# A META-ANALYSIS OF INCOME AND DEMOCRACY

Why we still have not reached consensus
after 60 years of research


UNIVERSITY OF BERGEN
*Faculty of Social Sciences*

Troy Saghaug Broderstad

---

Master Thesis
Department of Comparative Politics

Spring 2015

# Abstract

The purpose of this study is to examine the relationship between per capita income and democracy. Namely, do increases in income per capita affect a regime's level of democracy? The scholarly tradition has investigated this question over many years, and what we have learned about the actual association between these variables remains inconclusive. As opposed to producing yet another empirical analysis of this hypothesized relationship, this thesis sets forth to examine it via a new methodological approach. Applying the tools of meta-regression analysis to 33 individual empirical studies that investigate the relationship between income and democracy, this thesis finds that income has no statistically significant, quantitatively meaningful effect on democracy (understanding the latter as a graded concept). I also show that there is no theoretical or empirical reason to believe that this field of research is "haunted" by publication selection bias. Moreover, I identify a variety of systematic differences between these studies, i.e. study heterogeneity, which explain why it is the case that after 60 years of research we still have not reached a consensus.

## Acknowledgements

Firstly, I would like to thank my supervisor Michael Alvarez for his enthusiasm and endorsement. Throughout the entire project he has provided insightful feedback, firm guidance and a substantial amount of patience. For this I am truly obliged.

Secondly, I am very grateful to the Challenges and Transformations in Advanced Democracies-group for allowing me to present an early draft of the thesis. I would especially like to thank Research Fellow Erla Løvseth, Post-Doctoral Fellow Yvette Peters, Associate Professor Michaël R. Tatham and Håkon Jørgen Tranvåg for insightful comments. Regardless, any faults and/or shortcomings remain entirely my own.

Finally, if not for you, Ingri, I would be nowhere at all.

Bergen, June 2015

Troy Saghaug Broderstad

# Table of contents

# List of figures



# List of Tables

# List of Abbreviations

| | |
|---|---|
| 2SLS | Two-step least squares |
| 3SLS | Three-step least squares |
| CL | Civil liberties |
| TSCS | Time-series cross-sectional |
| FAT-PET-MRA | Funnel-asymmetry-test |
| FE | Fixed effects |
| FHI | Freedom House Index |
| GDP | Gross domestic product |
| GMM | Generalised method of moments |
| G-to-S | General-to-specific modelling |
| IV | Instrumental variable(s) |
| MAER | Meta-analysis of economic research network |
| MEML | Mixed effects multi-level |
| MRA | Meta-regression analysis |
| OECD | Organisation for Economic Co-operation and Development |
| OLS | Ordinary least squares |
| PL | Political liberties |
| PWT | Penn World Tables |
| RE | Random effects |
| WBD | World Bank Data |
| 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

Why do some countries have democratic regimes while others do not, and why do some democracies survive while others fail? These are amongst the core questions that scholars of political science, sociology, history, and economics have been trying to answer since Aristotle first asked the question some 2.500 years ago. After centuries of research on this phenomenon, the number of hypotheses concerning what causes democratisation is manifold, and might sometimes seem incomprehensible. This thesis sets out to investigate one of the most influential of these hypotheses, namely that democracy is a function of income, or more precisely, that *higher levels of income per capita (and increases therein) increase the likelihood of transitions toward democracy*.

In the modern era, all OECD countries are democratic, while many of the non-democracies are found in the poorer parts of the world, like sub-Saharan Africa and Southeast Asia. This positive cross-country correlation between income and democracy is very clear. Acemoglu, Johnson, Robinson & Yared (2008) show this cross-correlation in the 1990s and Coppedge (2012) shows the correlation for the 1950s to the 2000s.[1] This association is famously explained by Lipset (1959) and his renowned *modernisation theory*. Said theory is based on the observation that income – operating through a variety of channels such as education, urbanisation, industrialisation, etc. – is thought to bring with it the "political correlates of democracy" (Lipset 1959: 80). But as the discipline has developed, the question has remained: Are these variables *causally* related? Unfortunately, no consensus has yet developed: Many scholars find that income does in fact cause democratisation, whilst others claim there is no causal relationship.

Hence, the debate about the nature of the relationship between income and democracy continues. But rather than producing yet another empirical analysis of this phenomenon, accommodating every aspect of the debate, this thesis sets forth to examine this relationship in a meta-analytical framework. Through the rigorous statistical techniques of meta-regression analysis, the findings of the last fifteen years of quantitative research will be analysed and the association between income per capita and democracy will be tested in order to determine what we actually know. In this thesis I will integrate conflicting research findings in an attempt to sift out the noise and to access "the nuggets of "truth" that have

[1] Both figures are reproduced in the appendix in Figure 7.1 and 7.2, respectively.
[2] The original sample consisted of 603 estimates in total as it includes both dichotomous and continuous

settled at the bottom" (Stanley & Doucouliagos 2012: 2). Meta-regression analysis (MRA), when replicable and conducted properly, offers such a methodology. I therefore derive the following main research question:

*$Q_1$: What is the effect of per capita income on democracy?*

To answer this question using Meta-Regression Analysis, I have created an original meta-data set consisting of *33 quantitative studies* with *492 corresponding estimates* which measure the association between per capita income and democracy. When conducting the analysis I have chosen to focus exclusively on a graded conceptualisation of democracy. There are two reasons for this: (I) When collecting the sample upon which the original meta-data set was constructed, only about 15% of collected estimates used a dichotomous conceptualisation and measure.[2] (II) Graded and dichotomous conceptualisations of democracy are not comparable in a meta-regression framework. Consequently, in order to get the most out of the multiple meta-regression analysis, in terms of degrees of freedom, the graded conceptualisation of democracy was chosen.

Before answering the main research question, however, two subsidiary questions must be investigated in order to ensure the validity of the meta-regression analysis. Firstly, I must determine if publication selection bias is present in the literature, i.e. if editors, reviewers and authors treat significant results as more important than non-significant results or if a certain type of finding is dominantly reported in the literature. Publication bias is most often found in fields of research that are dominated by a high degree of consensus. Therefore, researchers who produce results that are contradictory to the dominant theory of the field are less likely to be accepted, and hence not get their results published (Stanley & Doucouliagos 2013). In this particular field of study I have no theoretical reason to expect that research that finds a positive relationship between income and democracy be selected in favour of research that finds no effect. This is because arguments both for and against an effect are present in the literature, as I will show below. Still, to ensure the robustness of the over-all conclusion of the meta-analysis, publication bias has to be controlled throughout the analysis.

Secondly, I will have to identify what factors lead researchers to report systematic differences, i.e. study *heterogeneity*. One of the reasons why the scholarly field has not reached a consensus in answering the question is that researchers in the quantitative field disagree on what statistical approach is the most feasible one (Acemoglu et al. 2008; Bobba &

[2] The original sample consisted of 603 estimates in total as it includes both dichotomous and continuous measures of democracy.

Coviello 2007; Lipset 1959; Barro 1999), which measure for democracy should be used (Boix 2011; Przeworski, Alvarez, Cheibub & Limongi 2000; Teorell 2010), how the sample should be composed in terms of years (Acemoglu et al. 2008; Boix 2011), and countries (Coppedge 2012; Hadenius 1992), and lastly, what control variables that should be included in the model specification (Ross 2009; Acemoglu, Johnson, Robinson & Yared 2005; Benyishay & Betancourt 2010). In order to answer the main research question, I therefore have to sort out how and why scholars still are not certain whether income causes democratisation or not. Thus, I derive the following research questions:

*$Q_{1a}$: Is publication selection bias present in the literature?*
*$Q_{1b}$: Are there systematic differences between studies that generate different results?*

To sum up: $Q_1$ is *de facto* the main question this thesis is trying to answer, but in order to answer this question with the methods of meta-regression analysis, I must first empirically confirm that publication bias is not a problem in this field of research ($Q_{1a}$). Secondly, I must control for possible systematic differences between studies that might affect the results ($Q_{1b}$). This then results in a second research question, very closely related to $Q_{1b}$ but more effectively introduced at this point

*$Q_2$: Why is it that we observe no consensus on this question across the scholarly literature?*

Hence after controlling for possible publication bias and exploring the sources of heterogeneity across studies, I can address research question $Q_2$ as well as the main research question $Q_1$. To provide a glimpse at the findings presented later in this thesis, I can report that I find that publication bias is not a problem in the literature on the income-democracy nexus, that there exist significant and consequential differences between the studies, and that this heterogeneity can account for much of the variation and differences in the study-outcomes. In conclusion, I will show that the multiple meta-regression analysis ultimately demonstrates that there is *no robust effect of income on democracy* when using what the literature deems as "best practice" research design.

## 1.1 Contribution

This thesis is complimentary to the scholarly tradition in two ways: methodological and theoretical. The thesis contributes methodologically to the growing field of meta-regression analysis in social science and economics (Stanley & Doucouliagos 2012). While meta-analysis in social science is a newly discovered method, quite few meta-regression analyses have been conducted. For example: Doucouliagos & Ulbasoglu (2008) look at the relationship between democracy and economic growth, Doucouliagos & Paldam (2008) study the effect of aid on democracy, Efendic, Pugh & Adnett (2011) look at institutions and economic performance, Ludvigsen (2010) examines economic voting, and more recently, Ahmadov (2014) looks at how oil affect democracy. Few meta-regression analyses have been conducted in the field of comparative politics, and the specific question chosen in this thesis is in great need of a robust consideration of the current state of knowledge in order to guide further research. This thesis draws on all of these works cited above, as well as Stanley & Doucouliagos (2012) and the Meta-Analysis of Economic Research Network (MAER) (Stanley et al. 2015), which contributes to state-of-the-art methodology in meta-regression analysis.

Theoretically, this thesis ultimately shows that modernisation theory offers "little, if any, explanatory power" (Przeworski et al. 2000: 80), understanding democracy as a graded concept. It also shows that different methodological choices, in terms of estimation technique, choice of dependent variable, choice of data set for the independent variables, attributes of the authors, and control variables chosen matter for the divergent research outcomes observed.

## 1.2 Outline of the thesis

In the *second chapter* I review the relevant literature in a chronological manner starting with Lipset (1959) and modernisation theory. After tracing some of the publications modernisation theory has generated, I will discuss the literature that claims income does not have an effect on democracy. To conclude chapter two I will identify several aspects that lead authors to arrive at different conclusions, i.e. study heterogeneity.

In the *third chapter* I discuss the methodology of meta-regression analysis and the development of the original meta-data set used in this thesis. I start by discussing how and why meta-regression analysis is the best-known way to review a field of quantitative research. After justifying the choice of method, a discussion of the dependant variable is conducted. The dependent variable in this study is the *partial correlation* between income and democracy

– a uniform measure of the association between these two variables. Utilising the partial correlation as our dependent variable allows us to compare the large variety of divergent results across studies included in my sample. A corresponding statistical measure of *study-quality* is also derived from this dependent variable. Following this, I show how I can use the tool of meta-regression analysis to answer the research questions outlined above. In a concluding section I discuss the data set and the operationalisation of the independent variables thought to capture the heterogeneity of the scholarly field in light of the relevant literature.

In the *fourth chapter*, I conduct the actual analysis starting with inspection of some descriptive statistics. Following this, I show that there is no practically significant amount of publication bias in the literature and that I cannot detect an effect of income on democracy when controlling only for publication bias. Further, I argue that controlling for publication bias is not sufficient, and that controls for study-heterogeneity, i.e. systematic differences between studies, must be undertaken. Therefore a multiple meta-regression analysis is conducted. Through this statistical model I show that sources of heterogeneity such as the variety of estimation methods, attributes of the authors and specification of the research design account for 60% of the observed variation in study-outcomes. After addressing publication bias and study heterogeneity, I show that when applying a proper research design, the effect of income on democracy proves to be non-existent.

In the *fifth chapter* I present the conclusion of the thesis. Here, I sum up the findings, discuss them, and present suggestions for further research.

## 2. Income and democracy

The purpose of this review is to present a brief account of the scholarly literature on the relationship between per capita income and democracy. Before examining the empirical evidence using a meta-regression analysis, I will discuss the ways in which the scholars in the field view the process of democratisation theoretically, and how the latter may or may not be caused by increased income per capita. In an introductory discussion I review the different views on this, and argue for the specific approach examined in this analysis. For methodological, as well as theoretical reasons, the chosen approach in the meta-analysis is to investigate the effect of income per capita on democracy, where democracy is understood as a graded concept, as opposed to a dichotomous one. As a consequence of this, the studies examined in the meta-analysis understand the process of democratisation as a movement upwards towards a "perfect" democracy, and the absence of movement backwards towards autocracy (Teorell 2010). In this literary review I will, however, use all available arguments and findings, regardless of conceptualisation or method, be it qualitative or quantitative.

With this theoretical backdrop, I will discuss the well-known modernisation theory, most commonly associated with Lipset (1959). This theory views income as a causal determinant of democracy, and has been supported throughout the scholarly tradition. After discussing this, I will briefly take account of the ways in which supporters of this theory have developed it over time. Secondly, I review the opponents of this theory. In this section I will introduce the argument that democracy is an endogenous concept and should be treated as such in empirical analysis. This means that democracy may affect income and income might affect democracy, simultaneously, and has spawned a considerable debate in the literature concerning research methodology. Whilst scholars arguing the case of the modernisation hypothesis also have treated democracy as endogenous, as discussed below, a substantial part of the scholarship find that when doing so in a rigorous manner, the effect of income on democracy is non-existent.

Lastly, I discuss the relevant sources of heterogeneity in the literature based on the review outlined above. Heterogeneity is understood here as the very factor(s) which create varying results in the research – the factors which explain why there is such variation in the findings across this large body of research, thus relating both to $Q_{1b}$ and $Q_2$. These same factors will enter in the multiple meta-regression analysis as independent variables to investigate what causes different studies to report different results. Examples of such factors which might explain why we observe contrasting findings regarding the effect of income

upon democracy include the following: the particular type of econometric technique employed; the composition of the sample; the conceptualisation of democracy; which data set is used to measure income per capita; author attributes; whether the main focus of the study is the income-democracy nexus; whether the first author is an economist or a political scientist; and whether the study is published in a peer-review journal or is an unpublished working paper. With this in mind, I will now present a survey of the literature, as described above.

## 2.1 Traditional views – Modernisation theory

Does income per capita cause the emergence of democracy? To approach this question, a clear understanding of democracy, and the process of democratisation, is needed. To begin with, the study of democratisation actually encompasses a variety of problems including the transition to democracy, transition to autocracy, and/or survival of democracies. In other words, the general question the scholars attempt to answer is: "What causes regimes to rise, endure and fall?" (Przeworski et al. 2000: 78). But how is democratisation itself conceptualised? Scholars that have studied the process of democratisation have, in general, either (I) seen democracy as a graded concept where countries become more democratic (sometimes a cut-off point which indicates a "full democracy" is identified); or (II) as a process in which countries transition to democracy, survive as a democracy, or transition to autocracy with necessary and sufficient criteria for each category. Given these contrasting approaches, how democracy is measured needs to be taken into account, and the different continuous and dichotomous approaches will be discussed at length below. In any case, regardless of the conceptualisation of democracy, the process of democratisation looks at whether a country's political regime moves towards democracy (understood as the absence of movements backwards, toward autocracy) (Teorell 2010).

It is important to observe that the very question: "*Does income have an effect upon democracy?*" has in practice several different meanings in the theoretical and empirical literature. First, in some studies it means the following: Do increases in per capita GDP over time increase the likelihood of a *transition* from a non-democratic regime to a democratic regime? This specific question is often addressed in studies that employ a dichotomous measure of democracy. The two best known studies of this type are Przeworski et al. (2000) and Boix & Stokes (2003) both of which utilise a dynamic probit model to estimate the coefficient which measures the effect of income upon transition to democracy. It is this approach which is perhaps the best-suited for testing modernisation theory, as discussed

below. This hypothesis is referred to in Przeworski et al. (2000) as an internal explanation: Economies develop, and the expectation is that regime change eventually occurs as a result. But Przeworski et al. (2000) likewise show that the relationship between income and democracy could be due to an external explanation: In this argument, income plays no role in affecting a regime transition (actually, democracy may emerges due to other exogenous factors), but income plays a role in stabilising democracy. Both of these arguments would contribute to identifying empirically a relationship between income and democracy.

Another approach widely encountered in the literature and the one which the studies included in this meta-analysis adopt, is to examine whether income has an effect upon democratisation: Do increases in per capita income tend to produce higher levels of democracy? Observe that in this approach there does not necessarily exist a cut-off point which marks a transition from a non-democratic regime to a democratic one. Rather a regime is scored either higher or lower on a continuous democracy scale. I repeat: It is this approach to studying the relationship between income and democracy which is characteristic of the studies included in the present meta-analysis. Note that this approach in fact encompasses a variety of potential causal mechanisms which the other approaches study explicitly. For example, the observed relationship between income and democracy might be due to any of the mechanisms identified above: (I) It might be due to income producing transitions (the modernisation hypothesis); (II) It might be due to income stabilising democracy; (III) It might be due to some other causal mechanism wherein income sets some process into motion or creates circumstances which facilitate the process of democratisation. Because any or all of these mechanisms could be in play to explain any observed relationship between income and democracy, I myself will refer to them more or less interchangeably in this thesis because the main focus is the association between the two variables income and democracy, and not the process of democratisation as a whole.

Having said this, I remind the reader that the main focus of this meta-analysis is the process of democratisation where democracy is taken as a graded concept, asking whether regimes are more likely to become more democratic as countries develop economically, i.e. as income per capita increases. As a consequence of this, and for methodological reasons, studies using a dichotomous independent variable (like Przeworski et al. 2000; Boix & Stokes 2003; Houle 2009; Ansell & Samuels 2010, and alike) are excluded from the meta-data set. [3]

[3] Studies using a dichotomous dependent variable are not easily compared with a continuous dependent variable, both because of different conceptualisations and of the difference in estimation technique; they can therefore not be included in the same meta-analysis (Stanley et al. 2012). This will be elaborated below.

Further, I focus on the school of political scientists that attribute regime dynamics to structural variables, like Lipset (1959) and Dahl (1971) and their intellectual heirs, as opposed to the scholars that pay more attention to interactions between various societal groups in the process of transition, like Moore (1966); and Rueschemeyer, Stephens & Stephens (1992), for instance (Alemán & Yang 2011). Nevertheless, this latter point does not have any effect on the meta-analysis per se.

Turning to the empirical research, the income-democracy nexus is the cornerstone of the well-known modernisation theory put forth by Lipset (1959) Building on ideas dating back to Aristotle, Lipset (1959) argued that improvements in economic standards would ultimately lead to democratisation. He therefore argued that democracy was both created and consolidated by the broad process of modernisation. This process involved changes in the factors of industrialisation, urbanisation, wealth and education, and said changes that are so closely interrelated that they form one common factor. In other words: These are factors subsumed under economic development and then carry with them the political correlate of democracy, he argued. Because of this reasoning, one of the central conclusions of modernisation theory is the following: Once an authoritarian country reaches a certain level of economic development, the regime becomes democratic. Likewise, Lipset (1960: 31) also concludes that: "The more well-to-do a nation, the greater the chances that it will sustain democracy", meaning that economic development both causes democracies to *emerge* and to *survive*.[4]

In this analysis, Lipset had dichotomous indicators of regime stability and continuous socioeconomic indicators. He therefore chose to compare the average values of socioeconomic indicators for stable democracies versus unstable democracies or dictatorships in Europe, and for democracies and unstable dictatorships versus stable dictatorships in Latin America. This type of analysis, however, was inappropriate for determining the impact of socioeconomic development on democracy because it held the regime categories constant and then allowed the development indicators to vary. A more appropriate way in which this could be done would by using logit or probit models, but this technique was not adopted in social science before 1975, even though it was invented, respectively, in the 1930s and 40s (Aldrich & Cnudde 1975). Coleman & Almond (1960) also committed the same methodological flaw in their investigation. However, if we are not concerned about causal direction and just the

[4] As Przeworski et al. (2000) argue, either or both of these effects of income upon democracy can explain the observed correlation between income and level of democracy.

correlation, the analysis demonstrated a positive bivariate association between democracy and many aspects of development, income per capita included.

When continuous indicators of democracy became available, social scientists in the 1960s made use of correlation and regression analysis. Cutright (1963) found that among other variables associated with modernisation theory, economic growth accounted for about two thirds of the variance in national political development in 77 countries. His data consisted of an indicator of the stability of formal democracy from 1940-1960. Given his era, the analysis stands out for reporting slope coefficients, the $R^2$, imputing missing values, plotting the regression line and its confidence interval on a scatterplot and analysing outliers. He also "predicted that the outliers would regress to the line of fit, this exaggerated the implications of his estimates", thus creating a bias (Coppedge 2012: 147). Amongst these early scholars, Adelman & Morris (1965) also stand out. They constructed an indicator of political Westernisation and showed trough a factor analysis that this was associated with per capita income in a nonlinear fashion. Revisiting this, Jackman (1973) found the relationship to be in fact logarithmic, i.e. curve-linear. In his sample of 60 developed and undeveloped countries, it seemed that the log of income was a strong predictor for level of democracy in 1960. Logging per capita GDP also eliminates heteroscedasticity and keeps the assumption of linearity that is crucial in regression analysis. In modern econometrics, this is standard in time-series cross-sectional analysis (TSCS) (Skog 2007; Beck & Katz 2001), but scholars of that time were not seemingly aware of this .

This discovery brought the scholarly scope away from modernisation as a concept that included many aspects, such as industrialisation, education, urbanisation, etc., and to focus more specifically on per capita income. In the early years of the discipline the process of modernisation was seen as a single syndrome that moved in parallel, which made it impossible to determine the direction of causality. It was also impossible to determine the direction as long as scholars had to work with cross-sectional data, which continued to be the methodological norm up until the 1990s. The publication of *Democracy and Development* (Hadenius 1992) proved, in retrospect, to be the most comprehensive cross-sectional analysis in the history of the discipline. He employed a sample of 130 developing countries and tested a great variety of predictors concerned with economic and social modernisation finding "no connection whatever [*sic*] between an asymmetrical degree of development and democracy" (Hadenius 1992: 90).[5] This also ran contrary to the dominant scholarly opinions at that time.

[5] See also Mainwaring et al. (2003) for a cross-sectional approach in Latin America with the same conclusion.

Huntington (1968) for instance, also hypothesised a curvilinear relationship between modernisation and political development. It should be noted, however, that Huntington (1968) was primarily concerned with political stability, and not democracy.

However, conclusions based on the cross-sectional approach are deemed as premature, due to the limitations of cross-sectional analysis. There are four reasons for this. (I) An important assumption in cross-sectional analysis is that differences across countries are equivalent to changes within countries. This is problematic because countries at the same level of socioeconomic modernisation are likely to be at different levels of democracy as a result of factors other than modernisation. (II) Cross-sectional analysis provides no basis for generalising its findings to other years. Due to the nature of causal relationships in general, it is possible that an association exists in one year but disappears in another. There is also a risk of spuriousness, i.e. that the causal relationship is random or due to some unspecified variable which is correlated with both income and democracy. (III) Cross-sectional estimates can be sensitive to sample composition. Hadenius (1992), for instance, included only developing countries, which might have led to biased results. Coppedge & Midlarsky (1997) found that which modernisation variables matter and whether their effects were positive or negative could vary greatly depending on which regions where included or included in the sample. (IV) In cross-sectional analysis it is impossible to disentangle the effects of variables that are strongly correlated with one another, known as the problem of multicollineariy. "It could be argued that multicollinearity gave rise to the very concept of modernisation, the essence of which was parallel change in a large number of collinear variables" (Coppedge 2012: 149).

While methodologically flawed, Lipset's conclusions have been reproduced in some major qualitative studies on democracy (Dahl 1971; Huntington 1991; Rueschemeyer et al. 1992). Diamond (1992) argued that income per capita was the most robust modernisation variable, thus establishing itself as the most important variable in research on modernisations effect on democracy. Moving from this, scholars in the 1980s began to develop democratisation data sets that contain observations on multiple countries and multiple years. Since about the year 2000 all analyses that have been published in leading journals have used time-series cross-sectional (TSCS) data (Coppedge 2012).[6] This also brought on the need for more advanced statistical techniques. Without it, the cross-national differences and changes within countries are simply pooled together as undifferentiated variance to be explained. If there is more variation across countries than there is within them, a regression with a simple

[6] Based on this fact (and the problems associated with cross-sectional analyses) this is therefore the main scope of this thesis' analysis.

pooled sample still explains cross-sectional variation in levels of democracy more than it explain change. Further, there are several problems in TSCS data related to the Gauss-Markov theorem.[7] Amongst these problems we find the following: It is likely that democracy is systematically over- or under-predicted in certain countries or years (fixed effects); causal or cluster heterogeneity is present; democracy is easier or harder to predict in some countries or years (heteroscedasticity); one country influence another country's level of democracy (spatial autocorrelation); the level of democracy in a country in one year is dependent on the level of democracy in another year (autocorrelation) (Coppedge 2012). These problems are unavoidable in TSCS, but there are ways of dealing with them, like cluster-robust standard errors and fixed and random effects specifications. The discussion about what econometric method is most feasible is vast in the literature and is thus a very important source of heterogeneity which much be addressed. Hence, I will summarise the debate in a subsequent section.

In the context of large-*n* studies using TSCS, Lipset's hypothesis (1959) receives empirical support from Barro (1999). He finds, among other effects associated with modernisation theory, that income is a positive determinant of democracy in a large sample of about 100 countries, developed and undeveloped, within the time period 1960-1995. He also includes an autoregressive term (AR1) in the model with 5- and 10-year lags. This statistical approach that will be discussed in full below, but in short it will suffice to say that it is used to address the possible simultaneous effects of democracy, i.e. endogeneity in the model.[8] He notes that the increase in various measures of the standard of living predicts a gradual increase of democracy. In contrast, democracies that arise without prior economic development tend not to last.

