
Constant	-.026	.046	-0.57	.571	-.119	.066	

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.183	Number of obs	915
F-test	7.634	Prob > F	0.000
Akaike crit. (AIC)	-302.375	Bayesian crit. (BIC)	-259.005

**** p<.01, ** p<.05, * p<.1*

9.2.9 FE G-to-S

Linear regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpa r~l	.073	.045	1.62	.112	-.017	.162	
olsfe	-.042	.09	-0.47	.639	-.221	.137	
endog	.067	.036	1.86	.068	-.005	.14	*
Barro1yes	-.409	.038	-10.72	0	-.485	-.332	***
Primaryschooling 1yes	.078	.079	0.99	.326	-.08	.236	
Tertiaryschooling 1~s	.208	.094	2.22	.03	.02	.396	**
Shareofbottomand mi~s	.328	.121	2.72	.009	.087	.569	***
Newmeta	.271	.106	2.55	.013	.059	.484	**
Study no. : base 1	0	.	.	.	.	.	
2	-.026	.001	-40.74	0	-.027	-.024	***
3	.154	.047	3.29	.002	.06	.248	***
4	.396	.018	21.80	0	.359	.432	***
5	-.061	.013	-4.72	0	-.087	-.035	***
6	.307	.017	17.62	0	.272	.342	***
7	-.008	.003	-2.76	.008	-.014	-.002	***
8	.183	.047	3.91	0	.09	.277	***
9	.031	.031	1.01	.316	-.03	.093	
10	-.356	.03	-12.04	0	-.415	-.297	***
11	.25	.012	20.71	0	.226	.275	***
12	-.345	.03	-11.54	0	-.405	-.286	***
13	.393	.006	62.26	0	.381	.406	***
14	-.124	.03	-4.18	0	-.183	-.064	***
15	.293	.052	5.66	0	.189	.396	***
16	-.14	.054	-2.59	.012	-.249	-.032	**
17	.056	.002	22.68	0	.051	.061	***
18	.035	.012	2.93	.005	.011	.059	***
19	-.322	.035	-9.31	0	-.391	-.253	***
20	-.065	.029	-2.20	.032	-.124	-.006	**
21	-.075	.029	-2.56	.013	-.134	-.016	**
22	-.102	.03	-3.44	.001	-.161	-.043	***
23	-.112	.029	-3.79	0	-.171	-.053	***

24	-.249	.099	-2.52	.015	-.447	-.051	**
25	.312	.03	10.57	0	.253	.371	***
26	-.156	.091	-1.72	.091	-.338	.026	*
27	-.11	.029	-3.74	0	-.169	-.051	***
28	.016	.012	1.34	.186	-.008	.04	
29	.647	.03	21.85	0	.588	.706	***
30	.414	.007	63.38	0	.4	.427	***
31	.842	.026	32.37	0	.79	.894	***
32	.704	.022	32.70	0	.661	.747	***
33	-.153	.03	-5.17	0	-.212	-.094	***
34	.291	.031	9.44	0	.229	.353	***
35	-.195	.062	-3.14	.003	-.319	-.071	***
36	-.006	.03	-0.20	.843	-.065	.053	
37	-.495	.11	-4.48	0	-.716	-.274	***
38	-.936	.063	-14.83	0	-1.062	-.81	***
39	-.455	.115	-3.96	0	-.685	-.225	***
40	-.621	.102	-6.12	0	-.824	-.418	***
41	-.106	.078	-1.36	.179	-.263	.05	
42	-1.212	.101	-11.97	0	-1.415	-1.01	***
43	-.594	.114	-5.21	0	-.822	-.366	***
44	-.315	.108	-2.91	.005	-.531	-.099	***
45	-.544	.084	-6.52	0	-.712	-.377	***
46	.159	.109	1.45	.151	-.059	.377	
47	-.424	.113	-3.76	0	-.649	-.198	***
48	-.785	.115	-6.80	0	-1.016	-.555	***
49	-.189	.115	-1.63	.107	-.42	.042	
50	-.572	.115	-4.96	0	-.802	-.341	***
51	-.229	.123	-1.87	.067	-.474	.016	*
52	-.53	.055	-9.68	0	-.64	-.421	***
53	-.362	.119	-3.04	.004	-.6	-.124	***
54	.258	.115	2.25	.028	.029	.488	**
55	.235	.099	2.38	.021	.037	.433	**
56	.045	.012	3.69	0	.02	.069	***
57	.781	.017	45.96	0	.747	.815	***
58o	0	.	.	.	.	.	
59	-.098	.047	-2.09	.04	-.191	-.004	**
60	.307	.029	10.41	0	.248	.366	***
61o	0	.	.	.	.	.	
Constant	.06	.029	2.05	.045	.001	.119	**

Mean dependent var	-0.044	SD dependent var	0.378
R-squared	0.635	Number of obs	915
F-test	.	Prob > F	.
Akaike crit. (AIC)	-95.378	Bayesian crit. (BIC)	-66.465

*** *p<.01*, ** *p<.05*, * *p<.1*

9.2.10 MEML G-to-S

Mixed-effects ML regression

Partialcorr	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Standarderrorofpar~l	.127	.183	0.69	.487	-.231	.485	
olsfe	-.09	.051	-1.75	.08	-.19	.011	*
endog	.03	.032	0.94	.347	-.032	.092	
Barro1yes	.033	.07	0.47	.639	-.105	.171	
Primaryschooling1yes	.088	.034	2.59	.01	.021	.154	***
Tertiaryschooling	.19	.04	4.75	0	.112	.269	***

1~s							
Shareofbottomand mi~s	.336	.056	5.96	0	.225	.446	***
Newmeta	-.07	.074	-0.95	.343	-.215	.075	
Constant	-.034	.051	-0.66	.507	-.135	.067	
Constant	-1.473	.108	.b	.b	.	.	
Constant	-1.442	.024	.b	.b	.	.	

Mean dependent var	-0.044	SD dependent var	0.378
Number of obs	915	Chi-square	69.931
Prob > chi2	0.000	Akaike crit. (AIC)	121.986

*** *p<.01*, ** *p<.05*, * *p<.1*