For about 40 years the modernisation theory stood the test of falsification, until Przeworski et al. (2000)[9] introduced their empirical investigation of democratisation. This will be discussed in length below, but the essential argument is that income does not cause autocracies to transition to democracies, but sustains established democracies. In an earlier work, Przeworski & Limongi (1997) claim that in poor countries the value of becoming a dictator is greater than in wealthy countries, "hence struggle for dictatorship is more attractive in poor countries" (Przeworski & Limongi 1997: 166). These conclusions created a debate in the scholarly field, and as a response to this, Boix & Stokes (2003), using game theoretic and

[7] See Gujarati Porter (2010) for discussion.
[8] Below, I will also discuss whether this is the correct specification to model endogeneity.
[9] See also Przeworski & Limongi 1997.

statistical tools, reason that when the status quo is dictatorship, economic growth increases the incentives for the ruling faction to democratise. They agree, however, that when democracy is the status quo income does increase the stability of democracy.

Moving from the theoretical argument to the empirical analysis, Boix & Stokes (2003) argue that the Przeworski et al. (2000) findings are biased because they (I) observe very few transitions to democracy at high levels of income and based on this infer that income does not cause such transitions, and (II) their sample is subject to selection problems because it does not include a long-run perspective. In 1950, when their observations starts the distribution of regimes was highly correlated with income and thus not random, according to Boix & Stokes (2003). Based on this reasoning, Boix & Stokes (2003) extended their sample to 1800-1949, in sharp contrast to the approaches discussed above. Their last appeal (III) is that the analysis made by Przeworski et al. (2000) suffers from omitted variables bias. In Boix & Stokes (2003), they show that controlling for exogenous effects of international political and factor endowments, economic development makes democratisation more likely.

The conclusion from Boix & Stokes (2003) is that income does not cause democracy per se, but that higher income forces other changes that accompany this change, in particular, income equality. They accordingly offer an explanation of the connection between development and democracy: As countries develop, income becomes more equally distributed, and the redistributive scheme that would win democratic, i.e. support from the median voter, would deprive the rich of less income than the one the median voter would support if income distribution were highly unequal. Hence, the wealthy class find a democratic tax structure to be less expensive for them as their country gets wealthier, and thus being more willing to accept democratisation. Although not defending modernisation theory, per se, Boix & Stokes (2003) provides a causal account for the positive relationship between income and democracy.

Following this, Hadenius & Teorell (2005b) as well as Coppedge (2003) supports the conclusion of Boix & Stokes (2003), although they give a general critique of the usage of the dichotomous dependent variable. In their analysis, Hadenius & Teorell (2005b) use a graded conceptualisation of democracy; looking at both developed and undeveloped countries from 1975-2001, and they also introduce a lagged democracy effect (viz. a lagged autoregressive term, AR1): This lag of the dependent variable at *t*-1 in the model is intended to control for the expectation that "democracy at an earlier time-point may have generated higher levels of economic development" (Hadenius & Teorell 2005b: 99). The coefficient of the logged and

lagged income term remains statistically significant when controlling for this, and leads to the conclusion that a high level of income per capita is conducive to higher levels of democracy.

The authors stress, nonetheless, that the effect of income on democracy is dependent on context and that the effect increases as countries more towards democracy. Yet, on the other hand, countries that fulfil all the basic criteria of democracy and have reached the democratic peak cannot move towards more democracy. Their result suggests that the less economically developed a country, the further it will move away from democracy and towards authoritarian rule. They conclude that Przeworski et al. (2000) are wrong when dismissing the modernisation theory, but right when it comes to the fact that economic development helps prevent authoritarian backslides. An interesting remark here might also be that Hadenius & Teorell (2005b) found this effect in the short-run.

More recently, both Boix (2009) and Boix (2011) find evidence of a positive effect of income on democratisation over a long period. The findings are robust to the inclusion of both country and year fixed effects, and utilise an AR1 term that suggest that development has a causal effect on democracy. Boix (2011) argues that income has a declining marginal impact on democratisation: In wealthy countries, any additional growth stabilises democracies, but does not increase the likelihood of a transition to democracy. Furthermore, the effect of income is strongly mediated by the structure of the international order and the ways in which great powers shape the resources of political fractions in small countries, a point which is also supported by Gleditsch & Ward (2006). Benhabib, Corvalan & Spiegel (2011) also recover a significant effect of income even when controlling for fixed effects, in contrast to the findings of Acemoglu et al. (2005); Acemoglu et al. (2008); Acemoglu, Johnson, Robinson & Yared (2009) which will be taken into account beneath.

Furthermore, Treisman (2011) demonstrates that there is evidence of a positive effect of income on democracy, more strongly over the medium run than over the short run. Gundlach & Paldam (2012), on the other hand, found that the relationship between democracy and economic development ran from the latter to the former. In other contemporary studies, Bobba & Coviello (2007) isolated a significant effect of income using a system-GMM (Generalised method of moments) estimator, which can account for possible simultaneity in an approach more sophisticated than the AR1 method.[10] The technique provides a computationally convenient method for estimating the parameters of statistical models based on the information in *population moment conditions*. Its popularity, especially

[10] See Arellano & Bond (1991) and Arellano & Bover (1995) for dicsussion.

in economics, is due to the fact that competing theories often imply that variables satisfy different sets of population moment conditions. The specific form of these population moment conditions depends on the context, but the generic form of the GMM estimator is the same in each case.

This method has grown in popularity in recent years.[11] Its popularity is mostly due to the fact that there are several advantages with this method: The estimator handles important modelling concerns such as fixed effects and endogeneity, while avoiding dynamic panel bias (Nickell 1981). The flexible GMM framework accommodates unbalanced panels and multiple endogenous variables, but there are major sources of bias if the instruments are misspecified (Bobba & Coviello 2007; Roodman 2009). An instrument is an explanatory variable that captures and identifies the causal effect of, in our case, income on democracy. Below, I give a specific example of how this is utilised in the literature. The same problem also arises in regular instrumental variable (IV) approaches, which, like GMM, are designed to account for some of the same problems (Roodman 2009). In modern political science, this is widely used: Murtin & Wacziarg (2014) find strong support for modernisation theory in their survey of 74 countries from 1870-2000, for instance. They also use system-GMM to address the issue of endogeneity and find robust evidence of an effect of democracy on income. In their analysis they incorporate recent advances in dynamic panel data methods, and address issues of weak instruments, instrument proliferation and violations of stationary assumptions that have been found to be in the literature and may therefore be a problem in the meta-analysis as well (Bobba & Coviello 2007).

To conclude this section, we see that the modernisation hypothesis has spawned a great deal of publications that claim income per capita is linked to democracy in a positive way. Although it seems like modernisation theory was studied using inappropriate methodological tools in its early phase, scholars have been able to reproduce the results with the appropriate methods. There are good arguments for this theory, as we have seen, but some academics remain unconvinced, and I will now proceed to summarise the most cited and relevant opponents of the modernisation theory.

[11] See Roodman (2009: 136) for an illustrative graph.

## 2.2 Conflicting views – Endogeneity

In this section, I will draw the lines of the main conflicting views in the field. I now move back in time, and trace the development in democratisation research from the perspective of the sceptics: Although Lipset's (1959) observation that democracy is related to economic development has generated the largest body of research on any topic in comparative politics, the conclusions, neither theory nor the facts, are clear (Przeworski et al. 2000). In an influential analysis, O'Donnell (1963) argues that at a certain stage of economic development in contemporary Latin America, further development produces not democracy but "bureaucratic-authoritarian" dictatorship. Diamond (1992) advocates a reconsidered version of the modernisation theory where the income per capita may be a surrogate for a broader measure of average human development and well-being that is in fact even more closely associated with democracy.

As we have seen, Przeworski et al. (2000) find that income does not cause transitions to democracy, but does hinder autocratic reversals. To investigate whether income causes *transition* to democracy, Przeworski et al. (2000) start with the observation that democracies are much more frequently observed in developed countries, and dictatorships in poor ones, although they fail to detect any threshold of development that would make the emergence of democracy predictable. Arguably, economic development may open the possibility for transition to democracy, even when the conditions for democracy are ripe and the outcomes of political conflicts are inseminated. They conclude that, "in sum, modernisation theory appears to have little, if any, explanatory power" (Przeworski et al. 2000: 80). The same pattern is found in Teorell (2010) and Persson & Tabellini (2009), yet also criticised (Boix & Stokes 2003; Epstein et al. 2006) on empirical grounds, but mostly based on their dichotomous measure of democracy.

### 2.2.1 The problem with endogeneity

In later work, Przeworski (2010) writes that identifying causality might be impossible, because there are no "primary causes" to drive history. In his view, this means that *everything* is endogenous, i.e. simultaneous. "Ideational, economic and political life evolved together, without even trying to detect which were causes and which the effects" (Przeworski 2010: XV). However, some scholars has set forth doing just that. Reviewing democracy in an endogenous perspective, Robinson (2006) argues that from an empirical point of view, the literature on democracy has only just "begun to wake up to the issue of the identification of

causal relationships" (Robinson 2006: 524). He states that, from Lipset (1959) to Przeworski et al. (2000), the scholarly tradition suffers from a *post hoc ergo propter hoc* fallacy. Scholars have paid little attention to the basic fact that the variables income per capita and democracy were jointly determined in equilibrium. These two variables are more likely correlated because the same factors that tend to make a society prosperous also tend to make it democratic, which in turn suggest that there might be omitted variable bias in the literature. Many aspects of the institutions and organisation of society maybe even culture and geography, will help to determine its prosperity and levels of democracy. These factors will be unobserved, and thus omitted from the explanatory models scholars estimate. One way to solve these problems is the inclusion of proper instruments. An instrument is an explanatory variable that captures and identifies the causal effect of income on democracy, as noted, and will, in turn, push the focus of the research towards endogeneity and causality. He also argues that scholars must "search for natural experiments" (Robinson 2006: 519-520), also known as real-world situation that mimic controlled experiments where there is a clearly exogenous source of variation. This is, however, a contested stance that I will discuss below.

Further, the majority of empirical work by Lipset (1959); Barro (1999); Przeworski et al. (2000); and Boix & Stokes (2003) either ignores the fact that income per capita might be endogenous, or attempts to instrument for it using lagged values of income per capita. Yet, if there are omitted variables that are present over time, such as the institutional organisation of society, for instance, then these variables will be correlated with lagged income as well, and this will not be a valid instrument. Furthermore, the estimation strategies, be it OLS or dynamic probit models, not only fail to account for the fact that income is endogenous, but also fail to isolate whether any estimated relationship between income and democracy is being driven by the cross-sectional variation or the time-series variation. Put differently: Is the question whether *richer countries are more democratic*, or as *countries become richer, do they become more democratic?* Answering this question is highly relevant for determining if the estimated relationship is likely to be causal or not (Robinson 2006). A simple and appealing strategy for looking at the latter issue is to introduce fixed effects into the model, as demonstrated by Acemoglu et al. (2005). They show that introducing fixed effects into models robustly removed any relationship between income and democracy, which will be discussed below.

Additionally, Acemoglu et al. (2005) discussed the use of an instrumental-variable strategy for estimating the causal effect of income on democracy, and thus dealing with the problem of endogeneity statistically. This is not a new idea, as Bollen & Jackman (1995) also

estimated a model of this kind using two-stage leas squares, but "they never discuss the exclusion restriction they make or whether they are plausible" (Robinson 2006: 521). Acemoglu et al. (2005), however, experiment with two potential instruments: (I) Past savings rate, and (II) changes in income of trading partners. The latter instrument is created by a matrix of trade share and constructs predicted income for each country using a trade-share-weighted average income of other countries. They show that this predicted income has considerable explanatory power for income, and argue that it should have no direct effect on democracy.

Both of these instrumental-variable techniques show no evidence of a causal relationship between income and democracy, and once income is instrumented to allow for its potential endogeneity, its estimated coefficient is indistinguishable from zero. These results, along with the evidence from the fixed-effects regression, suggest that the omitted variable bias is a probable explanation for the correlation between income and democracy. Yet, neither of these instruments are perfect, as Robinson (2006: 522) notes:

> […] [S]aving rates might be correlated with future anticipated regime changes; or democracy scores of a country's trading partners, which are correlated with their income levels, might have a direct effect on its democracy.

Nonetheless, Robinson (2006) argues that this moves the scholarly literature in the right direction.

In a subsequent article Acemoglu et al. (2008) take this approach further. Firstly, they focus on the problem of potential endogeneity. Secondly, they hypothesise that there is a potential omitted variable bias in the literature. Some other factors may determine both the nature of the political regime and the potential economic growth. To investigate the first hypothesis, the authors utilise two separate methodological strategies: The first is to control for country-specific factors effecting both income and democracy by including country fixed effects; the second strategy is to use instrumental variables (IV) regression to estimate the impact of income on democracy, as discussed above.

Clearly, since there is a cross-sectional relationship between the level of income and the level of democracy, at some point in time a relationship between growth rates must emerge. This assumption gives scholars a justification to hypothesise a relationship in the long run, looking at democracy from the 1500s to today. If the time-line is long enough, we can see that countries whose income per capita increases tend to become more democratic,

and this is why we can observe a cross-sectional relationship in present day.[12] There are two ways to interpret this finding, according to Robinson (2006): The first is that there is a causal effect of income on democracy, but it works over much longer time periods. It remains unclear, however, if this effect works through transition to democracy, or away from it. The second interpretation is that there is no causal effect at all, but over long periods of time, income and democracy are positively correlated because countries move along development paths where these two variables are jointly determined. The first view is, of course, a long-run version of the modernisation hypothesis, whilst the other focuses on how institutions of societies diverge as the result of historical critical junctures.

In the social science literature, the modernisation hypothesis has been very influential, as we have seen. But, according to Acemoglu et al. (2009) the critical juncture thesis is also used for reasons discussed above. A common understanding is that a juncture is critical because is places institutional arrangements on paths or trajectories, which are then very difficult to alter (Pierson 2004). Path dependence is a crucial causal mechanism for historical institutionalists, and critical junctures constitute the starting point for many path-dependent processes (Collier & Collier 1991; Mahoney 2001; Capoccia & Kelemen 2007). Moore (1966) demonstrates for instance, how the critical juncture thesis is used to describe regime transitions in England, Russia and Germany.[13]

In a subsequent study, Acemoglu et al. (2009) demonstrate that that the modernisation hypothesis is much weaker than some earlier work has found, as discussed above. They instead find support for the existence and importance of critical junctures. By introducing fixed effects to standard regression, the positive relationship between income and the level of democracy as well as the transition to and from democracy disappears. This is robust in both the sample period from 1960-2000 (which is the most commonly used) as well as the period 1875-2000. Furthermore, if income does not cause democracy, what should then be used as a primary explanation? The fact that including fixed effects removes the correlations between income and democracy leads Acemoglu et al. (2009) to conclude that relatively time-invariant, possibly historical factors are at the root of both the relative prosperity and the relative democratic experience of some countries. They therefore include historical variables, such as several colonisation variables, path of development, etc., in a pooled cross-sectional regression, and check if this removes the statistically significant association between income and democracy. They also use generalised method of moments-models (GMM) and

[12] See Figure 8 in Robinson (2006).
[13] For other macro-historical accounts, see also Rokkan & Lipset (1967) and Collier & Collier (1991).

instrumental variable (IV) approaches, as well as logit and hazard models. They conclude that although there is a positive correlation there are no significant causal effect of income and democracy. Omitted, historical factors appear to have shaped the divergent political and economic development paths, leading to the positive association between economic performance and democracy.

## 2.3 Heterogeneity – Systematic differences between studies

Based on this brief review of the literature on the vast subject of democratisation it is clear that there is a wide variety of possible explanations that are both theoretically and empirically plausible. It therefore seems justifiable to conclude that there is a large degree of heterogeneity in the literature. In order to answer whether income affects democracy ($Q_1$), we have to control for this observed heterogeneity ($Q_{1b}$). This will be done statistically using multiple meta-regression analysis and will let us, not only, address the two questions mentioned above, but also why we observe no consensus in the literature ($Q_2$). In order to explain this heterogeneity, the following characteristics listed below can be considered as theoretically important, explaining the difference between studies. Furthermore, it is precisely these features which are used as the *independent variables in the multiple meta-regression analysis*. The different attributes that are theoretically interesting to investigate fall into the following four categories: (I) Estimation differences: Do the different statistical methods create systematic variation in the literature? (II) Time and region differences: How does the different composition of the samples, in terms of countries included and time-perspective studied, affect the literature? (III) Knowledge-related differences: Do different attributes of the different authors and studies create systematic variation? (IV) Specification differences: Several measures of both the dependent variable (democracy) and several data sets for the independent variable (income) are available. Does the choice of measure have an effect on the study outcome? Many studies also control for different confounder variables, like oil, education, population, etc. The meta-analysis will investigate whether there is an omitted variable bias in the literature in this respect as well.

### 2.3.1 Estimation differences: Type of statistical technique

As discussed above, an important source of heterogeneity might be the estimation techniques. Although some scholars dismiss regressions using observational data entirely, relying solely on natural experiments (Summers 1991; Rodrik 2005; Seawright 2007; Sekhon 2009), fact remains that this is a somewhat extreme stance. The lack of external validity which this approach brings forth makes it problematic, especially in the democratisation research context, making these approaches ungeneralisable outside these small samples. Quantitative methodologists are well aware of the limitations that the Gauss-Markov theorem brings forth, and constantly try to diagnose problems and devise solutions. This gives us a rich diversity of econometric techniques applied in social science, and in this field of research especially.

The most basic choice made possible by the TSCS data sets is the differentiation of between-country and within-country estimates. Those who estimated between-country (random) effects approach tests the hypothesis that the higher the average income a country has compared to other countries, the higher its average level of democracy. This bears striking resemblance to the hypothesis estimated with cross-sectional data, and often find the same results (Coppedge 2012). The effects of income on democracy would be observable in the long term and across countries, but would not necessarily manifest itself in the short term, within countries. This means that the estimation technique focuses on explaining differences between countries rather than changes within countries. In contrast, the within-countries (fixed) effects test a radically different hypothesis. Namely that the richer a country is this year compared to itself in other years, the more democratic it is relative to itself in other years. In short: Small changes in national income cause small, immediate changes in the level of democracy. The estimators of this relationship transform the data so that each country's deviation from its mean income explains each country's deviation from its mean democracy score. The published results from these approaches are also quite different. Between-country estimates often show a significant positive effect of income on democracy (Coppedge 2012), but within-country estimates are far from conclusive. Some scholars using fixed effects have found a significant, positive effect (Brinks & Coppedge 2006), some a significant, negative effect (Rudra 2005), and some no significant effect at all (Colaresi & Thompson 2003). The most comprehensive tests report that with fixed effects, per capita income has no significant impact (Acemoglu et al. 2008). This will be addressed empirically in the meta-analysis.

Most of the studies published after year 2000 rely on TSCS analysis. Cross-sectional studies are not used in this literature anymore, as noted earlier. Since the studies included in

this analysis use a continuous measure of democracy, the standard methodological choice is OLS, but with several variations. There is a debate concerning methodology, but Beck & Katz (2001) argue that models with fixed effects, panel corrected standard errors, and lagged dependent variables (AR1) should be used. I therefore make the distinction between estimates from studies that adopt TSCS with a lagged independent variable and fixed effects, and those how do not. A second methodological variation is studies that treat democracy as an endogenous variable in a more rigorous way. Recall, however, the discussion of the AR1 approach above. Some therefore use estimation techniques like Generalised Method of Moments (GMM), Two- and Three-Steps Least Squares (2SLS, 3SLS), and other instrumental variable (IV) approaches. The meta-analysis will test whether estimation technique has an impact on the varying income-democracy effects found across the studies.

### 2.3.2 Countries and time

The regional focus of the studies may also be an important factor. Most studies use the whole world as a basis for the TSCS data, although the tendency in the early literature to provide detailed descriptions of which countries were included in the sample, this has been abandoned in the recent years, resulting in loss of data points in the multiple meta-regression analysis (Doucouliagos & Ulbasoglu 2008). It is possible, however, to identify four broad regional groupings: Africa, Asia, Latin America and the rest of the world (mainly the OECD). From this, I have separated estimates that include the OECD, i.e. developed countries, and the rest of the world, i.e. developing countries. Some studies also compose samples where former socialist countries are excluded. A similar approach is adopted to investigate democratisation over time. Most studies use the third wave of democracy as their primary focus, from 1950 up until today. Coppedge (2012) argues that this is practically the whole domain that one would want to use from democratisation studies because national, representative, full-adult-suffrage democracy barely existed before 1920 (Dahl 1989). Despite this, other studies look at democracy in the long run, and very long run, some dating back to the 1800s (Boix & Stokes 2003) or 1500s (Acemoglu et al. 2008). However, extending the sample to years before 1950 radically restricts the variety of independent variables that can be tested. I therefore test if estimates from regression using the long-run perspective systematically find different results than studies focusing of the third wave of democratisation. Some studies also pool their observations on a yearly basis, as opposed to a 5-year, 10-year, 20-year, 25-year and 50-year

periods. Whether this creates excess variation in the sample of studies will be addressed through a multiple meta-regression analysis.

### 2.3.3 Primary focus, discipline bias, knowledge, and publication effects

The main focus of this thesis is the relationship between income and democracy. Yet, some studies included in this meta-analysis focus on other factors affecting democracy, and use merely income as a control variable. Some studies may use the relationship between income and democracy as a point of departure, while other studies have taken other effects into account, be it inequality, conflict, natural resources, etc. Take for example the difference between Acemoglu et al. (2008); and Ansell & Samuels (2010), where the former is an example of a primary focus-study, and the latter is an example of a study that does not focus primarily on the income-democracy nexus, but rather inequality and democracy. This analysis will investigate if the scholars that focus specifically on the income-democracy nexus produce different results than scholars who do not.

Most of the studies carried out in this field are produced by either political scientists or economists, wherein about 57% of the individual regression estimates collected in the sample are produced by political scientist while the rest are produced by economists. As we have seen above, Acemoglu et al. (2009) argued that political scientists tend to favour the modernisation hypothesis, whilst the critical juncture theory was introduced as a competing theory from scholars with an economic background. In a meta-analysis in the context of globalisation on taxation, Adam, Kammas & Lagou (2013) suggest that there is a discipline bias between political scientists and economists.[14] One might argue that political scientists tend to favour the importance of the political procedures whereas economists tend to favour explanations based on market driving forces. Therefore authors from different discipline can produce different results on the outcome, due to scholarly focus and theoretical considerations. In this study, I will investigate if this is also the case in the research on democratisation.

Following the meta-studies of for instance Ahmadov (2014) and Doucouliagos & Ulbasoglu (2008), I will investigate whether the study is published, or not. This is a way to accommodate possible publication bias empirically (Borenstein 2009), which will be discussed below. Yet, there are many other ways in which this could be captured. It could be hypothesised that there are attributes with the specific journal that could be relevant (certain

[14] Kallager (2014) also show that political scientist tend to publish different results than economists in the research on globalisation effect on the welfare state.

journals might be biased towards publishing certain types of results, for example). In this case, I have focused on the differences between economists and unpublished work to try to capture this. Likewise, if the author has published in this field earlier, I will investigate if this has an effect on study outcome. It is possible to argue that researchers that have published previous work might be forced to be consistent with what they publish in order to maintain their integrity. The year of publication, with 1999 as the base category, is included to capture any time-specific trends in the literature. We could assume that a certain method or research design was particularly popular at some given point in time, and this variable would capture if time of publication has an effect on study-outcome. Some meta-analyses also focus on socio-economic variables, such as gender, nationality, funding, etc. and hypothesise that these could have an effect on the results they produce (Stanley 2008). In the field of democratisation research these factors are not seemingly important, and are therefore excluded from the present study.

### 2.3.4 Specification differences: Measurement, conceptualisation and aggregation

In the literature there are several ways to operationalise democracy within two main categories: Dichotomous or continuous. In the latter category, there are two common approaches to measures democracy in the field. Freedom House, a continuous measure with a focus on political rights and civil liberties; and Polity, a continuous measure focusing on the competitiveness and openness of elections, the nature of political participation, and the extent of checks on executive authority. Other continuous measures include Vanhanen (2000), Hadenius & Teorell (2005a) etc. In the dichotomous category, we find the ACLP (Przeworski et al. 2000) and BMR (Boix, Miller & Rosato 2013) approach. These are based on a minimalist definition of democracy and utilise a dichotomous democracy-dictatorship measure. Theoretically, there are good arguments for all measures and most studies include several to ensure robustness. The choice of measure also has consequences for the methodological possibilities when it comes to the statistical estimation technique. In our case, all collected estimates from the studies use either Polity or Freedom House. About 55% of the estimated regression coefficients use Polity and about 45% use Freedom House.[15] Both measures are thus most common in the literature, and I will now discuss some of the different attributes of these measures. But first I will make some general remarks on measurement in democratisation research.

[15] As discussed below, this may very well be a result of availability bias.

As an underlying understanding of the concept of democracy, there are three major distinctions that dominate modern political though when it comes to the forms of government, according to Przeworski et al. (2000): In Montesquieu's locus classicus *L'esprit des Lois* (2011 [1748]) there is a distinction between limited regimes and despotic regimes. In Kelsen's (1945) view, which can be traced back to Rousseau and Kant, he distinguishes between autonomy and heteronomy. Lastly, there is the Schumpeter's (2006 [1942]) innovation, which was to emphasise competition, or, in Dahl's (1971) terms: *Contestation*.

Keeping this in mind, Teorell (2010) identifies some mainstream approaches in the field of comparative politics wherein democracy is defined by a few basic criteria. Amongst these we find the holding of periodic, free, fair and effective elections to the legislative and/or chief executive offices of state, together with a bundle of continually upheld political rights, most notably freedom of association and opinion As a consequence of this, autocracy or authoritarianism means the absence of these basic characteristics. These basic-level characteristics are commonly agreed upon (O'Donnell 2001), but beyond this there is profound disagreement over the meaning of democracy. Different adjectives have been attributed to the democracy concept to highlight the different nuances in the concepts (see Andersen 2012; Boix 2011; Glaeser, Giacomo & Shleifer 2007; Shafiq 2010).

The question whether democracy is a graded concept, i.e. a property of which one can have more or less, or if it is an either/or phenomenon is very much debated in the literature (Adcock & Collier 2001). One of the most favoured dichotomous versions is the one of Przeworski et al. (2000) (Teorell 2010). Their specific coding rules, though contested, are based upon on a necessary and sufficient definition of democracy where a regime in which those who govern are selected through contested elections where the opposition has some chance of winning. The main attribute with these coding rules is that they are based upon an objective criterion, and do not leave room from a subjective evaluation of the regime in question. In some cases, this has proven problematic to keep the objectivity intact.[16] Critics of this measure also emphasise that "some chance of winning office" is vague, because it also requires an ambiguous cut-off point with regard to the probability of regime change (Teorell 2010). Since this measure is not included in the analysis, I refrain from conducting a further discussion of this concept in light of the continuous measures reviewed above.[17]

[16] Consult Knutsen & Wig (2015) for discussion.

[17] Of course this creates a possible bias in the analysis, because the entire literature is no captured by the data gathering process. While collecting and coding studies, I found about 15% of regression estimates from the studies collected used a dichotomous measure of democracy as the dependent variable, and chose thus to focus on democracy as a graded concept to exploit the full potential of the multiple meta-regression analysis.

The main focus of this thesis is, as noted, democracy as a graded concept, and we therefore continue with a discussion of the most common measures of democracy with a continuous conceptualisation. The ordinal scale measures for democracy also specify a set of criteria for judging regimes, but unlike dichotomous measure, they assume that democracy is continuous and provide scales that range from low to high level of democracy. The vast majority of the studies included in this meta-analysis are what Coppedge (2012: 33) call "data-driven research". This implies that data availability is the most important factor that determines which variables are chosen to measure what, and since time-series cross-sectional data analysis became the gold-standard of political science research (Beck & Katz 2001), the availability of the Polity and Freedom House data has dictated the direction of the research.

The Polity data series takes into account both the democratic and autocratic features of countries, while its combined score on democracy ranges from -10 for a full autocracy and to +10 for full democracy (Jaggers & Gurr 1995), though often normalised to ease interpretation in regression analysis. The indicator is based on five attributes concerning participation and contestation, subscribing to a minimalist conceptualisation of democracy. The Freedom House Index (FHI) is a continuous scale where a country receives the highest score if political rights come closest to the ideals suggested by a checklist of 25 questions, thus being a maximalist definition. The checklist asks whether there are free and fair election, whether those who are elected actually rule, if there are competitive parties, whether the opposition plays an important role and has actual power and whether minority groups have reasonable self-government or can participate in the government through informal consensus, amongst many other (Munck & Verkuilen 2002). The Freedom House Index (FHI) has two separate scales from political and civil liberties that range from 1 (full enjoyment of liberties) to 7 (full restriction of liberties), which have often featured in cross-national comparisons in some combined from as a measure of democracy. FHI measure this on two dimensions: political rights (PR) and civil liberties (CL).[18] The FHI is an average of PL and CL. Both PR and CL are indices constructed from large subset of indicators, which are formulated as control questions. Freedom House scores its indices on the basis of 25 such questions, each having several sub-questions. Ten of the check questions are related to political rights and 15 are related to civil liberties. The index is based on the presence, or absence, of different institutions, but it also seeks to account for the actual performance of these institutions.

[18] Some studies differentiate these two measures and run two separate regressions for PR and CL. In this meta-analysis this is treated like the study used Freedom House as the dependent variable equal to those who use the FHI (the average value of PL and CL).

"Therefore, this operationalization is presumably valid for substantive democracy definitions" (Knutsen 2010: 112).

While both Polity and FHI measure variation in democracy, criticism has focused on their less-than-transparent coding rules (especially Freedom House). They have, according to critics, an illogical form of aggregation into single indices that does not take into account trade-offs between the institutional dimensions and political rights dimensions, their inability to differentiate the democratic performance of those countries at the extreme ends of the spectrum, i.e. among mature democracy and highly authoritarian regimes, and the possible presence of ideological biases (Freedom House in particular) (Munck & Verkuilen 2002).

The key advantage with the Polity measure is that it clearly identifies what attributes it seeks to measure, namely offices and agenda setting. They also have clear and detailed coding rules, tests of inter-coder reliability and a comprehensive empirical scope. According to Munck & Verkuilen (2002), the Freedom House Index shares only the latter strength. Weaknesses with the Polity measure includes that it omits the aspect of participation. Problems of redundancy are also present. This refers to the fact that attributes at the same level of abstraction should measure mutually exclusive aspects of the immediately superior level of abstraction.[19] It also includes an inappropriate aggregation procedure (this implies that the theoretical link between attributes that are at the same level of abstraction are connected to the same overreaching attribute). In the case of Freedom House, the aggregation rule is "clear and explicit: Scores for the two attributes – political rights and civil rights – are generated by adding up the scores assigned to each of its respective components" (Munck & Verkuilen 2002: 25). Polity, on the other hand, has five attributes that are weighted differently by using different scales. Weaknesses with Freedom House are the maximalist definition, problems of conflation, measurement problems and an inappropriate aggregation procedure as well (Munck & Verkuilen 2002).

There continues to be a substantial debate over the different measures in the literature, but there seems to be no consensus on which measure is to be considered the most accurate. This thesis will not propose a standard for measuring democracy, but rather will empirically address the implication of choosing one or the other. Thus, it can be hypothesised that the choice of the dependent variable in primary studies will be of importance for the inference the authors can draw.

[19] See figure 1 in Munck et al (2001: 13) for the logical structure of concepts. See also Goertz (2006) for discussion.

### 2.3.5 Income data

The main independent variable in the studies summarised above is income per capita. This is the most frequent measure of modernisation in the literature, and all studies in the sample used in the meta-analysis utilise this conceptualisation of income. What might be interesting to investigate, however, might be the different datasets used to measure the GDP. In his article, Krieckhaus (2004) argues that it makes a difference whether World Bank data or Penn World Tables data are used when studying democracy and growth. Albeit this is not the focus of this study, the data used overlap and growth is a product derived from income per capita over time. The data sources used here are World Bank Data (WBD), Penn World Tables (PWT), and some also use Maddison (2003). Boix (2011), for instance, uses Maddison when studying pre 1950 and Penn World data when studying post 1950. Krieckhaus (2004) finds that differences in macroeconomic data do matter, and his findings, sobering as they may be, imply that we can call into question the "literature's implicit assumption that the choice of macroeconomic dataset is not problematic" (Krieckhaus 2004: 651). In further investigations he finds that influential cases, such as Iraq, included in the World Bank data is not included in the PWT data. Much of the difference between the two sets of findings was not due to the case of Iraq, however, suggesting that estimates of democracy's effect on growth is partially determined by which measure of growth is utilised. Levine & Renelt (1991) argue that the WBD is preferable, but recent scholarship has tended to employ the PWT data or Maddison, albeit usually without explicit justification (Krieckhaus 2004). Based on this discussion, I include this perspective of the independent variable as well in the meta-analytical framework.

### 2.3.6 Specification differences: Confounder variables

Some studies included in the meta-data use different control variables that income is thought to work through. Werger (2009) includes a measure for population, as do Acemoglu et al. (2005). In total, about 15% of study-estimates use this as a control variable. Many studies (about 21% of study-estimates in this sample) look at the relationship between oil and democracy, and income is here an important control variable. Ahmadov (2014) finds in his meta-analysis that income per capita has no significant effects, thus, oil is unlikely to hinder democracy by affecting income. He also writes that future meta-analysis on income per capita would "enable judging the substantive significance of the oil-democracy effect size in a comparative perspective" (Ahmadov 2014: 1263), and it is precisely that this thesis does. Further, 15% of study estimates also control for education and hypothesise an effect on

democracy (Che, Lu, Tao & Wang 2013; Cornell 2012; Bobba & Coviello 2007). The implications of this will be examined through the multiple meta-analysis.[20]

## 2.4 Summary

In this chapter a variety of empirical studies have been examined, and it seems sensible to conclude, after this brief survey, that the results found in the literature are conflicting. Most of the studies focus on the entire world, albeit some exclude developed countries and socialist countries. The time perspective is also mostly confined to 1960-2000, but with some variation. Some studies also look at data dating back to 1800s and 1500s. The measurement of democracy most frequently used is either Polity or Freedom House, which uses, respectively, a minimalist and maximalist definition of democracy. Regarding statistical estimation technique, we have seen that Beck & Katz (2001), amongst others, have advocated the use of TSCS with an autoregressive term (AR1), i.e. a lagged dependent variable, and fixed effects. Others claim that to address the possible endogeneity, instrumental variables (IV), GMM or 2SLS and alike must be utilised. Some frequent confounder variables will also be investigated.

Undertaking *yet another reassessment,* using either new or similar sets of data, specification and estimation techniques may or may not dissolve the confusion due to the theoretical preferences, data limitations and methodological choices. The *meta-analytic framework* and strategy applied below, on the other hand, capitalises on the existence of a sufficient amount of scholarship which thus permits a cautious integration and systematic examination within the integrated framework of the link between income and democracy, confounders that affect it and causal mechanisms, if any, through which this link works.

[20] Other controls are also frequent, but observations are too few (around 5% of study estimates) and the variables are assumed not to have an effect and thus omitted to prevent multicollinearity. These controls include age, conflict, religion, and inequality.

# 3. Methodology of meta-regression analysis

In this chapter I will discuss the use of a quantitative literature survey method, namely meta-analysis, or meta-regression analysis (MRA). There are several ways to apply such an analysis, but in this thesis I will argue that a meta-regression analysis is the most rigorous way to get an overview of the quantitative literature in this field of research. A meta-analysis can be understood in several different ways, which will be discussed below, but the meta-regression analysis is broadly defined as a set of statistical techniques that enable the summarising of a body of research with conflicting findings. A meta-regression analysis evaluates and explains carefully the between-study findings differences which stem from differences in samples, estimation technique and different specifications, i.e. study heterogeneity, as it is referred to in meta-analytical terms. This is a technique commonly used in the medical sciences, but increasingly more common in economics and social science.[21]

Not only does this technique allow the meta-analyst to access information buried in the pile of research in the income-democracy field ($Q_1$), the method can also address publication selection bias ($Q_{1a}$) and heterogeneity between studies ($Q_{1b}$). "Publication bias" can be defined in a number of different ways, but in short it refers to the potential problem of scholars publishing selected findings.[22] To sum up, a meta-analysis, and meta-regression analysis specifically, is the *analysis of other empirical, quantitative analyses, which attempts to integrate and explain the literature about some specific important phenomenon, address potential publication bias and explain the systematic differences between studies* (Stanley & Jarrell 1998: 301; Borenstein 2009: 358).

In this chapter I describe the statistical methods applied in this thesis: Firstly, I will discuss the meta-analysis and distinguish it from other forms of literature reviews, quantitative or otherwise. Also, I discuss the justification of the use of this quantitative

[21] For examples, see: Stanley & Jarrell 1998, Doucouliagos & Laroche 2003, Stanley 2004, Doucouliagos & Ulbasoglu 2008d, Doucouliagos & Paldam 2008a, Feld & Heckemeyer 2011, Stanley & Doucouliagos 2012, Ahmadov 2014). See also Figures 1.1. and 1.2. in Stanley et al. (2012) for a graphical display of the numbers of meta-analysis in economics over time.

[22] This is a term that describes the simple assumption that researchers are more prone to get, say, a positive finding accepted and published by scientific journals if the majority of the literature that is published finds the same positive effect. This implies that insignificant findings, or in some cases negative findings, are put in the drawer, in Rosenthal's (1979) wording. The consequence of this is that a field of study could be skewed towards finding a positive effect and one would then have to search even harder to find the negative or non-findings. Hunter & Schmidt (2004) refer to this as availability bias, while Rosenthal (1979) calls it the "file drawer problem". Stanley et al. (2012) suggest that the problem more accurately should be called "selective reporting bias", but also advocate the term "publication selection bias" or just "publication bias". The latter term is most frequently used in this thesis, but all refer to the same concept.

technique applied to an inconclusive literature, as opposed to a traditional literary review or research synthesis. Secondly, I discuss the process behind the collection of the data, calculations of the dependent and independent variables needed to conduct a meta-regression analysis, and how to summarise the relevant data in an effective and rigorous way. Thirdly, I address how the issue of publication bias can be inspected through a "funnel plot", where I investigate the distribution of negative and positive findings in a simple scatter plot. As I will explain in detail below, the assumption in such a plot is that the study estimates in total should appear as an inverted funnel in the graph when there is no publication bias, and skewed towards positive or negative findings when publication bias is present.

After an introduction to the theory behind the inspection of the funnel plot, I then discuss a statistical test constructed to ensure that the examination of the funnel plot is robust and unbiased. The same statistical method can also be used to test whether there is a genuine effect of income on democracy through the potential publication bias. Lastly, I introduce the multiple meta-regression analysis and discuss this technique in both statistical and practical terms. In a multiple meta-regression analysis, the researcher is able to take account of both publication bias and the systematic difference between studies to determine how methodological, sample and other specifications contribute to the lack of consensus in this field of research ($Q_2$). The multiple meta-regression analysis is consequently the full extension of meta-regression analysis.

## 3.1 Meta-analysis, meta-regression analysis, and research synthesis

The publication of findings in the sub-discipline of democratisation research is vast, and the piles of knowledge and data might sometimes seem incomprehensible. Therefore, the review of the research seems of utmost importance in the search for a sound conclusion. To do this, scholars often conduct a literary review, or research synthesis, to summarise previous research. This is done to determine what we actually know about some phenomenon, and to guide further research in the right direction based on prior experience. Before we discuss the chosen approach for this thesis further, we must consider the several ways in which this can be done. A distinction is made between literary review, research synthesis[23] and lastly – the chosen method of this thesis – meta-regression analysis. Literary review, being the broadest term, typically appears in independent scholarly works or as a brief introduction before the

[23] Also referred to it as a *narrative review* in the literature (Borenstein et al. 2009, Hunter et al. 2004, Stanley 2001, Stanley & Jarrel, 1989).

researcher reports new findings. A research synthesis, on the other hand, is the type of literary review that focuses on empirical studies only, and seeks to summarise past research by drawing overall conclusions form many separate studies that address related, or identical hypothesis. The distinction that makes a research synthesis stand out from a traditional literary review is just this: That it draws conclusions based in prior research and does not present new (empirical) evidence. The ultimate goal is thus to present the state of knowledge concerning the relation in question and highlight the questions that previous research has left unresolved. A meta-analysis is referred to as purely a quantitative synonym for a research synthesis and is used because there are some apparent limitations with the research synthesis as a way of summarising research (Cooper 2010).

The main problem with the traditional research synthesis is that it does not hold the methodological standards of current political science in terms of reproducibility, objectivity, unbiased selection and validity (Stanley 1989; Stanley 2001; Hunter & Schmidt 2004; Stanley 2005; Stanley & Doucouliagos 2012). This is because a research synthesis "relies on the researchers ability to digest the array of findings across studies and arrive at a pronouncement regarding the evidence using some unknown, and unknowable mental calculus" (Wilson 2001: 73). Furthermore, the research synthesis "impressionistically chooses which studies to include, what weights to attach to the results, and which factors that are responsible for the differences" (Stanley 1989: 300), resulting in subjective judgements of study quality.

Hence, there are several problems with this approach: Firstly, methodological selection bias often occurs (Stanley 2001). This implies that researchers might omit studies that are not in line with a preferred hypothesis. The main problem is not that a scholar dismisses a study for being misspecified, for that may indeed be quite justified. However, if such a judgement is made in a subjective and selective manner for the purpose of omitting unwelcome empirical evidence, it will cause severe bias in the conclusions drawn from this survey of the literature. Secondly, a review of this sort cannot account for the possibility of publication bias in the existing literature, which seems to be a regarded as a problem in most disciplines of empirical research (Stanley 2008).

To account for these problems scholars have suggested several other methods for conducting these types of surveys.[24] In this thesis, however, I will use a meta-regression analysis to summarise and evaluate studies from previous research done on the relationship between democracy and income. The meta-regression analysis is a statistical method that has

[24] See Borenstein et al. (2009), Hunter et al. (2004), and Cooper (2010) for other methods.

been found useful in doing just this (Stanley 1989; Stanley 2001). A meta-analysis is often defined as an analysis of analyses that aims not just to review a given set of studies in a respective topic descriptively, but rather it tries to analyse and test results from previous studies, and summarise its findings. This approach is frequently used in the medical sciences to determine whether a specific treatment and/or drug have the desired effect, as mentioned above (Smets & van Ham 2013). In medicine, as well as in social science, the method allows the analyst to integrate results from different studies and isolate the individual artefacts of the independent study, such as between-study data, specification and estimation differences. Consequently, the meta-analysis can yield a more powerful test of the effect than any single study can achieve independently (Borenstein 2009). Once I isolate publication bias and heterogeneity, I can draw several solid conclusions about the relationship between income and democracy. More specifically, the meta-analysis is used to eliminate the effects of sampling error, specification choices, methodological choices and publication bias, and then removes these effects from the empirical findings (Hunter & Schmidt 2004; Borenstein 2009).

In response to the critique of the traditional literature review, the meta-regression analysis is more precise because, firstly, it does not *a priori* judge studies based on their quality. It is thus more objective than the traditional literary review since studies included are solely based on a statistical measure of quality, which will be discussed below. We will therefore not encounter the problem of selection bias described above because the selection of studies included is designed to mimic a random sample. The meta-analysis is also able to address all studies related to the research question, and in that sense, be more wide-ranging than the traditional approach. This is because the meta-analyst strives to be as inclusive as possible when searching the literature for studies to include in the analysis. As noted, not only can meta-analysis identify publication bias, but it can also explain systematic variation across studies, and hence draw conclusions about the research design itself through the multiple meta-regression analysis. Lastly, the meta-analysis uses essentially the same econometric tools used by scholars that produce the empirical estimates in the literature on democratisation.

Like any other statistical tool, the meta-analysis can be misused, and it has its own weaknesses. Stanley (2001) identifies several of these weaknesses: First, in the context of the multiple meta-regression analysis, there can be disagreements on which study characteristics should be included in the analysis as independent variables. Secondly, a problem can arise if all studies contain the same misspecification. In this case, there is no way to distinguish or to estimate this common misspecification bias. Thirdly, meta-analyses are sometimes criticised

for considering all empirical studies, regardless of their quality. While valid, these criticisms are also present in the literature synthesis as well. The advantage of the statistical nature of the meta-analysis is that all of these choices made are explicitly addressed and their implications pursued statistically. The discussion of which study characteristics to include can therefore be examined statistically and compared in order to draw conclusions based on objective observations, whilst a literature synthesis cannot.

An extension of the critique posed above is that if the meta-analyst includes studies of poor quality alongside studies of good, this will result in a potential bias. As stated above, the inclusion of studies is not based on an *a priori* judgement of quality, but there are some different ways in which this could be done. The meta-analyst could for example search only highly ranked journals, since one could subjectively expect that studies from a highly ranked journal will have higher quality than a study published in a journal ranked low. Stanley & Doucouliagos (2012) argue in contrast that this quality check is inevitably objective. They propose to use statistical approaches to determine quality, or more specifically they suggest a statistical measure for the estimates precision, i.e. the standard error of the effect size.[25] This implies that each estimate collected from the studies is given a purely quantitative measure of how precise it is, rather than customising the collection of data to include only top-ranked studies.

Critiques have also argued that the meta-analysis compares "apples and oranges", meaning that the studies included in the analysis are too dissimilar for the researcher to draw sound conclusions. However, this will be the case in almost any review, or any study, for that matter. Card (2012) argues that including studies that are diverse in methodology, measures, and sample within a meta-analysis has the advantage of improving the generalisability of the conclusions. These differences can also be addressed in the multiple meta-regression analysis that further reduces this problem.

In conclusion, we can state that all of these problems are present in the traditional review of literature as well, as mentioned above. But the fact remains that the meta-regression analysis does a better job by addressing these problems explicitly. Many of the problems can also be handled through the statistical modelling in the multiple meta-regression analysis. Stanley (2008) argues that no more objective, comprehensive, or rigorous method exists for assessing and drawing useful inferences from a scholarly literature. After making the distinction between a more conventional, qualitative, literature review and the quantitative

[25] Specific calculation of the measure is discussed in full below.

approach of meta-regression analysis, I have argued that this is the best way in which to address an inconclusive literature. In the following section I will outline the process of data gathering and coding conducted to create the meta-data set. Afterwards, I discuss the calculation and the specific statistical methods used to complete such an analysis. Below the reporting guidelines of a meta-analysis are mapped out as suggested by Stanley et al. (2013). This is also the criterion upon which this thesis is based.

### 3.1.1 The steps of meta-regression analysis

The objective of this thesis is first and foremost, to establish whether there is an effect of income on democracy ($Q_1$), approaching this question via an understanding of how and why the existing literature has produced contrasting findings. In order to do this properly, we must control for publication bias ($Q_{1a}$) and heterogeneity ($Q_{1b}$) that might affect the literature, which in turn also allows us to answer why we observe no consensus on this question across the scholarly literature ($Q_2$). To test whether or not income affects democracy, I can use what Hunter & Schmidt (2004) call a "bare-bone meta-analysis" or what Feld & Heckemeyer (2011) call a "classical meta-analysis". The approach consists of calculation the average effect of income on democracy across studies, using different weighting techniques, which will be discussed in full below (Equation 1.3). There are some problems with the estimation techniques associated with these approaches, and results will be reported with caution, in accordance with Stanley & Doucouliagos (2012). Furthermore, we keep in mind that these averages do not suffice singlehandedly because it does not take possible publication bias and heterogeneity into account. Hence, I secondly want to examine whether publication bias is present and how this might affect the literature. Lastly, to realise the full potential of the method, I will model the systematic difference between studies, also referred to as heterogeneity. The latter part is examined using a multiple meta-regression analysis. In sum, I purpose the following steps for the analysis: (I) Search the relevant literature and selecting and coding of estimates; (II) summarising research, choose independent variables, and obtain the effect size; (III) accommodate publication bias; (IV) use meta-regression analysis to test beyond publication bias and to model heterogeneity; (V) guiding further research.

To summarise: The steps I-II are carried out investigating the pooled effect of per capita income on democracy across all collected studies. The term *effect size* is used here to describe the strength of the association between income and democracy. Below I will discuss how this can be used differently in different fields of study and how this specifically is

calculated and interpreted in this thesis. The independent variables are the individual characteristics of the study-estimates described in the chapter above. These steps lay out the foundation of the meta-regression analysis that follows in steps III-V. After searching the literature thoroughly, the researcher must collect all relevant data from the studies. This is the essential data such as the beta-coefficient, degrees of freedom and standard error or *t*-statistics. The next step is to obtain the effect size and perform other relevant calculations, which will be discussed in detail below. Following this, I will summarise the collected cross-study averages and time-series plots in order to inspect the development of the scholarly tradition. I will then test for publication bias both graphically and statistically.

Moving to the multiple meta-regression analysis, I will test for systematic variation across studies, i.e. heterogeneity, while controlling for potential publication bias detected above. This allows us to identify different features of the studies that result in systematic variation in the reported findings of the individual regression estimates collected from the studies. If the income-democracy relationship's individual effects have one parameter value, or several values distributed within a certain range, the meta-regression analysis is able to estimate these values once the individual study findings are isolated, and their individual artefacts are controlled for. In practice, this means that a binary control variable is included in the meta-regression analysis-model that control for any study-specific methodological or conceptual variation. It is also possible to assess the impact of confounder variables on the relationship between income and democracy and evaluate these relationships (Ahmadov 2014: 43-47).

The discussion that follows will examine the particular methods and calculations that are necessary to conduct a meta-regression analysis. The main contribution of such an analysis, as we have seen, is to contribute to making inference about the state of knowledge in a specific field, to detect and correct publication selection bias, and explain heterogeneity in empirical studies (Stanley & Doucouliagos 2012: 38).

## 3.2 Collection and coding of meta-analysis data

The overall objective of the searching of the literature is to obtain a representative, unbiased collection of studies that can be analysed through the methods of meta-analysis. The data in this meta-analysis represent the results and features of estimates of the relationship between income and democracy collected from the publicly available studies in English. They were obtained through a comprehensive review of *ISI Web of Knowledge, Google Scholar,*

*ProQuest, Jstor*, *Science Direct,* dissertations and personal correspondence with scholars conducting research in this field. The online search was conducted with a combination of the search words: "income", "democracy", "GPD", "income+democracy", "democracy+GDP", "income+democratisation" and "GDP+democratisation", which of course produced many results, some 1500 studies. Therefore, a more specific search was carried out to find econometric studies though *Citeweb* and *Web of Science*. The search was completed ultimo March 2015.

Regarding the content it is important to assess that the initial dataset only contains studies written in English, although it is recognisable that some studies of importance may be published in another language. Stanley & Doucouliagos (2012) note that exclusion of non-English studies might not be critical since most journals are written in English. In their survey of political science journals, Giles & Garand (2007) find that all the "top 90" journals are published in English. Therefore this exclusion will not be problematic for the overall meta-data sample. Furthermore, a clear understanding of the studies is of vital importance, and therefore an unqualified translation of the journal articles may cause confusion and lead to incorrect inferences.

After identifying a population of 33 studies, with 492 corresponding estimates, that could be meaningfully examined using the tools of meta-analysis, these studies were coded and included in the meta-data set. This implies that the studies used (at least) one measure of democracy as the dependent variable and (at least) one measure of income (GDP per capita). It is also critical that the necessary estimates are reported, such as standard errors, sample size, and coefficients that enable the construct of the dependent variable in the meta-analysis (Ahmadov 2014) (calculation and justification of the dependent and independent variables will be discussed below). The search identified several studies published as peer-reviewed journals, working papers, or books, and all were included in the meta-data set. This is in line with Borenstein's (2009) suggestion that inclusion of unpublished work is a way to explicitly address publication bias empirically, and to illustrate the difference between published studies and unpublished work. Yet, potential publication bias will also be addressed through statistical tests as a part of the meta-regression analysis, as mentioned earlier.

Furthermore, to ensure similarity and allow for a meaningful analysis, this thesis will include studies that measure political regimes using, for instance, the Polity scale or Freedom House measure of democracy. As discussed above, the theoretical debate concerning the measurement of democracy is an important aspect of this sub-field of political science, and the meta-regression analysis must consequently control for the effect of the different

conceptualisations and evaluate them.[26] Some studies also use other measures of democracy: e.g. Goldberg, Wibbels & Mvukiyehe (2008) use electoral competition. Ulfelder (2007) separates survival of authoritarianism from survival of democracy, like Przeworski et al. (2000) and Boix et al. (2013). This vast divide results in conceptually different measurements compared with other studies that measure regime on a democracy-autocracy continuum. Accordingly, they are not comparable to our chosen conceptualisation of democratisation as a graded concept. It is also not possible to calculate a common metric for studies using dichotomous and continuous dependent variables (Stanley & Doucouliagos 2012), as noted above. A way to accommodate this could be to separate the dependent variable of the meta-regression into a dichotomous variable, but as Stanley & Doucouliagos (2012: 16) note:

> Taking a continuous variable and arbitrarily dichotomizing it (e.g. significant or not significant) is likely to introduce a spurious structure into the data that does not correspond to any underlying reality.

There is a danger that the significant independent variables identified by a multiple meta-logit-regression analysis will reflect mere correlation with the publication selection process, rather than any genuine characteristic of the underlying phenomenon studied. The studies that analyse the process of democratisation with a binary dependent variable will thus have to be examined in a separate meta-regression analysis.

The impact of study quality on findings is treated as an empirical *a posteriori* question, as discussed above. Hence the inclusions of studies are not based on an *a priori* judgement of their quality. If studies do not show immediately evident measurement errors or severe design issues they are not to be excluded from the study. This is done to mimic the effect of drawing a random sample from the population, as discussed earlier. Consequently, the treatment of the impact of study quality on finding is of an empirical *a posteriori* question, rather than an *a priori* matter of opinion (Ahmadov 2014). A statistical estimate, referred to as the individual estimates precision (i.e. the standard error of the dependent variable: *SE*), is used as the indicator of quality. This is the most statistically valid approach as it is derived directly for the study's estimate and does not rely upon any additional judgment (Stanley & Doucouliagos 2012). Also, to avoid severe author dependency studies from the same author using the same dataset and/or the same time period are to be excluded from the meta-dataset. Replication-studies are also excluded for the same reason.

Based on these criteria, I derived 492 estimates from 33 quantitative studies, including 26 published journal articles, 5 working papers and 1 book. This was then coded into a dataset

[26] See Munck et al. (2002) and Hadenius et al. (2005)

that records the characteristics for each estimate that capture regional composition in the sample used to derive the estimate, period coverage, measurement, estimation, specification differences and epistemic effects. This approach is based on the Stanley et al. (2013, 2015) description of reporting guidelines for meta-regression analysis. After collecting the relevant data and coding the necessary variables, I move to the next step, which is the calculation of the dependent variable in the meta-analysis.[27]

### 3.3 Calculation of the dependent variable and summarising of the meta-data

The dependent variable in a meta-analysis is commonly referred to as the effect size, a measure that reflects the effect of the independent variable on the dependent, viz., for this case: How much (if at all) does income affect democracy? In order to conduct a meaningful meta-regression analysis and to accommodate the critique summarised in the previous sections, a common metric estimate must be calculated. As noted before, in the traditional literature this is referred to as an effect size. This effect size can be obtained in several ways and, depending on the field of research, there are many options.[28] Stanley & Doucouliagos (2012) note that the most commonly used effect sizes are elasticities, partial correlations and *t*-statistics. Elasticities are also a commonly used measure in economics. It measures the percentage change in some economic phenomenon arising from a percentage increase in some stimulus. Elasticises are often not reported directly and must therefore be calculated. Yet, they are not always possible to calculate and their standard errors can be hard to derive.[29] *t*-statistics are another measure that can be used. They are comparable across studies and easily collected from reported regression coefficients. Nevertheless, *t*-statistics can be difficult to interpret in a substantive matter (Stanley & Doucouliagos 2012: 28).

Consequently, in this thesis I will use *partial correlation* as the measure of effect size, which is a widely used measure for the effect size used in meta-analysis.[30] Partial correlation is a measure of strength and direction of the association between two variables holding other variables constant and thus providing a measure of association, *ceteris paribus* (Cohen 1992). This measure is chosen because it enables a wider and more comprehensive dataset, as will be shown in the discussion below. The measure is not commonly reported, and must therefore be

[27] An overall description of the selection procedure can be found in Figure 7.4 the appendix.
[28] For calculations of and further descriptions of alternative effect size measures see: Stanley and Doucouliagos (2012:22-29) and Hunter et al. (2004) chapter 7.
[29] See Stanley and Doucouliagos (2012:26-29) for calculation and discussion of elasticities.
[30] See for example Ahmadov (2014), Doucouliagos & Ulbasoglu (2008), Paldam & Doucouliagos (2008).

calculated, but the calculations of them are straightforward (Stanley & Doucouliagos 2012: 24-25):

**1.1.**

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

where *t* denotes the *t*-statistic of the appropriate multiple regression coefficient and *df* reports the degrees of freedom of this *t*-statistics. The standard error of the partial correlation is used as the measure for the precision (i.e. quality) of each estimate, as discussed above, and is given by:

**1.2.**

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

It should be noted that precision is the standard error of the partial correlations and not the standard error of the individual regression (Abdullah, Doucouliagos & Manning 2013).

The statistical significance of the partial correlation coefficients can be tested and calculated using the same *t*-statistics used for its associated regression coefficient, though the meta-analyst is rarely interested in the statistical significance of an individual reported effect. This applies whether we measure it by a regression coefficient or by *r*. This is due to the meta-analyst' explicit wish to investigate the underlying message of our accumulated scientific knowledge.

Partial correlations, being a measure of the association between two random variables, *ceteris paribus*, has the key advantage of being a unitless measure, allowing the partial correlation from one study to be readily compared to partial correlation of another study. Secondly, partial correlation can be valued for a larger set of estimates and studies than almost any other effect size measure.[31] On the other hand, an important drawback for the partial correlation-approach is that, like any simple correlation, its distribution is not normal when its value is close to – 1 or + 1 In political science this is rarely a problem, because none, or few, partial correlation will be close to these limits (Stanley & Doucouliagos 2012).[32] To understand the measures, the interpretation follows Cohen (1992) and his and his standard for interpretation partial correlations, *r*, where $r = 0.10$ is interpreted as a small association, medium if $r = 0.25$ and large if $r = 0.40$, or greater.

[31] For discussion of effect size see Stanley et al. (2012).

[32] Stanley et al. (2012) and Borenstein (2009) note that a normal remedy when this problem is present is to use Fisher's *Z*-transformation: $z = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right)$. However, through research, the same authors have found that the transformation makes little or no practical difference for the central findings.

The resulting correlations from the aforementioned process run from 0 to 1. Once the sign of correlation is changed to negative and if the original partial regression coefficient is negative, the partial correlation run from – 1 to + 1, ranging from a strong negative, to a strong positive association. In order to get an overview of the data, I can calculate a cumulative cross-study estimate of the relationship between income and democracy, the mean cross-study partial correlation, using different weights. This is also what was previously referred to as a "bare-bone meta-analysis":

**1.3**

$$r = \frac{\sum[N_{ij} r_{ij}]}{\sum N_{ij}}$$

where $r$ denotes the mean partial correlation between income on democracy measured by partial correlation, $r_{ij}$ is the standardised effect from the *i*th regression estimate of the *j*th study, and *N* is the samples size used as a weight, following Ahmadov (2014). The resulting measure is the best estimate of the extant empirical scholarship on the effect of income on democracy. Once $r$ is calculated, I can construct 95% confidence intervals to evaluate the accuracy of the mean partial correlation. The confidence intervals are also constructed using the fixed-effect (FE) and random-effect (RE) models, using sample size as weights. The FE model assumes the true effect to be the same across studies, whilst the RE model does not. These assumptions concerning the nature of the "true" effects is what separates these two terms in the meta-analysis framework, and are different from conventional econometric jargon (Feld & Heckemeyer 2011). Fixed effects assume that all of the reported estimates are drawn from the same population with a common mean. Hence, that the studies included in the meta-dataset are the whole population of studies (Hunter & Schmidt 2004), an assumption that seems unrealistic when there is heterogeneity present.[33] The random effects assume that estimates are drawn from several populations, i.e. that there is heterogeneity amongst studies (Stanley & Doucouliagos 2012). Random effects is therefore the preferable model, technically, but in the presence of publication bias all averages, weighted or not, are distorted (Stanley 2008). The FE/RE-discussion is also of importance when modelling the multiple meta-regression analysis, and will be discussed at length below. From these results, I will also derive the Cochran Q-test, which test the null-hypothesis that there exists no heterogeneity, i.e. systematic differences between studies. This test is thus very important for the

[33] Consult Kontopantelis & Reeves (2010) and Harris (2008) for calculations and further discussion.

multivariate meta-regression analysis, which bases itself on the assumption that this variation exists.

To sum up this section: Equations 1.1 and 1.2 provide us with the possibility to compute the statistical tests needed to examine publication bias and investigate the possible heterogeneity between studies. The partial correlation (Equation 1.1) is basically just a common metric describing the association between income and democracy, and its standard error (Equation 1.2) is a measure of the estimates precision. The last equation (1.3) lets us look at the cumulative average association between income and democracy across studies, which were referred to as a "classical meta-analysis" or "bare-bone meta-analysis" earlier. A time-series plot with the weighted average partial correlation over time will also be reported to get a grasp of the development of the research over time.

## 3.4 Publication bias

This section deals with the specifics of the meta-regression analysis (MRA), i.e. steps III-IV in the list posed above. After estimating the average effect size using the methods described above, we must ensure the robustness of these findings. The results produced by the estimated average effect size are premature in the presence of publication bias and heterogeneity, as noted. The next step in the meta-regression analysis is therefore to test for the present of publication selection bias. As we have noted earlier, a tendency to only publish papers with significant results is present in the social sciences, as well as in the medical sciences. Rosenthal called this the "file drawer-problem" (cited in Hunter & Schmidt 2004: 499), which, quite vividly, illustrates this problem of insignificant findings, or findings that run to the contrary of the existing literature in a specific field, and are put in a drawer never to be published. Further, Begg & Berlin (1988) also note that this occurs frequently in most empirical sciences. Gerber & Malhotra (2008) have found evidence of this in the field of political science journals and likewise, Doucouliagos & Stanley (2013) in their meta-meta-analysis of 87 different areas of research economics found evidence of substantial bias. In terms of testing if this bias is present, Hunter & Schmidt (2004) suggest several ways to conduct such analyses commonly used in the in the other empirical sciences, where publication bias is a very important concern. In literature on the income-democracy nexus, this problem has not been yet been addressed, therefore the investigation of publication bias in this sub-field of political science is of paramount importance. The polarisation of the studies on income-democracy might lead us to conclude that there is no publication bias,

since there are no apparent incentives to prefer one side of the debate to the other. Nevertheless, it is important that potential publication bias is addressed before continuing further, and in the presence of publication bias the meta-regression analysis itself becomes biased.

Publication bias can arise as a result of many factors. According to Callot & Paldam (2010), there are four different *priors* that can lead to censoring of the results, which in turn creates publication bias. In short these are described as (I) theoretical priors, where some part of the $\beta$-range cannot be true, as dictated by theory. Consequently, in large data-samples they ought to be impossible. This leads to censoring in small samples. (II) Further, when a result is not politically and/or morally pleasant, this is called a political prior. An example of this is Doucouliagos & Paldam (2008, 2013) who find, in their meta-analysis, that development aid is not effective. Here they reason that this is a field where many researchers, and perhaps also journals, are reluctant to publish negative results, since most researcher wish to make a positive contribution to the "laudable enterprise of development aid" (Stanley & Doucouliagos 2012: 52). (III) Researchers may also exhibit economic priors. This suggests that they work in areas where they have interests, research grants or funding for example, and censor accordingly. The last prior (IV) is referred to as a polishing: In order to reach marketable results, unclear results are censored. This is also investigated further by Brodeur, Lé, Sangnier & Zylberberg (2012), who find that the top-ranked journals are most interested in studies rejecting the null-hypothesis. When the research process leads to a study reporting a result, it is consequently generated by the "stopping rule", which concludes the search. This rule differs from study to study, and it is often difficult even for the author to know the decisive stopping rule for a paper. Callot & Paldam (2010) find three such stopping rules: (I) The results fulfil the priors of the research; (II) The results are statistically satisfactory, i.e. it satisfies the current econometric norm; and (III) the results are deemed to be marketable for economic journals.

After discussing different ways in which publication bias can arise, we now move to the discussion on how to detect and correct it. Disregarding all the possible statistical tests for a while, "the simplest and most commonly used method to detect publication selection is an informal examination of a funnel plot" (Sutton et al. 2000 quoted in Stanley & Doucouliagos 2012: 53). A funnel graph is a scatter-plot of all empirical estimates (partial correlations in our case) of a given phenomenon and these estimates' precisions, i.e. the inverse of these estimates' standard errors ($1/SE$). Studies with less precision (and thus large standard errors) are at the bottom of the graph and will produce estimates that are more spread out, and vice

versa. In practice, this means that when the sample is large, the standard error will be small and the confidence in the estimate will consequently be high, which gives us better reasons to trust that this is the correct estimate. So, studies with larger sample sizes have smaller standard errors, and are thus more precise. Studies with less precise estimate, i.e. studies with larger standard errors, will have lower precision. This is because studies with smaller samples, and hence larger standard errors (*SE*) will need to run and re-run their models more intensely to achieve statistically significant results. More precise studies, i.e. those with smaller *SE*, will require less searching and less selection to obtain the desired significant results (Doucouliagos & Stanley 2013). Accordingly, the tell-tale signal of publication selection is a systematic relation between reported effects and their standard errors (Stanley 2005).

If publication bias is not present, theories of heteroscedasticity dictate that we should expect to observe an inverse funnel in the scatter-plot because we have a theoretical expectation of estimates from studies being negative, zero and positive, with different levels of precision (Stanley 2008). Put simply: In the absence of publication selection, estimates will vary randomly, hence symmetrically, around the one "true" effect in the shape of an inverse funnel. One problem with the funnel plot-approach, however, is that the interpretation is of a subjective nature. We therefore need statistical tests to determine how to test for symmetry. The statistical test used in this thesis is the funnel-asymmetry-test (FAT), which tests if the funnel plot is symmetrical. This test also produces the precision-effect-Test (PET), which test whether or not there is a genuine underlying effect beyond the potential distortion due to publication selection bias (Stanley 2005; Stanley 2008). The funnel-asymmetry-test and precision-effect-test meta-regression analysis (FAT-PET-MRA) is considered to be the preferred method in modern meta-regression analysis. However, other approaches are available but will not be discussed at length here due to the superiority of the FAT-PET-MRA (Stanley 2008; Stanley & Doucouliagos 2012).[34]

The funnel-asymmetry-test and precision-effect-test meta-regression analysis can be expressed by using weighted least squares-model (WLS), to accommodate heteroscedasticity, with the precision squared ($1/SE^2$) as weights:

**1.4.**

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

[34] Amongst these options, we find the failsafe *n*-method, trim and fill-method and meta-significance-testing. For a thorough discussion of these, see Stanley (2005, 2008), Stanley et al. (2012) and Rosenthal (1979).

where $r_i$ denotes the reported estimates of the partial correlation, and *SE* its standard errors. Publication bias is tested for by checking the symmetry of the (funnel-asymmetry-test), i.e. if the slope of $\beta_1$ is statistically and practically different from zero, we can say that there is publication bias (Test: $H_0$: $\beta_1 = 0$) To test whether there is a genuine effect beyond potential publication bias (precision-effect-testing), we test if $\beta_0$ is statistically and practically different from zero (test: $H_0$: $\beta_0 = 0$) (Stanley & Doucouliagos 2012). In order to interpret the results of the funnel-asymmetry-test (FAT) Doucouliagos & Stanley (2013: 320-21) suggest the following guidelines:

1. If FAT is statistically insignificant or if $|\hat{\beta}_1| < 1$, then selectivity is "little to modest"
2. If FAT is statistically significant and if $1 \leq |\hat{\beta}_1| \leq 2$ then selectivity is "substantial"
3. If FAT is statistically significant and if $|\hat{\beta}_1| > 2$ then there is "severe" selectivity

The results from the precision-effect-test are interpreted following the guidelines of Cohen (1992), in the same manner as partial correlations.

If publication bias is present, the precision-effect estimate with standard error (PEESE) meta-regression analysis is the preferred option to account for this. The FAT-PET-PEESE meta-regression analysis provides a better estimate for the underlying "true" effect when there is an effect. However this is not true when there is no empirical effect and only publication selection. The PEESE-model takes the following form:

**1.5.**

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

where the standard errors from the partial correlation are included in both terms. It should be stressed that PEESE should only be used if publication bias is found when modelling FAT-PET MRA, or if the PET-term shows a genuine empirical effect (reject $H_0$: $\beta_0 = 0$).

In sum: The statistical test for the funnel asymmetry (FAT) fits a least squares to the funnel plot and checks if the line has a significant positive or negative slope, and if it does we can say that there is publication bias present in the literature, according to the guidelines above. Also, because of heteroscedasticity, a regular OLS estimation does not suffice. Other options, like fixed and random effects and cluster-robust WLS models will also be used, but this will be discussed in full below. The precision-effect test (PET) serves as a test of whether or not there is a genuine underlying empirical effect beyond the potential distortion due to publication bias and the precision-effect estimate with standard error (PEESE) is a robustness check testing the same genuine empirical effect. Simple meta-regression analysis models of publication selection often provide an adequate overall estimate of an empirical effect

corrected for publication bias, but one can never categorically rule out the possibility that some strong systematic research heterogeneity will overwhelm any single-value research summary, making it irrelevant and perhaps misleading (Stanley & Doucouliagos 2012). Therefore it is important to explicitly deal with potential heterogeneity. In the following section we will discuss the ramifications of this assumption. Recall that we must address both publication bias ($Q_{1a}$) and heterogeneity ($Q_{1b}$) in order to answer this thesis' main research question ($Q_1$).

## 3.5 The multiple meta-regression-analysis

In this part, I will discuss the methodology behind the multiple meta-regression analysis, which is an extension of the funnel-asymmetry-test and precision-effect-test meta-regression analysis Equation 1.4 cited above.[35] The objective of this part of the meta-analysis is to address and explain how the systematic variation that exists between studies, i.e. study heterogeneity, affects study outcome, whilst simultaneously controlling for publication bias. The question about whether this heterogeneity is a feature of the underlying data generating process, or if it is an outcome of the research design process as a whole, will be addressed through this statistical model. The former implies that there is an underlying distribution of the income-democracy parameter values which are negative in certain situations and positive in other, and the latter implies that reported differences are the result of artefacts, such as differences in econometric specification, sample compositions, and the like (Doucouliagos & Ulbasoglu 2008). As pointed out earlier, the Cochran Q-test serves as an empirical parameter for existing heterogeneity, as well as our theoretical foundation outline in the review of the literature. To conduct a successful multiple meta-regression analysis, as in most empirical research, the researcher needs to understand and explain the observed variation in the empirical world, which in this case are the studies gathered. And in the field of social science this is always much greater than what should be expected to stem from random sampling error alone. This variation needs to be accessed to ensure the validity of the meta-analysis itself. As in all empirical research, the omission of important variables might be a source of bias in the analysis. To accommodate this and explain this excess research variation, a multivariate meta-regression analysis should be utilised (Stanley & Doucouliagos 2012).

Given that this is the extension of the bivariate funnel-asymmetry-test and precision-effect-test meta-regression analysis, the multiple meta-regression analysis includes several

[35] Specific calculations are shown below in equation 1.6.

other independent variables. The individual characteristics of the studies that we hypothesise to have an effect on the partial correlation between income and democracy, i.e. the dependent variable in the meta-regression analysis, constitutes the independent variables. The multiple meta-regression analysis examines their impact on the relationship between income and democracy. In a word: The *characteristics of the studie*s are the *independent variables* which are hypothesised to explain the variation in the findings regarding the relation between income and democracy that we observe across these study-estimates. The meta-regression analysis will measure the importance of these independent variables Stanley & Doucouliagos (2012) make a distinction between *essential* and *typical* variables. *Essential* variables are those necessary for conducting the meta-analysis such as effect size (partial correlations in our case) and precision (*SE* of the partial correlations*)*. *Typical* variables are study characteristics made, such as regional composition in the sample, period data, measurement, estimation difference and other specifications. The *typical* variables thus contain information on study-specific choices made by authors in relation to methodological questions, conceptualisation, etc. The *typical* variables are mostly coded as binary variables and the essential variables are coded as continuous variables.[36]

Moving back to the specific estimation technique, a discussion needs to be carried out in order to determine which statistical method can best account for this heterogeneity in findings that we observe across our sample of studies. The meta-analyst has a full arsenal of econometric techniques available to answer the research question, but there is a considerable debate concerning the choice of estimation method. Due to heteroscedasticity that stems from different estimation techniques, data sets, sample size and the variance of the estimated coefficient that may not be equal in the sample of empirical studies collected, the use of regular OLS should not be carried out (Stanley 1989). Consequently, there is a discussion about which estimation technique to use when conduction the multiple meta-regression analysis. A central debate in the literature is whether to use *fixed effects* (FE), *random effec*ts (RE) or a model with weighted least squared (WLS), using precision squared as weights ($1/SE^2$), when modelling the multiple meta-regression analysis. This debate is vast in the literature, and need to be summarised. Below, I discuss the differences between RE and FE models before concluding that the use of WLS might be viewed as the most viable solution. I also discuss the methodological consequences of using fixed effects clustered by studies with

[36] A full list of the independent variables will be provided in a subsequent section.

author dummy variables (FE) and mixed effects multilevel models (MEML)[37] to assess the possible hierarchical nature of the meta-data, i.e. that estimates are clustered in studies

Let us now proceed to develop the discussion of fixed and random effects models in the context in of the multiple meta-regression analysis. Firstly, a random effects model can be useful if study differences result from both sampling errors as well as random differences between studies (Doucouliagos & Stanley 2008). The RE model also assumes that the studies included are a sample of the whole population, and it can therefore be argued to be more realistic than the assumptions associated with the FE model, which assumes that all estimates are drawn from the same population with the same mean. In dealing with heterogeneity in social science, we can assume that all excess heterogeneity is random and independent of all of our moderator variables, including the standard error. These RE models incorporate an additional term to the meta-regression model, which allows for any between-study random variation. A crucial assumption, however, is that, as in any RE model, these added random effects are *independent of all the explanatory variables, which is very unlikely to be true in meta-regression models*. In imprecise studies, e.g. those with large standard error where it requires a greater effort to find statistical significance, *the random effect will routinely be correlated with the standard error in cases with publication selection bias*. Random effects are, in these cases, partly, the result of these greater efforts to select and report desired estimates.

Stanley & Doucouliagos (2012: 83) show through simulations that there is a "positive correlation between random, yet selected, heterogeneity and the estimate's publication selection". This expected correlation increases with the incidence of publication selection. The same simulations also show that random effects meta-regression analysis causes larger biases than fixed effects meta-regression analysis. In the medical sciences there is no culture, even though heterogeneity is present, for investigating it because of the fact that their estimates come from controlled experiments. They therefore assume that any observed excess heterogeneity is random. This conclusion does not hold in the social sciences, and "although this area of study merits further study" (Stanley et al. 2012: 84), the use of RE models are not recommended in meta-regression analysis because the observed heterogeneity is not random.

Following this discussion, I conclude with that I am best advised to rely on a model with weighted least squares (WLS), using precision squared as weights ($1/SE^2$), as recommended by Stanley & Doucouliagos (2012, 2013a, 2013b). The WLS model allows for

[37] Random effects multilevel models will be referred to as mixed effects multilevel (MEML), because they incorporate both fixed and random effects.

research dimensions that explain both the reported heterogeneity amongst the results and ensures that any simple meta-regression analysis finding is robust to more comprehensive analysis (Stanley & Doucouliagos 2012). WLS models are also efficient if estimation models are heteroscedastic, and the nature of heteroscedasticity is known, which is the case in meta-regression analysis (Feld & Heckemeyer 2011).

FE, RE and WLS can be modelled to address the hierarchical data structure with estimates clustered within studies, are also important to ensure the robustness of the findings of the meta-regression analysis. The question about which model is the *true* model based on this discussion might be difficult, but "fortunately, there are objective criteria for meta-regression analysis selection" (Stanley & Doucouliagos 2012: 102). First, it is important to address the different analytical levels in the data. I therefore use the Breusch-Pagan Lagrange Multiplier (LM) test to test for significant study-level effects (Stanley & Doucouliagos 2012). Theoretically, and empirically, we would expect there to be study-level effects, but are they *fixed* or *random*? To answer this second question, a Hausman specification test can be used. This test makes it possible to "differentiate between an estimator that is consistent under both alternative specification (FE) and a second estimator that is consistent but also efficient under one specification (RE)" (Stanley & Doucouliagos 2012: 103). Recall, nonetheless, the discussion above: All RE models assume that the random effect is uncorrelated with the independent variables; but publication selection bias is likely to make selected random effect *correlated* with the standard error, which is one explanatory variable needed to be investigated in any meta-regression analysis. It is therefore recommended that fixed effects with study dummy variables clustered by studies (FE), and *not* random effects multilevel (MEML) be used to address the multilevel structure of the data. This is due to the fact that the fixed effects model is a "generalisation of the random effects model because it allow the study-level effects to be correlated with the independent variables" (Stanley & Doucouliagos 2012: 103)

Following this discussion, Equation 1.6 expands the funnel-asymmetry-test and precision-effect-test meta-regression analysis equation (eq. 1.4.) to be a multivariate model by adding variables *Z*- and *K*-variables in order to explain and allow for the exploration of heterogeneity in the reported results. This equation describes the standard WLS model used as a starting point for the multiple MRA. The *Z*-variables are study characteristics coded as dummy variables, and *K*-variables are related to publication bias, here reported through partial correlation and the standard error of the partial correlations.

**1.6.**

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

Further, I will use the general-to-specific (G-to-S) modelling approach in order to approach the "correct" model, as recommended by Stanley & Doucouliagos (2012). This approach consists of taking the different dimensions that might potentially affect the magnitude of the reported results into account. Given that the analyst were to include 10 moderator variables in the meta-regression analysis, the possible variation of models would exceed 10 million possible models, and thus it is necessary to rethink this approach. *The G-to-S begins with all explanatory variables included in the equation, and then excludes variables that are the least statistically significant, instrumentally one at a time, until only the statistically significant remain.* This is not ideal, of course, but its strength is that the model proceeds from a very general model in a more statistically valid fashion. Another sensible approach is to report only the meta-regression analysis model that includes all coded moderator variables. Assuming that there are sufficient degrees of freedom to run the all-inclusive meta-regression analysis, it would be difficult to see through the fog of multicollinearity (Stanley & Doucouliagos 2012: 91). Thus, the general-to-specific approach is utilised for the reasons discussed above, and to avoid a "garbage bin" model.[38]

To answer the main research question concerning the relationship between the two variables income and democracy ($Q_1$), we must first answer if there is publication bias ($Q_{1a}$) and heterogeneity ($Q_{1b}$) in the literature. This is done by computing predictions from the multiple meta-regression analysis by inserting specific values for the independent variables in the model. These predictions give us an unbiased estimate of what the nature of the association between these two variables really is. The combined procedures described in this chapter permit us to answer both what the effect of per capita income on democracy ($Q_1$) and why we observe no consensus on this question in the literature ($Q_2$).

## 3.6 Summary

This section describes the way in which this analysis is carried out. I started this chapter with a discussion of the meta-analysis in relation to traditional literature reviews and surveys, concluding that the meta-regression analysis is the most thorough and scientifically rigorous way to conduct such an analysis. Further, I discussed the way the calculation of the essential

[38] For illustration of the overall estimation procedure, see Figure 7.7 in the appendix.

variables for the meta-regression analysis is to be carried out and showed how the average effect size could be calculated, how publication bias should be addressed, and reviewed the statistical foundations of both the funnel-asymmetry-test and precision-effect-test meta-regression snalysis and multiple meta-regression-analysis.

To answer the $Q_1$, "*What is the effect of par capita income on democracy?*" I first have to address publication bias ($Q_{1a}$) and heterogeneity ($Q_{1b}$). Therefore, the chosen approach in this thesis to answer these research questions is the following: I will summarise the data through both a time-series plot and weighted averages to get an overview of the data. I will then display and interpret the distribution of the partial correlation between income and democracy using a funnel plot, which is the simple scatter-plot discussed above. Recall that this is done to answer the question about the potential publication bias. Following this I will conduct the statistical test (FAT-PET-MRA) necessary to ensure that the right conclusion is drawn from the inspection of the funnel plot. Then I will then move towards the multiple meta-regression analysis.

Firstly, I will estimate a full WLS-model, as depicted above. Secondly, I will use the general-to-specific approach to trim down the number of moderator, or independent, variables. Thirdly, due to the multiple estimates per study, I conduct a Breusch-Pagan LM test to find study level effects. If these are present, which we have theoretical reasons to expect, then fourthly, I run a cluster-robust WLS, fixed effects with study dummy variables and mixed effects multilevel model. To determine if the study-level effects are "fixed" or "random", I use the Hausman-test to identify the correct specification (viz., whether I should employ FE or RE). Fifthly, I report all meta-regression analysis model specifications and focus on those research dimensions that have consistent findings across alternative meta-regression analysis models, this will allow me to answer $Q_2$: "*Why is it that we observe no consensus on this question across the scholarly literature?*" Lastly, to answer $Q_1$, I predict the correct effect from the multiple meta-regression analysis when both publication bias and heterogeneity is controlled for.

## 3.7 Meta-data set

Before proceeding to the actual analysis, I will now discuss the meta-data set and the independent variables included in the study. This lays out the foundation for the meta-regression analysis that follows in the actual analysis. This chapter presents the original data that has been collected for the purpose of this thesis. I will first present descriptive statistics of

the dependent variable, sorted by studies, and then take account of all the different independent variables used in the multiple meta-regression analysis. The number of studies collected is 33, with 492 individual regression estimates from which partial correlations have been derived.[39] In a meta-meta analysis of meta-analysis carried out in the field of economics, Doucouliagos & Stanley (2013) found that the average number of studies included in meta-analysis is 41, with a median of 35. Consequently, this sample is within the bounds of the average meta-data se,t and the 492 individual observations offer sufficient degrees of freedom to utilise the full potential of the multiple meta-regression analysis. Below I present a list of the articles included in the meta-data set with summary statistics of the dependent variable in the meta-analysis, viz. partial correlations.

Table 3.1 below shows the 33 studies that are included in the meta-data are shown with corresponding descriptive statistics of their respective partial correlations. In some cases, studies do not provide sufficient information to calculate the partial correlation because of missing information on *t*-values, standard errors or *p*-values. This might be a substantial problem in some meta-analysis, because the reporting standards across journals and books could vary. In the construction of this data, missing information has not been a severe problem. The 33 studies included in the sample contained the sufficient data described above. This includes *t*-statistics, regression coefficients standard errors or *t*-values. If *t*-values are not reported, they are calculated by dividing the regression coefficient by the corresponding standard error:

**1.7**

$$t = \frac{\beta}{SE}$$

If the *p*-values and degrees of freedom are reported, the calculation of *t*-statistics is also possible. However, if studies only report significance levels of 1, 5, or 10%, a standardised *t*-value has to be assigned to them by the meta-analyst and this may introduce some measurement error in the meta-data (Stanley & Doucouliagos 2012). All 33 studies in this sample, however, reported sufficient estimates, so this possible bias does not occur in this analysis. For approaches on how to handle this, see for example Greenberg, Michalopoulos & Robins (2003).

As Table 1 below clearly shows, the collected estimates from studies vary. Acemoglu et al. (2008), for instance has 53 estimates, as opposed to Hariri (2015), which only has two. On average, studies contain 15 regression coefficients. We also observe that the partial

[39] For a visualisation of the selection procedure, see Figure 7.6 in the appendix.

correlations vary from negative to positive across studies with – 0.254 as the minimum value and 0.737 as the maximum value. The overall unweighted mean partial correlation is positive 0.082, which suggests that, on average, the relationship between income per capita and democracy is slightly positive. In the terms of Cohen (1992), this is a very small positive association. This unweighted mean is nevertheless premature because it does not take publication bias and study heterogeneity into account.

**Table 3.1: Distribution of dependent variable sorted by studies included in meta-data**

| Author(s) | No. of estimates | Min | Max | Mean | Std.div. |
|---|---|---|---|---|---|
| Acemoglu et al. (2005) | 12 | -.0695 | .0812 | .0055 | .0439 |
| Acemoglu et al. (2008) | 53 | -.2456 | .5130 | .0173 | .1388 |
| Acemoglu et al. (2009) | 12 | -.0977 | .2414 | .00001 | .1039 |
| Albertu et al. (2012) | 12 | -.0814 | .0964 | .0053 | .0407 |
| Ansell et al. (2010) | 12 | -.0533 | .0441 | .0029 | .03587 |
| Aslaksen (2010) | 5 | .0207 | .2491 | .0974 | .0891 |
| Barro (1999) | 2 | .3378 | .3567 | .3473 | .0134 |
| BenYishay (2014) | 32 | -.0981 | .0692 | .0180 | .0375 |
| Boix (2009) | 23 | -.0052 | .2598 | .0798 | .0596 |
| Boix (2011) | 25 | -.0139 | .4090 | .1455 | .1301 |
| Booba et al. (2007) | 3 | .0013 | .0014 | .0013 | .00001 |
| Brinks et al. (2006) | 3 | -.0505 | .0732 | .0295 | .0694 |
| Cervellati et al. (2014) | 42 | -.1781 | .4197 | .0203 | .1425 |
| Che et al. (2012) | 41 | -.0585 | .4153 | .1627 | .0856 |
| Colaresi et al. (2003) | 3 | .0149 | .0183 | .0168 | .0017 |
| Cornell (2012) | 8 | .0141 | .0664 | .0451 | .0167 |
| Faria et al. (2014) | 10 | .0001 | .0745 | .010 | .0231 |
| Hadenius & Teorell (2005) | 3 | -.0087 | .0597 | .0282 | .0346 |
| Hariri (2012) | 3 | .2962 | .3833 | .332 | .0451 |
| Hariri (2014) | 2 | .2873 | .3927 | .340 | .0745 |
| Heid et al. (2012) | 15 | -.0631 | .2281 | .1030 | .0965 |
| Herb (2004) | 4 | .0352 | .1120 | .0726 | .0314 |
| Högström (2013) | 3 | -.0125 | .1837 | .1082 | .1056 |
| Jensen & Paldam (2005) | 32 | .1687 | .6196 | .3692 | .1139 |
| Knack (2004) | 4 | -.0354 | .1676 | .0696 | .0979 |
| Moral-Benito et al. (2012) | 18 | -.1703 | .1678 | -.0089 | .1091 |
| Oskarsson & Ottosen (2010) | 16 | -.0128 | .7370 | .1644 | .1749 |
| Ross (2001) | 24 | -.0476 | .4033 | .1477 | .0927 |
| Rudra (2005) | 7 | -.0900 | .0763 | -.0107 | .0725 |
| Schimmelfennig et al. (2008) | 7 | .0145 | .0265 | .0193 | .0047 |
| Skaaing et al. (2012) | 8 | .0784 | .1336 | .1023 | .0199 |
| Teorell (2010) | 4 | -.0449 | .0337 | -.0085 | .0323 |
| Treisman (2011) | 25 | -.0491 | .2244 | .0481 | .0631 |
| Werger (2009) | 19 | -.0878 | .1675 | -.0042 | .0545 |
| **Nr. of studies: 33** **Sum:** | **492** | **-.2454** | **.7370** | **.0822** | **.1397** |

PC: Partial correlation between income and democracy. Std.div.: Standard deviation.

### 3.7.1 Independent variables

Before conducting the multiple meta-regression analysis, a review of the independent variables are in due order. In the multiple meta-regression analysis the independent variables, or moderator variables, are the individual attributes of the studies or its authors. These independent variables are the sample chosen, statistical estimation methods and other specifications, as outlined in the review of the literature above. In the process of collecting data from the studies, potential sources of heterogeneity were coded in the meta-data set. Their selection is based on the survey of the existing literature and discussion made in a previous chapter to try to determine whether there are systematic differences between studies, answering $Q_{1b}$ and $Q_2$. Some of the variables, like the partial correlation, number of countries included and year of publication are continuous variables. The remaining variables are binary ones. Here, the value 1 is given to all specific estimates that use a specific method, sample, measure of democracy, etc. All these variables are summarised in Table 3.1 below. The variables are mostly self-explanatory, and are divided between essential variables, variables that are necessary for conducting the meta-regression analysis, like the partial correlation and the standard error of the partial correlation, and typical variables, i.e. attributes with the studies, as discussed above. The essential data are continuous variables and correspond to the *Z*-variables described in Equation 1.6. The typical data are most often binary variables and relate to the *K*-variable in the same equation. As hypothesised, these variables are assumed to explain why studies find a variety of different results when researching the relationship between income and democracy.

## Table 3.2: Description of variables

| Variable | Description | Min | Max | Mean | Std. Dev. |
|---|---|---|---|---|---|
| **Essential variables** | | | | | |
| Partial correlation | Partial correlation between democracy and income | -.245 | .737 | .0822 | .1397 |
| Standard error | Std. err. of partial correlation | .0047 | .0001 | .0027 | .0432 |
| **Estimation differences** | | | | | |
| OLS FE | BV: 1 = Uses OLS with fixed effects & AR1 | 0 | 1 | .3780 | .4853 |
| Endogenous | BV: 1 = Democracy is endogenous | 0 | 1 | .3069 | .4616 |
| **Time and region** | | | | | |
| Annual | BV: 1 = Uses annual observations | 0 | 1 | .380 | .4859 |
| Five year | BV: 1 = Uses five year pooling of observations | 0 | 1 | .4085 | .492 |
| Pre1950 | BV: 1 = Uses pre 1950 | 0 | 1 | .4349 | .4962 |
| Countries | Number of countries used | 7 | 184 | 110.14 | 36.99 |
| OECD | BV: 1 = Includes OECD countries | 0 | 1 | .9268 | .2606 |
| No socialist | BV: 1 = Excludes socialist countries | 0 | 1 | .0853 | .2797 |
| **Knowledge effects** | | | | | |
| Income focus | BV: 1 = Income-democracy nexus is primary focus | 0 | 1 | .7113 | .4535 |
| Political scientist | BV: 1 = Researcher is political scientist | 0 | 1 | .5731 | .4951 |
| Unpublished | BV: 1 = The study is an unpublished work | 0 | 1 | .1646 | .3712 |
| Prior | BV: 1 = If the first author has published in field previously | 0 | 1 | .8495 | .3578 |
| Publication Year | Year of publication with 1999 as base | 1 | 13 | 9.047 | 3.173 |
| **Specifications differences** | | | | | |
| FHI | BV: 1 = Uses Freedom House Index (Ref. Polity) | 0 | 1 | .4349 | .4962 |
| Maddison | BV: 1 = Uses Maddison data (Ref. World Bank) | 0 | 1 | .1138 | .3179 |
| Penn | BV: 1 = Uses Penn World tables (Ref. World bank) | 0 | 1 | .6219 | .4853 |
| Education | BV: 1 = Includes education control | 0 | 1 | .1565 | .3637 |
| Population | BV: 1 = Includes population control | 0 | 1 | .1544 | .3617 |
| Inequality | BV: 1 = Includes inequality control | 0 | 1 | .0569 | .2319 |
| Oil | BV: 1 = Includes oil control | 0 | 1 | .2154 | .4115 |

BV: Binary variable. OLS: Ordinary least squares. FE: Fixed effects AR1: Autoregressive term.
OECD: Organisation of economic co-operation and development

As the second section of the shows, I have included variables that measure the effects of different estimation methods. 38% of the collected estimates use OLS regression with robust standard errors, fixed effects and an autoregressive term (AR1). As Beck & Katz (2001) argue, this is the preferred approach in TSCS. The variable *OLS FE* is given the value 1 if this methodological standard is employed. However, some 30% of the collected estimates specify potential endogeneity in a different way (recall the discussion above about the problems with the AR1 approach). These estimates are produced by methods such as Generalised method of moments, (GMM), instrumental variables (IV) and two- and three-

stage least squares (2SLS, 3SLS). The variable *Endogenous* is coded 1 if studies use these statistical methods. The *OLS FE* and the *Endogenous* variables are mutually exclusive, but in the values that are coded 0 in the *OLS FE* variable compares this specific methodological variation, as opposed to all other, random effects models and GMM alike.

In the next section of the table, I present time and regional specifications that I have coded. Most of the studies focus on the entire world during the third wave of democratisation, but some also include, or exclude, countries and time periods. The variable *OECD* captures whether OECD countries are included in the sample or not. This is used as to separate the developing world from the developed. The subsequent variable, *No socialist*, identifies studies that exclude (former) socialist countries. Additionally, the variables *Annual* and *Five year* identify the way the data is averaged over time. As we have seen, data pooling varies from annually to 50-year intervals. These variables capture whether the averaging of data is a source of heterogeneity. The variable *Pre 1950* checks whether studies that look at democracy *in the long run* display different results than studies looking at the individual decades from 1950-2000.

I have also included so-called "knowledge effects", which capture attributes of the authors and studies. The variable *income focus* is coded 1 if the article explicitly states that the relationship between income and democracy is the main focus. *Political scientist* is coded one if the first author of the study is a political scientist, zero if the first author is an economist. *Unpublished* is coded 1 if the study is not published in a peer-reviewed journal, but is rather a working paper, thesis, etc. *Prior* is coded one if the first author has published previously in the democratisation research with a focus on income and democracy. *Publication year* is also included. This is a continuous measure of when the study is published with 1999 as base.

Additionally, I have included some variables that capture specification differences across studies concerning conceptualisation of the dependent variable, measurement of the independent variable, and what control variables are used. Firstly, the variable *FHI* relates to the measurement of democracy. There is, as we have seen, an extensive debate in the literature concerning which measure of democracy should be used. In the tradition of continuous conceptualisation of democracy, the Freedom House Index (FHI) and Polity are measures used across the studies in our sample.[40] Given the substantially different conceptualisation of these two measures, I take into account what empirical effects the

[40] Distinctions are not made between different editions of FHI or Polity because these conceptualisations are assumed to be, to a sufficient degree, the same over time.

different conceptualisations have. Studies that use FHI are thus coded 1, and studies that use Polity are coded 0. The use of different measures of democracy might be an important source of variation in the empirical results (Bollen 1990; Munck & Verkuilen 2002). Thus, I use *FHI* as a variable to check whether studies that use this index tend to find different results, as compared to those that use Polity.

I now move to consider the independent variable in the study, the measure of income. All studies used in this analysis measure income in GDP, but there is some variation related to which dataset is used to derive the GDP. In the sample we include studies that use Maddison, Penn World Tables or World Bank Data. I have thus created the variables *Maddison* and *Penn*, which is coded 1 if studies use either of these measures, and use World Bank data as a reference category since the variables capture the whole population of studies. Further, I have included some possible control variables. These include *Population, Inequality* and *Oil*. Each of these is coded 1 if studies use one, or more, of these as a control variable in their regression.

When conducting a meta-regression analysis there may be many possible control variables, and they can often exhaust the degrees of freedom. It is therefore necessary to omit some potential meta-regression analysis variables (Stanley & Doucouliagos 2012). Another important measure to be taken to accommodate this problem is to code the variables broadly, which reduces the omitted variable bias problem. However, given the theoretical considerations and the degrees of freedom available, the range of moderator variables included here seems reasonable.

### 3.7.2 Different meta-data sets

Many meta-analyses make distinctions between *independent-set, best-set, average-set* and *all-set.* The independent-set consists of estimates that are deemed conceptually different, the best-set is the estimate favoured by the author(s), the average-set is constructed by taking an average of all effect sizes reported by each study and the all-set consists of all relevant estimates reported. In this thesis the all-set is the preferred dataset. The *all-set* consists of all relevant estimates reported in each of the individual studies. This increases the number of observations, but also results in a potential interdependence between data points, which needs to be accounted for by the appropriate statistical method, as discussed above. The advantage of using the all-set approach is that it offers more estimates to explain the large variation (heterogeneity) typically found between studies. It also does not contribute to selection bias

produced by the meta-analyst. Another reason for choosing this dataset over, for instance, best-set, is that the regression estimates favoured by the author are not explicitly stated.

Multiple sampling allows for more powerful tests and more accurate estimates due to a larger underlying sample. Furthermore, single sampling greatly reduces the degrees of freedom available to the multiple meta-regression analysis, which makes it difficult to account for the research dimensions due to a lack of degrees of freedom (Feld & Heckemeyer 2011). Many meta-analysts prefer to also include an average-set, following Stanley (2001); Doucouliagos & Paldam (2008); and Ahmadov (2014), to demonstrate robustness. According to Stanley & Doucouliagos (2012) and Feld & Heckemeyer (2011) conventional practice has evolved to the use of the all-set as the standard dataset. But for the sake of robustness, the average-set will also be used, to some extent when looking at the average partial correlations, since it is a natural extension of the all-set. But for the multiple meta-regression analysis I focus exclusively on the all-set data.

## 3.9 Summary

In this section I have presented the meta-data set and discussed some relevant descriptive statistics related to both the articles included in the meta-analysis and the independent variables of the multiple meta-regression analysis. We saw that the individual estimates derived from studies varied both in terms of how many estimates where reported and what values that actually was reported. In sum, Table 3.1 showed 33 studies with 492 corresponding partial correlations. In Table 3.2 the descriptive statistics and coding of the variables was explained and discussed. Based on the previous discussion of the existing literature, four dimensions that lead authors to produce varying results was identified. These independent variables are used in the multiple meta-regression analysis in order to explain, statistically, what creates systematic differences between studies, which in order helps answer explain the main research posed in this thesis: "*What is the effect of per capita income on democracy?*" In the concluding part of the chapter I discussed several versions of the meta-data and concluded that the main focus will be on the *all-set* data. In the next section we move to the actual analysis of the empirical data.

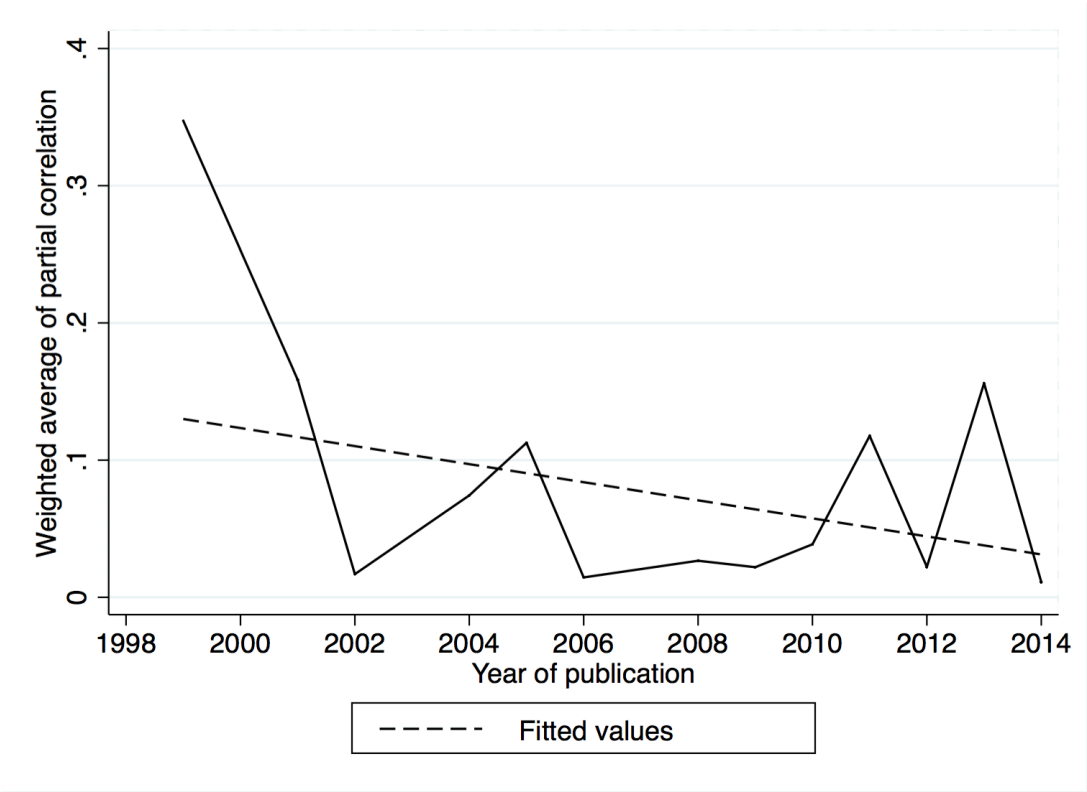
# 4. Meta analysis of income and democracy

In this section I will conduct the actual analysis that will help us to answer the research question posed in the introduction. To answer this question we must conduct a series of tests, which were laid out in the chapter on meta-analysis methodology. In an attempt to answer $Q_1$: "*Is there an effect of per capita income on democracy?*", I show the average partial correlation between these two variables. This will gives us an overview of the general association between these two variables in the literature. These averages are biased because they are not able to control for publication bias ($Q_{1a}$) or heterogeneity ($Q_{1b}$), as mentioned in the introduction. These two important factors must be controlled for in order to answer $Q_1$ thoroughly. Therefor this chapter will go through several phases: First, I start with a graphical overview of the development of the literature over time, discuss some trends that are observed in the literature and compute said average partial correlation. Secondly, I address publication bias, both through a funnel plot and the funnel-asymmetry-test. Here I show that there is no significant publication bias present in the literature, because the funnel plot represents a symmetrical funnel. The funnel-asymmetry-test confirms this observation and shows that the publication bias in the literature is very small and practically indistinguishable from zero, thus answering $Q_{1a}$. Next, the precision-effect-test (PET), which is conducted simultaneously with the funnel-asymmetry-test, is not able to detect any real effect of income on democracy when the potential publication bias is controlled for. Thirdly, I expand this model to a multivariate meta-regression analysis and include variables that control for study heterogeneity. This answers both $Q_{1b}$ "*Are there systematic differences between studies that generate different results?*" and $Q_2$: "*Why is it that we observe no consensus on this question across the scholarly literature?*", because we can empirically address what generates the different findings in the literature. In this part, I show that there are methodological, sample variation, author attribute, and specification differences that create much of the variation and confusion present in the literature. From this model I predict several research designs that the literature deem to be "best practice" and show that when using proper methodological specification there is no statistically or practically significant effect of income per capita on democracy, thus answering $Q_1$ in a valid fashion.

## 4.1 Descriptive statistics

As noted above, Figure 4.1 shows the times-series pattern of the income-democracy effect. The graph displays a time-series plot of the cumulative annual average partial correlation, weighted with the sample size. First, we observe that historically, the relationship between income and democracy has, on average, never been reported to have a negative value. Further, this shows that the early literature, starting with Barro (1999), reported relatively large, positive association. As more evidence has accumulated, the fitted values show that the effect has diminished from the late 2000's. It is possible to theorise that the publication of Przeworski et al. (2000) and its widespread impact influenced subsequent studies, bringing on a decline of the effect, but their findings are not included in this graphical display. Following this reasoning, the time-line suggest that the publication of Boix & Stokes (2003) and its the corresponding impact of its findings which challenged Przeworski et al. could have contributed to the rise in the partial correlation, while most recently, and Acemoglu et al. (2008); Acemoglu et al. (2009) contributed to the patterns found in the literature from 2010 until today. However, this is not sound inference and will remain merely a speculative interpretation. In this timespan the studies have reported a relationship around + 0.1, which indicate a small positive effect (Cohen 1992). It is clear that the income-democracy effect is either unstable and has declined over time, as seen by the fitted values. Additionally, it is important to keep in mind that the sample is not balanced, but for illustrative purposes the time-series plot is used to display the literature, as proposed by Doucouliagos & Ulbasoglu (2008).

**Figure 4.1: Weighted average partial correlations over time**




In Table 4.1, I present relevant summary statistics of the partial correlations between income and democracy. Column 1 displays descriptive statistics for the *all-set* data, including all 492 collected estimates from the 33 studies. And Column 2 reports the *average-set* data. In the *average set* one average estimate per study is calculated, resulting in 33 estimates from 33 studies. As discussed above, this is a natural extension of the *all-set* data and is included for the sake of robustness, and to deal with study dependence that stem from the fact that studies report more individual estimates than others. In the multiple meta-regression analysis the *all-set* data are preferable because it allows for more degrees of freedom. Columns 3 and 4 report the average for the groups of estimates derived after controlling for the effect of two key methodological variations: The studies that use fixed effects regressions with an autoregressive term and studies that use other statistical techniques to model endogeneity, as discussed above. The weighted averages and confidence intervals are constructed with, respectively, sample size as weighted (N), fixed effects (FE) model, and random effects (RE), using Equation 1.3, as discussed above. These individual estimation techniques are indicated in the vertical rows with their corresponding confidence interval below. Summing up available empirical evidence, all means indicate that there is a positive effect of income on

democracy, but in most cases this effect is very small and not practically different from zero. This indicates that income increases the level of democracy, but the effect is small in meta-analytical terms. *Ceteris paribus*, an increase of income will thus increase the partial correlation by values ranging from + 0.041 to + 0.082, depending on weighting. The results from Columns 3 and 4 positive, although the studies that use fixed effects find a smaller effect than the studies that use GMM or IV. In Cohen's (1992) terms, the association remains very small, and indistinguishable from zero. Additionally, the table reports Cochran's Q-test. The test is significant, suggesting the presence of heterogeneity between studies thus answering $Q_{1a}$. It also confirms the assumption outlined above where existing literature was reviewed.

**Table 4.1: Descriptive statistics, income-democracy effects**

| | (1) All-set | (2) Average-set | (3) All-set, *OLS FE* | (4) All-set, *Endogeneity* |
|---|---|---|---|---|
| Unweighted average | .082 | .082 | .033 | .054 |
| Weighted average (N) | .057 | .060 | .019 | .059 |
| Weighted average (FE) | .041 | .049 | .005 | .056 |
| Weighted average (RE) | .082 | .081 | .033 | .055 |
| 95% CI (N) | .048 – .066 | .053 – .066 | .010 – .029 | .041 – .077 |
| 95% CI (FE) | .041 – .041 | .049 – .049 | .005 – .005 | .056 – .056 |
| 95% CI (RE) | .075 – .089 | .076 – .086 | .027 – .040 | .039 – .070 |
| Cochrans's Q-test | p = .000 | p = .000 | p = .000 | p = .000 |
| Nr. of studies | 33 | 33 | 14 | 13 |
| *n* | 492 | 33 | 186 | 151 |

Note. N = A weighted average calculated by the sample size; FE = uses fixed effects assumptions to estimate confidence intervals; RE = uses random effects assumptions to estimate confidence intervals.
Q-test for heterogeneity, $H_0$ = no heterogeneity

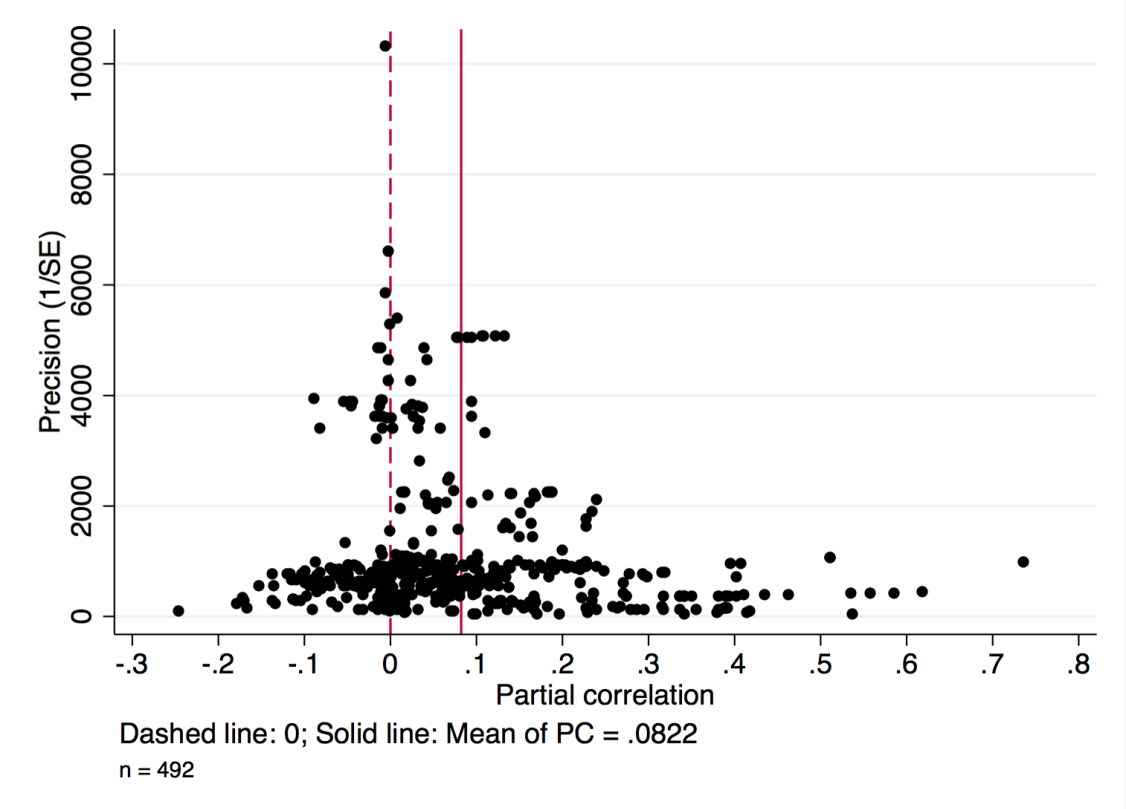
## 4.2 Publication bias

However, these averages are premature, according to Stanley & Doucouliagos (2012), because they do not control for publication bias or heterogeneity. Therefore the inference drawn from this table should remain conservative. Where there is important heterogeneity and potential publication bias, any measure of an average effect size will not capture the true nature of the phenomenon in question. Thus, advanced methods are crucial to further investigate publication bias and heterogeneity in order to answer whether there is an effect of income on democracy. In the following paragraph these analyses are carried out.[41]

[41] Several other samples of study-estimates could have been constructed, including more specific research designs than only methodological variation. These averages are premature, as noted; therefore I will use the multiple meta-regression analysis in order to investigate the "best practice" research design in the income-democracy context.

Figure 4.2 displays a funnel graph with the empirical estimates of the studies (viz. the partial correlations), plotted against these estimates' precisions, i.e. the inverse standard errors (1/*SE*). Because a measure of the variability of each estimate (1/*SE*) is placed on the vertical axis, those *estimates towards bottom have larger standard errors* and will, therefore, be widely dispersed. In contrast, *the more precise estimates, found at the top*, will be more compactly distributed. In the top of the funnel plot, we find Treisman (2011). His large sample size of some 10.000 observations stands out in terms of precision. The partial correlation of the estimate is also calculated to be slightly negative $(-0.005)$. The rest of the population is found with a precision between 6.000 with a cluster between 50 and 2.000. We also observe a cluster of values between – 0.2 and + 0.4, whilst some outliers report partial correlations spanning from approximately + 0.5 to + 0.7. An estimate from Oskarsson & Ottosen (2010) find the partial correlation between income and democracy to be + 0.737, with what seems to be the average precision. On the negative side we find Acemoglu et al. (2008) with – 0.24 and Moral-Benito & Bartolucci (2012) with – 0.17 with very low precision.

**Figure 4.2: Funnel plot of standard error by partial correlation**




Based on this plot, I can identify the two types of "outliers": Studies with high partial correlations and low precision, and studies low partial correlations and high precision, placed

far away from the vast majority for estimates in the funnel plot. The former can be kept with little or no harm to the results, since precision is used as weights in the multiple meta-regression analysis.[42] These leverage points in the data are not a result of coding error; they must therefore be retained because they are genuinely informative about the literature in question. Further, such "outliers" are in fact not outliers, but rather influential data points. Unless a valid and independent reason for the removal of these leverage points can be found, they will have to be kept in the data set. Robust meta-regression techniques will nevertheless minimise the undue influence if only one, or few values in the literature (Stanley & Doucouliagos 2012).

Moving to the interpretation: The funnel plot seems symmetrical; the plot has the shape of an inverse funnel, indicating no publication bias in the literature, but the estimates are somewhat skewed towards reporting positive results, which may suggest the presence of some publication bias.[43] In line with the theoretical expectations derived from the literature review, the scholars who study this phenomenon tend to report results varying from positive to non-existing association. The funnel plot also substantiates the Q-test from the table above, indicating that there is heterogeneity present that causes scholars to report varying results. This is also in line with the theoretical expectations. The problem with this interpretation, however, is that a simple visual interpretation might be biased in itself, as noted earlier. We must therefore use objective statistical tests that correspond to this funnel graph. These meta-regression analysis tests can identify publication selection and a genuine effect beyond publication selection, should it exist (Borenstein 2009; Stanley & Doucouliagos 2012). In the following section, I will conduct a funnel asymmetry test (FAT) and precision-effect test (PET) to determine whether the inference drawn from the funnel plot and the weighted means are robust to objective statistical testing.[44]

[42] Especially studies with high precision could be a sign of coding error. After a reassessment of the entire meta-data set was conducted, I confirmed that all estimates are coded properly. Neither precision nor partial correlations were subject to coding errors.

[43] For examples of other interpretations of funnel plots, see: Doucouliagos & Laroche (2003), Stanley (2005), Doucouliagos & Stanley (2008), Doucouliagos & Paldam (2008), Doucouliagos & Ulbasoglu (2008), Feld & Heckmeyer (2011) and Ahmadov (2014).

[44] As noted, the meta-analysis is not concerned with "counting" significant results (vote counting), but the over-all state of knowledge. Vote counting is not recommended, a discussion on this found in Stanley & Doucouliagos (2012) and Borenstein (2009), for instance.

### 4.2.1 Funnel-asymmetry-test and precision-effect-test meta-regression analysis

In Table 4.1, I report the results of the meta-regression model (Equation 1.4.). First, I employ a funnel-asymmetry-test to identify the presence of publication selection bias. $H_0: \beta_1 = 0$ serves as a test of whether or not there is publication bias; if we cannot reject $H_0$, we can infer with that there is no publication bias. As mentioned, this is because the regression model fits a line to the funnel plot, and if the line has an upward or downward slope we can infer that there is publication bias in the literature. Since this test examines whether the funnel graph is asymmetric, it is named funnel-asymmetry test (FAT). Further, the precision-effect testing (PET) checks $H_0: \beta_0 = 0$. This serves as a test of whether or not there is a genuine underlying empirical effect beyond the potential distortion due to publication selection. If $H_0$ is not rejected, the regression model is not able to detect a genuine effect of income on democracy when controlling for possible publication bias (Stanley 2005; Stanley 2008).

In the table above I have fitted four models using weighted least squares estimation with robust (WLS R) and cluster robust standard errors (WLS CR) clustered by each individual study, as suggested by Stanley & Doucouliagos (2012). The WLS models are chosen given the relatively large span of the partial correlations extracted from the studies producing heteroskedasticity. In this context, little weight is given to coefficients with large standard errors, while precise estimates are given a larger weight (Stanley & Doucouliagos 2012). The two latter models, fixed effects panel with dummy variables for each author, clustered by study (FE) and mixed effects multilevel (MEML) are also reported. In the three latter models I have allowed the partial correlations to be nested within studies. Clustering should be used whenever multiple estimates are coded per study, due to author dependency. The last model is a robust regression to accommodate possible problems with outliers (Doucouliagos & Paldam 2008; Doucouliagos & Paldam 2013).

In Table 4.2, the first row identifies the estimation method used, as described above. The two following rows are, respectively, funnel-asymmetry-test and the precision-effect test. The two last rows identify the number of observation and number of studies included. Notice that all coefficients are derived from the *all-set* data. Moving to the substantive interpretation, we notice that the funnel-asymmetry-test (FAT) term in row two in all of the models is significantly larger than zero (hence reject $H_0: \beta_0 = 0$), but the bias is too small to make a practical difference. We therefore conclude that the graphical inspection of the funnel plot (Figure 4.2) is correct: The funnel plot is symmetrical, but somewhat skewed towards publishing positive results. This conclusion is drawn based on the guidelines for interpreting

the FAT-term (as cited above). Doucouliagos & Stanley (2013: 320) state that if "FAT $|\beta_1|$ <1.0, [then] selectivity is 'little to modest'". In a meta-analysis of union-productivity literature Doucouliagos, Laroche & Stanley (2005) found publication selection to be "little to modest" ($\beta_1$= 0.56), and concluded that publication bias is not, while statistically significant, practically significant. Since $\beta_1$ varies from positive 0.077 to 0.179, depending on estimation method, *I can therefore conclude that publication bias is not a problem in the income-democracy literature*.

Furthermore we see that the precision-effect-test term is not statistically significant across models (recall we are testing $H_0$: $\beta_0$= 0). This implies that there is no statistical evidence of any effect of income per capita on democracy when publication bias is controlled for (accept $H_0$: $\beta_0$= 0). The estimate of $\beta_0$ is a correct estimate of empirical effect, but it has its problems. When there is no effect, it is biased upwards, in magnitude, and when there is an effect, it is biased downwards (Doucouliagos & Stanley 2008). Consequently, we are better off just assuming that the effect is zero if a research area fail to pass the PET (i.e. we cannot reject $H_0$: $\beta_0$= 0). The genuine effect of income on democracy will also be addressed in the multiple meta-regression analysis where both publication bias and heterogeneity is taken into account making the conclusion even more robust.

**Table 4.2: Funnel-asymmetry-test and precision-effect-test meta-regression Analysis**

| | WLS Robust | WLS CR | FE | MEML | Robust |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| FAT | 0.179 | 0.179 | 0.083 | 0.077 | 0.063 |
| Intercept: $\beta_1$ | (4.97) | (2.36) | (15.7) | (4.36) | (9.00) |
| PET | 0.122 | 0.122 | -0.77 | -0.050 | -6.25e |
| $1/SE_i$: $\beta_0$ | (0.08) | (0.02) | (-0.71) | (-0.05) | (-1.46) |
| Nr. of obs. (*n*) | 492 | 492 | 492 | 492 | 492 |
| Nr. of studies (*k*) | 33 | 33 | 33 | 33 | 33 |

*Notes*: The brackets hold *t*-ratios, using robust standard errors, clustered standard errors or robust regression, respectively. FAT measures the degree of publication bias. PET measures the effect of income on democracy corrected for publication bias. Model 1 and 2 uses weighted leas squares (WLS) regression using precision squared as weights with robust standard errors Model 2 reports cluster-robust (CR) standard errors clustered by each individual study. Model 3 uses fixed effects panel with study dummies (FE) and Model 4 uses mixed effects multilevel (MEML). Both are clustered by study. Model 5 is a robust regression

To summarise the findings of the funnel-asymmetry-test and precision-effect-test meta-regression analysis, I have shown that there is some publication selection bias in the literature, but not of practical importance, and we fail to detect an empirical effect of income

on democracy. Since the precision-effect-test does not reach statistical significance, i.e. income has no effect on democracy when controlling for publication bias, it is neither necessary, nor recommended, to regress the FAT-PET-PESSE MRA to address the genuine empirical effect (recall Equation. 1.5.). However, it is important to note that these tests do have their weaknesses. The FAT is known to have low power (Egger, Schneider & Minder 1997; Stanley 2008). I will therefore also control for publication bias in the multiple meta-regression analysis to ensure the robustness of the findings of Table 4.2. The PET is usually powerful enough, but it can have inflated Type I-errors and mistakenly detect effects that are not there (Stanley 2008). These inflated Type I-errors tend to occur when the degree of unexplained excess heterogeneity in the manalysis model is high (Stanley & Doucouliagos 2012). Put simply, this means that the Achilles heel of the PET is that it can find a genuine effect too often, but since these results run contrary to this bias we assume that the results are valid.

A more general critique could be that it is impossible in empirical sciences to make inferences about unobserved phenomenon, and this may be the case when dealing with publication bias. If some findings, controversial to the scholarly field are put in the drawer instead of published, that is something the scholarly field may never be able to observe. Thus it is unmanageable to produce a sound conclusion about something that one might never observe empirically. However, due to the polarisation of the funnel plot and the Funnel-Asymmetry-Test, I am able to conclude that since the estimates from the individual regressions that have been published have varying results, there is no reason to expect that some findings are systematically excluded from the literature. This is the basic assumption behind these approaches, and it enables the meta-analysis investigate something that we might never be able to observe. Nevertheless, there might be some studies in the discipline that never see the light of day; but this remains mere speculations.

It is also of paramount importance that we conduct a multiple meta-regression analysis to try to explain the excess heterogeneity, as emphasised earlier. The funnel plot and the Cochran's Q test do, as noted, reveal notable heterogeneity across studies, thus answering $Q_{1a}$ and $Q_{1b.}$ In order to why the scholarly community is unable to reach consent about income's effect on democracy ($Q_2$) and to further assess the robustness of the simple FAT-PET-MRA, indepdendet variables, i.e. study characteristics, are included in the multiple meta-regression analysis below. When both publication bias and heterogeneity is controlled for, the main research question concerning the level of association between these variables can be evaluated.

## 4.3 Multiple meta-regression-analysis of income-democracy literature

In this section I will expand the equation from the FAT-PET-MRA to include variables that not only captures publication bias, but also control for heterogeneity between studies. In other words, I specify a regression model where the explanatory variables are those that are hypothesised to explain the variation in findings across my sample of studies. I include the independent variables from Table 3.2, which I reproduce below.

### Table 3.2: Description of variables (reproduced)

| Variable | Description | Variable | Description |
|---|---|---|---|
| **Essential variables** | | **Knowledge effects** | |
| Partial correlation | Partial correlation between democracy and income | Income focus | BV: 1 = Income-democracy nexus is primary focus |
| Standard error | Std. err. of partial correlation | Political scientist | BV: 1 = Researcher is political scientist |
| | | Unpublished | BV: 1 = The study is an unpublished work |
| **Estimation differences** | | | |
| OLS FE | BV: 1 = Uses OLS with fixed effects, & AR1 | Prior | BV: 1 = If the first author has published in field previously |
| Endogenous | BV: 1 = Democracy is endogenous | Publication Year | Year of publication with 1999 as base |
| **Time and region** | | **Specifications differences** | |
| Annual | BV: 1 = Uses annual observations | FHI | BV: 1 = Uses Freedom House Index (Ref. Polity) |
| Five year | BV: 1 = Uses five year pooling of observations | Maddison | BV: 1 = Uses Maddison data (Ref. World Bank) |
| Pre1950 | BV: 1 = Uses pre 1960 | Penn | BV: 1 = Uses Penn World tables (Ref. World bank) |
| Countries | Number of countries used | Education | BV: 1 = Includes education control |
| OECD | BV: 1 = Includes OECD countries | Population | BV: 1 = Includes population control |
| No socialist | BV: 1 = Excludes socialist countries | Inequality | BV: 1 = Includes inequality control |
| | | Oil | BV: 1 = Includes oil control |

BV: Binary variable. OLS: Ordinary least squares. FE: Fixed effects AR1: Autoregressive term. OECD: Organisation of economic co-operation and development

In Table 4.3, I have used Equation 1.6, and given $Z = 1…20$ which are the independent variables, and $K$ = SE is the variable for publication bias. Thus, the $Z$-variables are related to heterogeneity and the $k$-variable is related to publication bias.[45] In this equation

[45] If the presence publication bias was severe, several $k$ variables must be included in the form of interaction term hypothesis that $SE$ could interact with all $z$-variables. Since publication bias is not a problem, this specific feature of the multiple MRA is not used. See Stanley & Doucouliagos. (2012) for discussion and examples.

I have used precision squared of each study as weights, i.e. the squared inverse standard error ($1/SE^2$) of each partial correlations to account for the inherent heteroscedasticity. To account for author dependency, and thus also heteroscedasticity, robust standard errors cluster by studies are also estimated, as in the FAT-PET-MRA.

**1.6 (reproduced)**

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

This equation is the starting-point of the analysis: In Model 1 I have estimated a weighted least squares (WLS) model with robust standard errors with the full set of possible explanatory variables. Model 2 also uses WLS, but with cluster-robust standard errors (CL) clustered by study. This is done to account for author dependency, which is acute when multiple estimates from the same study are included in the meta-data. Model 2 is similar to Model 1, and will produce corresponding beta-coefficients. However, the standard errors are computed in a manner to account for any potential dependence among the estimates within the specified cluster, studies in our case, which gives a more conservative significance test (Stanley & Doucouliagos 2012).

To accommodate the standards in empirical science, several models with fixed effects and random effects are included to ensure robustness in the findings. Models 3 and 4 are, respectively, fixed effects regression with estimates clustered by study (FE) and mixed (random) effects (MEML) multilevel models. In the FE model I have, in addition to using cluster-robust standard errors, included author dummies for each study.[46] In the MEML the partial correlations are assumed to be nested within studies.

I now proceed to discuss models 5-8. In these models, I have used the same estimation methods, but the variables are selected though the general-to-specific (G-to-S) procedure described above using WLS with robust standard errors. In these models we move from a general model to a specific one in a statistically valid manner. As mentioned, this approach consists of removing variables one at a time based on statistical significance, in a completely instrumental way. Using a two-tailed test and a significance level of 10%, the G-to-S approach resulted in 10 statistically significant *Z*-variables. The *K*-variable for publication bias is only significant at the 10% level, and is included to control for potential publication bias throughout the analysis.

[46] These author-dummies are significant in a joint significance test (p = 0.000).

In terms of a preferred model, I refrain from taking any stance and focus on the dimensions that are robust across models. Fortunately, there are objective statistical tests that can guide us: First of all the Breusch-Pagan/Cook-Weisberg test for multicollinearity suggest that the modes are not homoscedastic ($p$=0.000), therefore a standard OLS approach to the data is not appropriate. This is in accordance with our theoretical expectations discussed above. Since OLS is neither theoretically nor statistically feasible estimation method, the models have been excluded in favour of the WLS, fixed effects and mixed effects models. Recall that Stanley & Doucouliagos (2012a, 2013b) advocate for the use of WLS models, but are also agnostic in terms of choosing a *real* meta-regression model. As noted, there are statistical test that are objective when choosing estimation technique. This is due to concerns of dependencies among estimates within studies, the best econometric practice would suggest that we use either cluster-robust or a panel model. The Breusch-Pagan Lagrange multiplier test shows that there is clear evidence of study level effects (58.85; $p < 0.0001$). The question now becomes: Are these effects fixed or random? The Hausman test gives a clear rejection of the mixed effects meta-regression analysis in factor of the fixed-effects version (262.56; $p < 0.0001$). Recall that all "random-effects" models assume that the random effects are uncorrelated with the independent variables. However, publication selection bias is likely to make selected random effects correlated with the standard error, which is one explanatory variable needed to be investigated in any meta-regression analysis application.[47] Regardless of estimation procedure, Stanley & Doucouliagos (2012) recommend that the "meta-analysts focus on those results that are consistent across the multiple WLS, fixed effects, and cluster-robust MRAs" (Stanley & Doucouliagos 2012: 103). For a visualisation of this entire process described above, see Figure 7.7 in the appendix.

[47] See Feld and Heckemeyer (2011) for a comprehensive discussion.

## Table 4.3 Multiple meta-regression analysis of income-democracy research

| **Heterogeneity ($Z$-variables)** | **WLS R Full (1)** | **WLS 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 ($\beta_1$) | -0.146 | -0.146 | 0.413 | -0.052 | -0.027 | -0.027 | 0.163 | 0.029 |
| | (-2.387) | (-0.726) | (2.077) | (-0.735) | (-0.712) | (-0.859) | (1.636) | (0.598) |
| **Estimation differences** | | | | | | | | |
| OLS FE | -0.183 | -0.183 | -0.141 | -0.129 | -0.135 | -0.135 | -0.141 | -0.121 |
| | (-6.291) | (-2.603) | (-2.520) | (-6.826) | (-2.553) | (-3.313) | (-2.521) | (-6.299) |
| Endogeneity | -0.257 | -0.257 | -0.147 | -0.132 | -0.176 | -0.176 | -0.147 | -0.127 |
| | (-8.589) | (-2.386) | (-2.234) | (-6.864) | (-3.198) | (-3.519) | (-2.220) | (-6.660) |
| **Time and region** | | | | | | | | |
| Annual | 0.0785 | 0.0785 | 0.0249 | 0.0150 | 0.053 | 0.053 | 0.039 | 0.034 |
| | (2.886) | (1.850) | (0.606) | (0.729) | (1.665) | (2.184) | (1.113) | (1.855) |
| Five year | 0.099 | 0.099 | -0.0112 | 0.001 | | | | |
| | (4.099) | (1.674) | (-0.908) | (0.064) | | | | |
| Pre 1950 | 0.0675 | 0.0675 | 0.0262 | 0.0372 | | | | |
| | (2.349) | (2.282) | (1.322) | (2.204) | | | | |
| Countries | 0.001 | 0.001 | 0.0001 | 0.0001 | | | | |
| | (1.985) | (0.663) | (0.196) | (0.939) | | | | |
| OECD | 0.128 | 0.128 | -0.060 | 0.033 | 0.186 | 0.186 | -0.059 | 0.067 |
| | (2.583) | (1.474) | (-0.914) | (0.777) | (2.854) | (2.707) | (-1.034) | (1.523) |
| No socialist | -0.102 | -0.102 | 0.0100 | 0.0281 | | | | |
| | (-2.491) | (-1.835) | (0.257) | (1.180) | | | | |
| **Knowledge effects** | | | | | | | | |
| Income focus | 0.047 | 0.047 | -0.204 | 0.004 | | | | |
| | (1.056) | (0.637) | (-2.422) | (0.180) | | | | |
| Political scientist | 0.115 | 0.115 | -0.0609 | 0.0480 | 0.152 | 0.152 | -0.083 | 0.033 |
| | (5.345) | (2.522) | (-0.752) | (1.641) | (2.825) | (2.423) | (-1.387) | (0.977) |
| Unpublished | 0.0459 | 0.0459 | -0.143 | 0.127 | | | | |
| | (1.016) | (0.610) | (-2.635) | (3.290) | | | | |
| Prior | 0.0296 | 0.0296 | 0.0318 | 0.0554 | | | | |
| | (0.835) | (0.332) | (0.946) | (1.586) | | | | |
| Publication year | 0.006 | 0.006 | -0.022 | -0.0001 | | | | |
| | (2.229) | (0.690) | (-1.391) | (-0.0348) | | | | |
| **Specification differences** | | | | | | | | |
| FHI | 0.113 | 0.113 | 0.035 | 0.038 | 0.108 | 0.108 | 0.038 | 0.037 |
| | (5.157) | (1.711) | (1.133) | (2.626) | (2.820) | (2.357) | (1.273) | (2.482) |
| Maddison | 0.142 | 0.142 | 0.469 | 0.011 | 0.166 | 0.166 | 0.339 | 0.0504 |
| | (3.405) | (1.233) | (4.678) | (0.244) | (3.319) | (3.546) | (16.33) | (0.945) |
| Penn | -0.007 | -0.007 | 0.075 | 0.042 | | | | |
| | (-0.165) | (-0.0914) | (1.295) | (1.264) | | | | |

*Continues*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Education | -0.0984 | -0.0984 | 0.0199 | 0.0337 | -0.081 | -0.081 | 0.007 | 0.0071 |
| | (-1.006) | (-1.308) | (0.764) | (1.134) | (-1.851) | (-1.225) | (0.286) | (0.232) |
| Population | -0.105 | -0.105 | 0.000831 | -0.0178 | -0.122 | -0.122 | 0.0021 | -0.003 |
| | (-1.112) | (-1.692) | (0.0502) | (-0.660) | (-3.168) | (-1.851) | (0.175) | (-0.118) |
| Oil | -0.173 | -0.173 | -0.0130 | -0.0131 | -0.179 | -0.179 | -0.012 | -0.019 |
| | (-11.83) | (-6.573) | (-2.321) | (-0.557) | (-3.798) | (-8.120) | (-2.518) | (-0.763) |
| **Publication bias (*K-variable*)** | | | | | | | | |
| SE | -3.606 | -3.606 | -0.708 | 0.695 | -2.953 | -2.953 | -0.396 | 0.186 |
| | (-5.273) | (-1.203) | (-0.454) | (0.561) | (-1.820) | (-1.607) | (-0.343) | (0.176) |
| $R^2$ | 0.642 | 0.642 | 0.587 | - | 0.612 | 0.612 | 0.584 | - |
| $MLR^2$ level 1 | - | - | - | 0.345 | - | - | - | 0.176 |
| $MLR^2$ level 2 | - | - | - | 0.543 | - | - | - | 0.229 |
| F-test | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| No. obs (*n)* | 492 | 492 | 492 | 492 | 492 | 492 | 492 | 492 |
| Nr. Studies (*k)* | 33 | 33 | 33 | 33 | 33 | 33 | 33 | 33 |

Notes: *t*-statistics in parentheses. Dependent variable is partial correlations (PC). Model 1 is a weighted least squares (WLS) regression with robust standard errors where all control variables is included, using precision squared as weights ($1/SE^2$). Model 2 is estimated with WLS cluster robust standard errors clustered by studies with precision squared as weights, using all explanatory variables. Model 3 and 4 are, respectively, fixed effects panel (FE) and mixed effects multilevel regression (MEML) where partial correlations are assumed to be nested within studies. Models 5-8 use the same estimation methods as 1-4, respectively, but are modelled after the general-to-specific (G-to-S) approach advocated by Stanley et al. (2012). $MLR^2$ is Snijders & Bosker (2012) $R^2$ for MEML models.

From inspecting table 4.3, we see that the coefficients in all models are highly consistent. In terms of size they vary somewhat, depending on estimation technique, but we can only observe few insignificant sign-changes, i.e. a coefficient changes sign from + to – orvice versa. We can also observe that the WLS robust and WLS cluster-robust models display very similar coefficients, which is in accordance with the discussion above. The FE and MEML models report more conservative results, but effects are mostly consistent throughout. In the full model, many of the coefficients do not reach statistical significance, which may be due to high multicollinearity commonly associated with multiple meta-regression analysis (Doucouliagos & Paldam 2008). A VIF test of Model 1 confirms that multicollinearity is not a problem in this model; thus we can trust the general-to-specific approach.[48]

As mentioned, the G-to-S approach identifies 10 independent variables that are significant on the 10%-level. This is of course a very liberal level of significance and when the interpretations of the effects I will focus of results significant at the 1%-level. Notice, for instance, that the *SE* coefficient is significant at the 10%-level in model 5, but not significant in the other models. The same goes for the *Annual* coefficient. Further, the variables that

[48] Mean VIF = 8.60. Midtbø (2012) states, as a rule of thumb, that a mean VIF below 10 indicated that multicollinearity is not a problem.

measure *Estimation differences* are both significant across models. Two *Time* and *region* variables, *Annual* and *Western,* are also significant. *Specification differences* also create heterogeneity across studies, *FHI, Maddison, Education Population* and *Oil* is identified as important dimensions by the G-to-S approach. All these variables, including the intercept, are also jointly significant and show that the multiple meta-regression analysis identifies genuine, systematic patters among reported findings in the literature.[49]

Before continuing, it is important to note that I confirm the findings from the bivariate FAT-PET-MRA that publication bias is not a problem in the literature when controlling for heterogeneity. The coefficient does not reach statistical significance in any of the models, regardless of estimation technique. In terms of explanatory power, this multiple meta-regression analysis preforms quite well. The $R^2$ varies around approximately 0.60 which means that 60% of all the variation in the literature is explained. This is somewhat larger than the average of what Nelson & Kennedy (2009) and Stanley & Doucouliagos (2012) find is the most commonly reported $R^2$, namely ca. 50%. The multilevel $R^2$ finds that 17% of the variance on study-estimate level, and 23% of the study-level variance is explained

In the following section I will make a substantive interpretation of the models produced in table 4.3, with focus on models 5-8, but first a note on how to interpret the multiple meta-regression analysis: *When one of the independent variables is significant and negative, or positive, the coefficient indicates that the variables increases (reduces) the size of the measured partial correlation, which, in turn, increases (reduced) the effect of income of democracy* (Costa-Font, De-Albuquerque & Doucouliagos 2014). *If an independent variable is not statistically significant, this indicates that the specific study characteristic fail to explain the heterogeneity of the empirical findings* (Adam et al. 2013).

### 4.3.1 Estimation differences

We can see from examining the two coefficients on estimation techniques that all the *OLS FE* and *Endogeneity* coefficients are negative, of consistent size, and consistently statistically significant across all models. These two factors, which characterise heterogeneity across studies, are very important. *Thus, when both variables are given the value zero – which signifies studies wherein estimates are computed with random effects, without an autoregressive term, etc. – scholars find a larger positive association between income and democracy.* To repeat: These two sources of heterogeneity are meaningful influences upon

[49] $F(11, 480) = 17.08$; $p < 0.0001$.

whether one discovers a larger or small, more or less insignificant, effect of income upon democracy – utilising fixed effects AR1 and/or endogeneity specifications reduces the partial correlation between income and democracy, wherein endogenous models have the larger effect of the two.

As noted, the *Endogeneity* variable is also negative and significant across all models. There is a sizable literature on treating democracy as endogenous, but as the literature review suggests, using OLS with fixed effects and an autoregressive term is not adequate. Hence, estimations that stem from the use instrumental variables, GMM, etc. are the effect captured by the *Endogeneity* variable. However, in some cases, authors have argued that the methodological approaches used in, for instance, Acemoglu et al. (2008) merit criticism. Moral-Benito & Bartolucci (2012), Che et al. (2013) and Heid, Langer & Larch (2012) find that some of the instruments and estimation techniques used therein utilise invalid instruments. Hence, rather than suggesting that failing to control for endogeneity overstates the income-democracy effect, the meta-regression analysis may be detecting the findings of a literature that has not adequately modelled the endogeneity of democracy (Doucouliagos & Ulbasoglu 2008). At this point, we simply do not know; because: (I) We are unsure whether failing to control for endogeneity necessarily creates bias, and; (II) Whether those who do control for endogeneity perform it accurately and hence whether their (quite different) estimates are more or less close to the "true" coefficient.

### 4.3.2 Time and region

In the specification of Models 5-8 above, we see that amongst the variables within the "Time and Region" category, *Annual* and *OECD* are the only ones *not* excluded by the G-to-S approach. In the full models, we see that some other variables are significant, but as the exclusion process proceeds, the variables are removed. Take, for example, the variable *Pre 1950,* which indicates whether the researcher examines a longer time series of democracy, extending back in history. Models 1-4 indicate borderline statistical significance of this variable with a positive coefficient (indicating that such an approach is associated with finding a larger positive effect of income on democracy). However, the G-to-S procedure results in the exclusion of this variable on the basis of inadequate influence. This would imply that the claims of, for instance Boix & Stokes (2003) and Acemoglu et al. (2008) about the long-run effects of democracy are not supported by the findings here based upon these meta-data. Their argument about omitted historical variables may indeed be valid, but this meta-

analysis provides no alternative explanatory theory concerning whether or not there is a meaningful differences in results between scholars looking at post-1950 contra pre-1950.

Moving on to the *Country* variable, we see that the number of countries included in the analysis does not create systematic differences either. This coefficient is barely statistically significant in the full models, and the effect is undistinguishable from zero. This may reflect the growing use of the fixed effects specification in the research field. In sharp contrast, we see that the *OECD* variable is significant: Including OECD countries seem to produce significantly larger partial correlations, as illustrated by the WLS Models 5 and 6. By contrast, Models 7 and 8 produce different findings on this variable and thus create some contradictions which would require further work to disentangle. The variation could stem from the fact that the fixed effects estimator assumes that all collected partial correlations are drawn from a population with the same mean, as discussed above.

The variable *Annual* is coded one for those studies with annual data (as opposed, for example, to five-year averaged data). We see that such studies with annual data find a stronger relationship between income and democracy. However, this finding is significant on at the 10% level, and across models. The results thus suggest that using annual data does not have a significant effect on partial correlations. The variable *No socialist* indicates that excluding socialist countries does not have a significant effect either, although the negative sign of the coefficient does suggest that the exclusion of socialist countries has a negative effect on partial correlations. This effect is not robust across models, and the change from a negative sign in Models 1 and 2 to a positive sign in Models 3 and 4 is not consistent.

### 4.3.3 Knowledge effects

Attributes of the specific authors and particular studies do not seem to produce systematic differences. Take the *Income focus* variable, for example. We see that studies which are explicitly focused upon the income-democracy relationship (as opposed to those which focus on other variables but nonetheless include an income variable in their model specification) do not create excess variation in the data. The slightly positive coefficient of the *Income focus* variable suggests that scholars focusing on the income-democracy relationship would seem to report a more substantial association between income and democracy, but the coefficient does not at any point reach statistical significance and may therefore be ruled out as an explanatory variable. The variable *Prior,* which measures whether the first author has made previous contributions to democratisation research, show a positive coefficient, but is robustly

insignificant across models, and is therefore not included in the G-to-S models. The variable *Unpublished* measures whether the article or book is published in a peer-reviewed journal or otherwise – it too has low explanatory power. However significant it appears to be in Model 1, its effect was phased out by the G-to-S approach and thus is not included in the final model confirming that there is no publication bias in the literature. The effect of the coefficient is nonetheless very small, and subject to a sign reversal in the fixed effects model (Model 3).

The *Publication year* variable turns out not to be a source of heterogeneity either. This variable is commonly included in meta-analyses as a means for accounting for potential trends or path dependencies in research (Stanley & Doucouliagos 2012). Model 1 produces a small positive, significant coefficient, but this variable is eventually shown not to matter; hence, whether a study is published in this decade or a previous one does not matter for the outcome. This seems contra-intuitive if we refer back to Table 4.1, but the varying average effect over the last decade may contribute to the linear prediction of this variable.

The variable *Political scientist* show that studies published by political scientists produce more positive estimates than those produced by economists, according to models 4 and 6. This heterogeneity can be contextualised in the discussion above. Acemoglu et al. (2009) argued that the economic theory of critical junctures was a more feasible theory for explaining the relationship between income and democracy. This may confirm the discipline bias that Adam et al. (2013) found in the welfare state literature – and such bias seems to travel to the research on determinants of democratisation as well.

### 4.3.4 Specification differences

In this section, I find that several significant effects that are consistent across all the models. First, we notice in the *FHI* variable that using the Freedom House Index (FHI) produces significantly more positive results than using Polity (which, according to our discussion above is a more valid measure). Further, I find evidence that the choice of data set for income matter, as Krieckhaus (2004) argued. In our case, using *Maddison* data produce a significantly more positive effect between income and democracy, as compared to Penn world tables and World Bank data. The variable for *Penn* World Tables is slightly negative in the WLS models, but changes in the fixed effects and mixed effects models. Nonetheless, the coefficients never reach statistical significance and are thus excluded from models 5-8.

Looking at the control variables, I find some evidence to indicate that when studies control for *Education, Population* or *Oil* the results are affected. However, if we compare

results across the models, we see that we cannot infer that the effect of *Education* is robust, and neither is the *Population* variable. The *Oil* variable, however, is significantly negative in Models 5-8. This indicates that controlling for oil has a negative effect upon the chance that one will find a sizeable, positive significant effect of income upon democracy. Hence confirming education and population controls do not appear to have a specific effect as cofounder variables, while oil does. Ahmadov (2014) finds that oil has an over-all negative effect on democracy.[50] He also introduces modernisation control variables, but finds no evidence that oil hinders income per capita, which in turn, hinders democracy. In this thesis, however, I find robust evidence that supports the finding of Ahmadov (2014) that oil, in fact, hinders democracy indirectly via per capita income.

### 4.3.4 Publication bias

The G-to-S method determined that the effect of the publication bias variable *SE* (viz. the standard error of the partial correlation) is significant enough to keep in the model specification, although it is significant only at the 10%-level ($t = -1.82$; $p > 0.05$) in Model 5. In the Model 1, the coefficient is negative and highly significant, yet as the selection process moves forward, the results confirms the finding from the bivariate the funnel plot and funnel-asymmetry-test: That publication bias does not influence the literature. In an early phase of the research process, several *K*-variables were included as interaction terms with the other moderator variables, *Z*-variables. This yielded non-conclusive results, and created model bias due to the lack of publication bias in this literature. The funnel plot shows that studies with low precision sometimes report negative insignificant partial correlations, (recall that studies with high precision often report results that are indistinguishable from zero. Besides all this, the literature is not biased by publication selection). This observation is important, especially since Doucouliagos & Paldam (2013) claims several research areas in political science and economics are "haunted" by publication bias, and that publication bias is the norm in empirical studies (Stanley 2008). Hunter & Schmidt (2004) are somewhat more optimistic, and state that publication bias may not exist in some field of research (of which this field of research is an example).

[50] Ahmodov (2014) does not, however, address publication bias in a statistically valid fashion. I have therefore conduced a reassessment of Ahmodov's data (original data received after personal correspondence with the author). My statistical analysis confirms his conclusion from a graphical inspection of a funnel plot that there is no publication bias in the oil-democracy literature. I conducted a funnel-asymmetry-test and found no evidence of publication bias (accept $H_0$: $\beta_1 = 0$; $t = 0.26$; $p > 0.792$).

## 4.4 Predicting the correct effect from the multiple meta-regression analysis

The preceding analyses have now placed me in a position where I am able to address the main research question ($Q_1$) presented at the outset of this thesis: *What is the effect of per capita income on democracy?* But in order to answer this question, we must look further than the coefficients that account for the publication bias and observed heterogeneity in the multiple meta-regression analysis. These coefficients help us to understand *why we observe heterogeneity in the range of findings produced across our sample of studies*, thus answering why we observe no consensus in the scholarly field ($Q_2$). But these findings – however interesting and important they are – do not tell us whether income has an effect upon democracy. Rather, these coefficients only they only tell us why some think it does and some think it does not: they allow the researcher to determine the different ways in which choices of research design, estimation method and model specification affect study outcomes.

In order to answer $Q_1$, I proceed in the following manner: I refer the reader back to Table 4.3, statistical Model 6 – it is this model which can be argued to be the "best model", and hence the model which I will utilise to produce the "best estimate" of the effect of income upon democracy. Model 6 may be considered the best suited for computing such predictions because the assumptions behind both the FE and RE model might be flawed in the context of social science, and it addresses the possible author dependency of the data in a valid fashion (Stanley & Doucouliagos 2012, 2013a, 2013b).

In order to produce a "best estimate" of the effect of income upon democracy from the multiple meta-regression analysis, I insert values for all of the variables included in the specification of Model 6. Recall that these variables include the following. In addition to the intercept term, the model includes: (I) A binary variable which indicates whether the study utilised a fixed effects OLS estimation (*OLS FE*); (II) A dummy variable that indicates whether the study corrected for possible endogeneity bias (*Endogeneity);* (III) A binary variable that indicates whether the study analysed annual data as opposed to data where more than one year were averaged (*Annual*); (IV) A variable which indicates whether the study included OECD countries in its sample (*OECD);* (V) A variable which indicates whether the author of the study is a political scientist, as opposed to an economist (*Political Scientist)*; (VI) A variable that flags the use of the Freedom House measure as opposed to Polity (*FHI*) measure of democracy; (VII) A variable that flags the use of economic data produced by Madisson as opposed to Penn World Tables or the World Bank (*Madisson*); (VIII) A variable that measures (*Education*); (IX) A variable that indicates whether the study included an *Education* control variable; (X) A variable that indicates whether the study included a

*Population* control variable; (XI) A variable that indicates whether the study included an *Oil* control variable; (XII) A variable that tests for Publication Bias (*SE*).

To compute the predicted effect of income upon democracy, I choose specific values for each of these independent *Z*- and *K*-variables for the fitted statistical model (StataCorp 2013). This allows me to estimate a predicted effect (Pr(outcome)) of the income-democracy nexus when both publication bias ($Q_{1a}$) and heterogeneity ($Q_{1b}$) are controlled for. To minimise any effect from potentially questionable judgment on my part, I use the sample means of each individual variable for generating predictions. I represent this vector of sample means of each respective independent variable by: $\bar{\chi}$.

Note that in order to be able to predict a "best practice" study, we must filter out the possible publication selection bias and make the individual estimates obtain "perfect" precision, which implies that the standard error of the correlation coefficient that measures the relationship between income and democracy is equal to zero, or: $SE = 0$. In short: A study approaches perfection with no estimation error and publication bias.

In the following paragraphs, I will proceed in a series of stages, from one model to another. In each respective stage, I will assign different values to selected independent variables in order to determine whether the finding of the effect of income upon democracy is thereby altered. I do this in order to employ several versions of what the literature has concluded that can be viewed as "best practice" research when studying the relationship between income and democracy. I now proceed to calculate the estimated relationship of income upon democracy.

### 4.4.1 Prediction model 1

In this model, I use the sample means ($\bar{\chi}$) of each independent variable in Model 6 from Table 4.3 to predict the strength of the association between income and democracy, while assuming a "perfect" study wherein the standard error of the estimated correlation coefficient is equal to zero. Hence:

$$\Pr(\text{outcome}) : \text{SE}=0,\ \text{Remaining } Z\text{-variables} = \bar{\chi}$$

Doing so produces an estimate of the income-democracy nexus of +0.182 with a 95% confidence interval (0.16 - 0.20). In Cohens's (1992) terms this is still a small effect, despite the fact that it is larger than what the weighted averages (Table 4.1) and funnel-asymmetry-test and precision-effect-test meta-regression analysis (FAT-PET-MRA, Table 4.2) found. Recall that the weighted averages neither controlled for neither publication bias ($Q_{1a}$) or

heterogeneity ($Q_{1b}$), and the FAT-PET-MRA only took publication bias into account. Thus, the predictions from the multiple meta-regression analysis are best suited for answering the main research question.

### 4.4.2 Prediction model 2

Necessarily, critics will point out that using the sample mean values of the independent variables as a base for predicting the actual effect of income and democracy does not represent "best practice" research. I am also convinced that this is true, and therefore proceeds to Model 2. As Beck & Katz (2001) argue, using OLS with fixed effects and AR1 is the best approach in cross-sectional time-series data (but see also Clark & Linzer (2014); Bell & Jones (2015)). Other scholars claim that GMM, IV or other approaches are more sophisticated statistical methods in terms of modelling endogeneity (Robinson 2006, Faria et al. 2014). Hence, in this second model, all the other variables from Model 6 in Table 4.3 are held at their sample means ($\bar{\chi}$), but the variable *Endogeneity* is given the value 1 and the variable *OLS FE* is given the value 0, because these two categories are mutually exclusive. Recall that we also retain the assumption that $SE = 0$ in order to mimic the *perfect* study in terms of precision and publication bias:

$$\Pr(\text{outcome}) : \text{SE=0, Endogeneity = 1, OLS FE = 0,Remaining Z-variables} = \bar{\chi}$$

This yields a statistically significant, slightly smaller size effect of $+0.166\ (0.02 - 0.311)$.

### 4.4.3 Prediction model 3

Next, in order to code *OLS FE* using AR1 with the value 1, I by necessity code the *Endogeneity* variable 0.

$$\Pr(\text{outcome}) : \text{SE=0, Endogeneity = 0, OLS FE = 1, Remaining Z-variables} = \bar{\chi}$$

In this case, the multiple meta-regression analysis predicts a statistically significant effect of slightly larger size equal to $0.20\ (0.101 - 0.313)$. This latter predicted effect approaches what Cohen (1992) deem to be a medium effect between income and democracy.

### 4.4.4 Prediction model 4

The three predictions presented above might however, be premature, because some specifications might be considered "better" practice than others. Take for instance the discussion above of measuring democracy and measuring income per capita. Based on this discussion, it seems that using the Polity Index is more robust than FHI as a measure for

democracy (nonetheless, not perfect; see Munck & Verkuilen (2002). Contrary to Krieckhaus (2004), I do not find a systematic difference between Penn World Tables and World Bank data set. However, using Maddison data show a systematically more positive effect throughout the literature. And as advised by Krieckhaus (2004), I will use WBD in the prediction (recall that WBD is the reference category of for the *Maddison* variable). Hence, I next generate predictions using a model with OLS fixed effects and AR1, Polity measure of democracy (recall also that *Polity* is the reference category of for the FHI variable) and World Bank Data:

$$\Pr(\text{outcome}) : \text{SE} = 0, \text{Endogeneity} = 0, \text{OLS FE} = 1, \text{ Maddison} = 0, \text{ FHI} = 0, \text{Remaining Z-variables} = \bar{\chi}$$

This yields a predicted statistically significant, but quantitatively very small effect of +0.091 $(\text{CI} = 0.006–0.177)$.

### 4.4.5 Prediction model 5

In this model, I keep the model above, but assume that we include controls for possible endogeneity (which implies that we must code *OLS FE* as 0):

$$\Pr(\text{outcome}) : \text{SE} = 0, \text{Endogeneity} = 1, \text{OLS FE} = 0, \text{Maddison} = 0, \text{ FHI} = 0, \text{ Remaining Z-variables} = \bar{\chi}$$

Now, the meta-regression analysis predicts a statistically insignificant effect of very small size equal to +0.050 $(\text{CI} = -0.071 – 0.173)$, which indeed is a very small practical association.

### 4.4.6 Prediction model 6

I refer the reader back to Model 4, where we assume a fixed effects model (without endogeneity corrections), but now add a control for oil:

$$\Pr(\text{outcome}) : \text{SE}=0, \text{Endogeneity} = 0, \text{OLS FE} = 1, \text{ Maddison} = 0, \text{FHI} = 0, \text{ Oil} = 1, \text{ Remaining Z-variables} = \bar{\chi}$$

This model generates a prediction of a statistically insignificant, quantitatively small effect of –0.049 (CI = – 0.171 – 0.073).

### 4.4.7 Prediction model 7

I refer the reader back to Model 5, where we do *not* assume a fixed effects model, but *do* allow for endogeneity corrections, and include a control for oil:

$$\Pr(\text{outcome}) : \text{SE}=0, \text{Endogeneity} = 1, \text{OLS FE} = 0, \text{Maddison} = 0, \text{ FHI}=0, \text{ Oil} = 1, \text{ Remaining Z-variables} = \bar{\chi}$$

This model generates a quantitatively small, statistically not significant prediction equal to – 0.090 (CI = – 0.248 – 0.067). Models 6 and 7 might be considered to be "best practice" research because it captures much of the concerns outlined in the review of the relevant literature above. Nevertheless, as shown by these predictions, regardless of which of these two statistical estimation methods is actually "better", the predicted effect of income on democracy remains the same: it does not exist; it is of small quantitative magnitude and does not reach statistical significance.

## 4.5 Summary

In this chapter the actual meta-analysis has been conducted in a manner laid out by the chapter on methodology. In short, we can say that this exercise yielded three important results: ($Q_1$) The multiple meta-regression analysis finds *no robust effect of income on democracy*; ($Q_{1a}$) There is not *practically significant publication bias* in the literature, and it has no practical consequences; ($Q_{1b}$) There is study heterogeneity which *does* result in systematic differences across the studies. This specific heterogeneity also explains why we do not observe a consensus across in the field ($Q_2$). The multiple meta-regression analysis identified 10 variables that affect study outcome, and explained 60% of all the variation in the meta-data. As outlined in the literature review, estimation differences, sample composition, attributions of the authors and model specification differences have created the large variations which we observe across studies. The Table 4.4 below sums up the findings of this thesis. In the next chapter I will discuss the findings in light of the literature, address the over-all implication of the results of this thesis, and suggest several paths for further research.

## Table 4.4: Summary of findings

| Research question | Finding | Inference drawn from |
|---|---|---|
| $Q_1$: What is the effect of income per capita on democracy? | The association between democracy and income is indistinguishable from zero when appropriate research design is used | Table 4.1, 4.2, 4.3, all columns, prediction models 1-7 |
| $Q_{1a}$: Is publication selection bias present in the literature? | There is no publication bias in the literature | Table 4.2 and 4.3, all columns |
| $Q_{1b}$: Are there systematic differences between studies, and does it affect the literature? | There is heterogeneity between studies that can account for, and explain, much of the variation and differences of outcomes in published studies. | Table 4.2 and 4.3, all columns |
| $Q_2$: Why is it that we observe no consensus on this question across the scholarly literature? | Disagreement across the heterogeneity-dimensions identified by this thesis leads to the lack of consensus in the field. The list below describes how different dimensions create variation in study-outcome | Table 4.2 and 4.3, all columns |
| **Heterogeneity** | | |
| **Estimation differences** | | |
| OLS FE | Negative affect on PC | Table 4.3, all columns |
| Endogeneity | Negative affect on PC | Table 4.3, all columns |
| **Time and region** | | |
| Annual | Positive affect on PC | Table 4.3, Model 1, 2, 5 & 6 |
| OECD | Positive affect on PC | Table 4.3, all columns |
| **Knowledge effects** | | |
| Political scientist | Positive affect on PC | Table 4.3, model 5 & 6 |
| **Specification differences** | | |
| FHI | Positive affect on PC | Table 4.3, all columns |
| Maddison | Positive affect on PC | Table 4.3, models 1-7 |
| Education | Negative affect on PC | Table 4.3, model 5 |
| Population | Negative affect on PC | Table 4.3, model 5 |
| Oil | Negative affect on PC | Table 4.3, Models 1,2,3,5,6,&7 |

PC: Partial correlation between income and democracy; FHI: Freedom House Index; FE: Fixed effects
OLS: Ordinary least squares; OECD: Organisation of Economic Co-Operation and Development;

# 5. Discussion and conclusion

In this thesis I have investigated the following research question: ($Q_1$) "*What is the effect of income per capita on democracy?*" To investigate this relationship, I have collected 33 studies, with 492 corresponding estimates of this effect in an original data set and used the techniques of the meta-regression analysis to answer the main research question. As the review of the literature showed, after almost 60 years of research, there is still no consensus on income's effect on democracy. Hence, this field needs an examination of what we actually know about income as a determinant of democratisation. Instead of producing yet another analysis of this relationship, I have chosen to examine it in meta-analytical terms to determine the actual direction of this relationship and hopefully guide further research in the right direction. Before being able to answer the main research question ($Q_1$), I used the techniques of meta-regression analysis to address possible publication bias ($Q_{1a}$) and heterogeneity ($Q_{1b}$). To answer these two sub questions, I used the FAT-PET-MRA to show that there was no substantial publication bias in the literature and the multiple meta-regression analysis to show that heterogeneity accounted for 60% of the variation of reported results in the literature. As an extension of this, it was also necessary to deal with $Q_2$: "*Why is it that we observe no consensus on this question across the scholarly literature*?" With respect to this question, this thesis finds that it is heterogeneity across studies which contribute to the lack of consensus in the scholarly field. Moving back now to ($Q_1$), I used multiple meta-regression analysis to compute several "best practice" research designs for investigating the relationship between income and democracy. These predictions, controlling for both publication bias and heterogeneity, show that this thesis finds no robust effect of *income on democracy*. Having said this, I now proceed to discuss the sources of heterogeneity that explains why there is no consensus in the field.

## 5.1 Discussion of the findings

As Stanley & Doucouliagos (2012) note, one of the natural applications of the meta-regression analysis is to test rival theories that differ in terms of direction, magnitude and the nature of the distribution of these effects. In this thesis, I have tested the modernisation hypothesis. However, not even the most comprehensive and rigorous meta-analysis can by itself provide a definitive or sophisticated falsification of a theory – at least not in the Popperian sense. Popper (1959) agues that a hypothesis is only falsified if we discover a

reproducible effect which disproves the theory. In other words, we can *only* accept the falsification of a hypothesis if a second empirical hypothesis, which describes the same effect the first hypothesis is trying to explain, is proposed and verified. This kind of hypothesis may be called a falsifying hypothesis and is not present in the meta-analysis presented here. The main interest in this analysis is therefore to test the most central aspect of the modernisation hypothesis: The meta-analysis has focused on the value of the association between two key variables of this theory, and the practical significance of this effect. Thus, we can conclude that the analysis presented in this thesis confirms the findings of Acemoglu et al. (2008); Acemoglu et al. (2009), and also Przeworski et al. (2000) (within the framework of continuous conceptualisation of democracy), amongst others. At the same time, it is indeed possible that the criticisms (posed by Faria, Montesinos-Yufa & Morales (2014), Bobba & Coviello (2007) and others) of the GMM and IV approaches used in Acemoglu et al. (2009) could be valid – Acemoglu et al. (2008) might be using poor instruments and and misspecified GMM techniques. However, the second methodological variation in the literature investigated in this thesis, fixed effects AR1, still finds that the effect of income on democracy is not distinguishable from zero. Therefore while I am unable to falsify the modernization hypothesis, my data, methods and analysis lead me to conclude that income is not a robust determinant of democratisation. This finding is the central contribution made by this thesis.

### 5.1.1. Estimation method

The discussion of whether fixed effects is one of the most appropriate estimation methods can be recalled from the discussion of the literature above. The argument is the following: A major source of potential bias in a quantitative analysis of democracy's effect on income per capita is country-specific: Historical factors that can potentially influence both political and economic development. The fixed effects approach will remove them as a source of bias potential bias. Acemoglu et al. (2008: 810) use the comparison between USA and Colombia to illustrate this point:

> Consider, for example, the comparison of the United States and Colombia. The United States is both richer and more democratic, so a simple cross-country comparison, as well as the existing empirical strategies in the literature, which do[es] not control for fixed country effects, would suggest that higher per capita income causes democracy.

Furthermore, the idea of fixed effects is to move beyond this comparison and investigate the "within-country" variation. Sticking to the example, fixed effects allows us to investigate

whether Colombia is more likely to become democratic as it becomes richer. This greatly improves inference on the causal effect of income on democracy, and is, in fact, more closely related to the modernisation theory of Lipset (1959), which draws our attention to the expectation that individual countries should become more democratic if they are richer, not simply that rich countries should be democratic. On the contrary, Gonick & Robert M. (1988), Burkhart & Lewis-Beck (1994), Burkhart (1997) Ross (2001) and Herb (2004) used random effects exclusively.[51] Their findings should thus be interpreted as findings about predictions of long-term cross-national variation.

When variation in the independent variable is primarily within units, that is, the units are relatively similar to one another on average, the choice of random versus fixed effects only matters at extremely high levels of correlation between the independent variable and the unit effect. Under these conditions, the appropriate model should be guided by the researcher's goals, and in this context it would always be the variation within units that the researcher would like to investigate. When the independent variable exhibits only minimal within-unit variation there is a more nuanced set of considerations. In any particular dataset, the random-effects model will tend to produce superior estimates of beta coefficients when there are few units or observations per unit, and when the correlation between the independent variable and unit effect is relatively low. Otherwise, the fixed-effects model may be preferable, as the random effects model does not induce sufficiently high variance reduction to offset its increase in bias (Clark & Linzer 2014).

Consequently, when the researcher wants to estimate between-country effects, i.e. random effects, (s)he would want to use random effects specification. In contrast, when a within-country effect is the main focus, the fixed effects specification should be used. In some cases, the between-country effects could be interesting to investigate, but in the context of income-democracy nexus it might be a flawed approach. As the findings in Table 4.3 clearly state, whether studies focus on the income-democracy nexus or not do not generate systematic differences in study-outcomes. The fact that the multiple meta-regression analysis generates these findings might be an indication that most scholars are trying to explain within-country effects in their studies, thus making fixed effects specification the most appropriate econometric technique.

However, as suggested by Robinson (2006), the scholarly literature should also focus on investigating the simultaneous causality between these two variables. Unfortunately, as

[51] The latter two are included in the meta-data.

Przeworski (2010) notes, this may be a near impossible task. Thus, this must be done in a rigorous manner, and not by using weak instruments, because weak IV-approaches can understate or overstate the effect of income on democracy. As seen in Table 4.1 the averages of the partial correlations that use an *Endogeneity* specification are somewhat larger that the averages drawn from the sample that uses *OLS FE*. This may indicate that when instruments are not valid and thus endogeneity is not modelled properly, this produces larger effects (recall that these averages do not control for publication bias or heterogeneity). Therefore, rather than suggesting that controlling for endogeneity understates the income-democracy effect, the multiple meta-regression analysis may be detecting the findings of a literature which unfortunately has not modelled relationship properly. Nonetheless, I again remind the reader that the fixed effects AR1 method finds similar results; therefore, it seems reasonable, in the end, to infer that the approach which controls for endogeneity does not produce severe bias. Also, studies that criticise for instance Acemoglu et al. (2009), such as Faria et al. (2014), and Bobba & Coviello (2007) are included in the analysis – and nonetheless the multiple meta-regression analysis still predicts a non-existent effect of income per capita on democracy.

### 5.1.2 Polity and Freedom House Index

Munck & Verkuilen (2002) argues that the difficulties with the Polity index are less problematic than FHI, and may therefore be the most conservative and rigorous measure of democracy. However, as most authors in this field are aware, the inclusion of several measures of democracy, also dichotomous ones, should be adopted in order to ensure robustness in the findings. In this case, "best practice" research would be to use more robust continuous indicators, like Polity, and also use dichotomous indicators, like ACLP (Przeworski et al. 2000) or BMR (Boix et al. 2013).[52] Scholars not rely on one measure of democracy exclusively, because, as this meta-analysis has shown, they create systematic differences in reported results. The results of the meta-analysis show that authors should report results using FHI exclusively with caution.

[52] See also Knutsen & Wig (2015).

### 5.1.3 Income data

As emphasised by Krieckhaus (2004), the choice of data set used to measure income per capita has an impact on study out-come. In this meta-analysis, however, I fail to discover systematic differences between the two data sets Krieckhaus (2004) identifies, namely PWT and WBD. I am thus unable to confirm his claim after putting his arguments to a more rigorous tests. Yet, a genuine systematic heterogeneity occurs when researchers use Maddison data. In modern day academia there are, to my knowledge, few arguments for choosing one data set over the other. As noted, Levine & Renelt (1991) argue for the use of WBD, but a reassessment of this is needed to determine what actually creates the systematic differences between these data sets. Based on the results of this analysis I suggest that future research should consider the choice of data set carefully (or use several to check for robustness), guided by theoretical reasoning and not based solely on availability.

### 5.1.4 Sample composition

The fact that the variable capturing estimates that are derived from samples that includes the OECD countries is positive show that the association between income and level of democracy increases if the study includes one or more OECD country. Nevertheless, recall the averages which were computed in chapter four: When the entire world is included, the relationship between income and democracy was *still* non-existent. As Coppedge & Midlarsky (1997) argued, the modernisation variables that show an effect could depend greatly on regional samples, and this meta-analysis confirms this finding. Whether this proves that other determinants of democratisation are also affected by sample composition remains uncertain in this meta-analysis, since I have focused exclusively on the relationship between income and democracy. Using a fixed effects specification also controls for this possible sample bias. Consequently, this does not affect the over-all results. Recall also that the variables measuring how many countries were included in the individual study remained insignificant, confirming that when looking at within-country (fixed) effects, the number of countries included in the sample is irrelevant. The positive *OECD* coefficient may therefore be said to pertain to studies using between-country specifications and thus finds a larger association between income and democracy. Researchers may also be aware that the use of annual data (as opposed to multi-year averages) does not produce any systematic differences in study-outcomes; neither does looking at democracy in the long run. Recall the argument of Dahl (1971) presented above concerning democracy being a concept that might have long-term

determinants: My analysis does not detect any systematic difference in the findings generated by studies that try to test this argument; I fail to detect any long-term effects.

### 5.1.5 Confounder variables

If we consider some aspects of the modernisation hypothesis included in this analysis, like education or population, we see that education provides no robust effect across the multiple meta-regression analysis. This implies that income, working through education, does not have explanatory power, in line with the findings of Acemoglu et al. (2005), contrary to the findings of Glaeser et al. (2007). Many studies also use population size as a control, following Acemoglu et al. (2008); Heid et al. (2012); Högström (2013); and Werger (2009) who argues that the population density matters for democracy in the long run. This is also one of the variables commonly associated with modernisation theory. These claims are not supported by the multiple meta-regression analysis, and show no robust significance across all models. In Model 5 it is slightly significant, but it does not hold when study-dependency is taken into account by the cluster-robust and multilevel approaches. As noted earlier, *Oil* has a significant negative effect confirming the findings of the rentier-state theorist who arrived at the same conclusion through empirical tests in a comparative perspective (Aslaksen 2010; Ross 2001; Ulfelder 2007). The results indicate that oil hinders democracy by affecting income per capita, thereby disconfirming the findings of Ross (2009). Based on this finding, scholars studying the relationship between income and democracy must pay attention to how oil affects democracy and take account of it in empirical analyses.

## 5.2 Concluding remarks and suggestions for further research

This meta-analysis contributes to the body of research on income and democracy by offering information that can be used to shape future econometric models: It identifies several study characteristics, suggesting that these variables systematically influence the results reported in the studies included in this analysis. Some of these variables included in the meta-regression analysis are important in order to guide future research on the income-democracy nexus. When conducting the actual meta-regression analysis I have followed the guidelines of the MAER network (Stanley et al. 2015) and Stanley & Doucouliagos (2012), while drawing on several other individual meta-analyses from economics, social science and medicine alike. Nevertheless, these methods might change over time when more research on the underlying assumptions of the meta-regression analysis is conducted. An illustrative example of this is

the change from relying on fixed effects to WLS modelling. Meta-analysis is a field of research that is in constant evolution, and as both the econometric techniques and theory of meta-regression evolve, the analysis of these data must be reassessed to ensure that the overall conclusion remains valid.

There are certainly systematic differences between the studies included in this meta-analysis and it may very well be the case that the methodological shortcomings of some studies are responsible for some of the discrepancies observed between them. At the same time, as discussed earlier, there is no established "best practice" when it comes to research on the income-democracy nexus. Nonetheless, this meta-analysis has shown that the choice of dependent variable and data set from which the measure of income is gathered, as well as the estimation technique applied, the sample composition, and the choice of control variables all play important roles in accounting for cross-study heterogeneity.

Further, a reassessment of the systematic differences between Penn World Tables, World Bank Data and Maddison data should be conducted to determine which data set is most preferable with regards to what the scholar in question wants to investigate, for there are systematic differences between these sets. It could also be interesting to develop a theoretical framework for author affiliation and study whether authors with different educations or backgrounds produce different results, and not only the differences between political scientists and economists.[53] More sophisticated tests for long-run determinants could also be employed: The multiple meta-regression analysis could, for instance, test the effect of using colonisation control in a long-run perspective. Even though this analysis confirms the findings of Przeworski et al. (2000), and not for instance Boix & Stokes (2003) with democracy as a continuous variable, it is of paramount importance to investigate the literature that uses a dichotomous conceptualisation of democracy in a meta-analytical framework in order to identify whether higher income prevents authoritarian reversals.

This thesis has also contributed methodology to the scholarly field by applying meta-analysis to a topic that has not been subject to such an analysis before. I hope that the original sample and the unique methodological approach can contribute towards the scholarly community finally reaching a consensus concerning this question. Theoretically, it contributes to the literature by showing that modernisation theory has no explanatory power in the context of democratisation with democracy as a graded concept. Moreover, the findings show that some aspects of this relationship need additional research. Theory-generating qualitative case

[53] For example, exploring whether authors trained in the US produce systematically different results than researchers from Europe.

studies could be utilised to investigate casual processes of the within-country variation and perhaps produce other theories that can be used as a falsifying hypothesis for the modernisation theory. In terms of quantitative methodology, a "best practice" in terms of methodology must be determined. Appropriate estimation techniques should be applied when investigating within-country effects. To guarantee this, additional explorations and simulations must be conducted in order to ensure the robustness of the statistical approaches that treat democracy as an endogenous concept. Proper specification of the GMM and/or IV approaches must thus be used.

Additionally, a clustering of the reported estimates from individual studies on three levels could also be investigated using statistical modelling. This means that it should be investigated whether it is possible to look at the collected estimates as clustered in studies, and studies being clustered within authors. This will investigate the hypothesis that individual attributes of the authors matter (for example, that it is expected that the same authors show consistency in their publications). I would argue that this has to be investigated using a mixed effects multilevel model.[54] This, however, might be problematic due to the assumptions of mixed/random effects models in meta-analyses, as discussed above. But it might be interesting to think of published studies in these terms and develop statistical techniques to accommodate these problems.

In terms of future meta-analysis in comparative politics, further research could also investigate the income-democracy nexus in a comparative perspective, looking at further economic, social and other determinants of democratisation to investigate whether income has a cofounder effect on democracy in a meta-analytical framework. Also, additional confounder variables, commonly included in the literature, such as religion, inequality, etc., should be addressed in a multiple meta-regression analysis after updating and expanding the meta-data set. As new publications on the relationship between income and democracy emerge, the original data set used in this thesis would have to be updated in order to remain valid and reliable., but per now the conclusions reached in this thesis remains valid and reliable.

[54] See Steenbergen & Jones (2002); Luke (2004); Rabe-Hesketh & Skrondal (2012); and Snijders & Bosker (2012) for discussion.

# 6. Bibliography


Abdullah, A., Doucouliagos, H. & Manning, E. (2013): "Does education reduce income inequality? A meta-regression analysis". *Journal of Economic Surveys*.

Acemoglu, D., Johnson, S., Robinson, J. A. & Yared, P. (2005): "From education to democracy?". *The American Economic Review,* 95 (2).

Acemoglu, D., Johnson, S., Robinson, J. A. & Yared, P. (2008): "Income and democracy". *American Economic Review,* 98 (3): 808-42.

Acemoglu, D., Johnson, S., Robinson, J. A. & Yared, P. (2009): "Reevaluating the modernization hypothesis". *Journal of Monetary Economics,* 56 (8): 1043-1058.

Adam, A., Kammas, P. & Lagou, A. (2013): "The effect of globalization on capital taxation: What have we learned after 20 years of empirical studies?". *Journal of Macroeconomics,* 35 (0): 199-209.

Adcock, R. & Collier, D. (2001): "Measurement validity: A shared standard for qualitative and quantitative research". *American Political Science Review,* 95 (3): 529-46.

Adelman, I. & Morris, C. (1965): "A factor analysis of the interrelationship between social and political variables and per capita gross national product". *Quarterly Journal of Economics* (79): 555-78.

Ahmadov, A. K. (2014): "Oil, democracy, and context: A meta-analysis". *Comparative Political Studies,* 47 (9): 1238-1267.

Aldrich, J. & Cnudde, C. F. (1975): "Probing the bounds of conventional wisdom: A comparison of regression, probit, and discriminant analysis". *American Journal of Political Science,* 19 (3): 571-608.

Alemán, J. & Yang, D., Y. (2011): "A duration analysis of democratic transitions and authoritarian backslides". *Comparative Political Studies,* 44 (9).

Andersen, R. (2012): "Support for democracy in cross-national perspective: The detrimental effect of economic inequality". *Research in Social Stratification and Mobility,* 30 (4): 389-402.

Ansell, B. & Samuels, D. (2010): "Inequality and democratization: A contractarian approach". *Comparative Political Studies,* 43 (12).

Arellano, M. & Bond, S. (1991): "Some tests of specification for panel data - Monte-Carlo evidence and an application to employment equations". *Rev. Econ. Stud.,* 58 (2): 277-297.

Arellano, M. & Bover, O. (1995): "Another look at the instrumental variable estimation of error-components models". *Journal of econometrics,* 68 (1): 29-51.

Aslaksen, S. (2010): "Oil and democracy: More than a cross-country correlation?". *Journal of Peace Research,* 47 (4).

Barro, R. (1999): "Determinants of democracy". *Journal of political economy,* 107 (S6).

Beck, N. & Katz, J. N. (2001): "Time-series cross-section data: What have we learned in the past few years?". *Annu. Rev. Polit. Sci,* 4: 271-93.

Begg, C. B. & Berlin, J. A. (1988): "Publication bias – a problem in interpreting medical data". *Journal of the Royal Statistical Society Series a-Statistics in Society,* 151 (419-63).

Bell, A. & Jones, K. (2015): "Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data". *Political Science Research and Methods,* 3 (01): 133-153.

Benhabib, J., Corvalan, A. & Spiegel, M. (2011): Reestablishing the income-democracy nexus. *NBER Working Paper*

Benyishay, A. & Betancourt, R. R. (2010): "Civil liberties and economic development". *Journal of Institutional Economics,* 6 (03): 281-304.
Bobba, M. & Coviello, D. (2007): "Weak instruments and weak identification in estimating the effects of education on democracy". *Economics Letters,* 96 (3): 301–306.
Boix, C. (2009): Development and democratisation. *IBEI Working Papers.*
Boix, C. (2011): "Democracy, development, and the international system". *The American Political Science Review,* 105 (4): 809-828.
Boix, C., Miller, M. K. & Rosato, S. (2013): "A complete data set of political regimes, 1800-2007". *Comparative Political Studies* 46 (12): 1523-54.
Boix, C. & Stokes, S. (2003): "Endogenous democratisation". *World Politics,* 55 (4): 517-549.
Bollen, K. A. (1990): "Political democracy:Conceptual and measurement traps". *Studies in Comparative International Development,* 25 (1): 7-24.
Bollen, K. A. & Jackman, R. W. (1995): "Income inequality and democratization revisited: Comment on Müller". *American Sociological Review,* 60 (6): 983-989.
Borenstein, M., et al. (2009): *Introduction to meta-analysis.* Chichester: Wiley.
Brinks, D. & Coppedge, M. (2006): "Diffusion is no illusion: Neighbor emulation in the third wave of democracy". *Comparative Political Studies,* 39 (4): 463-489.
Brodeur, A., Lé, M., Sangnier, M. & Zylberberg, Y. (2012): Star wars: The empirics strike back. *Paris School of Economics Working Paper* Paris School of Economics: Paris School of Economics.
Burkhart, R., E. (1997): "Comparative democracy and income distribution: Shape and direction of the causal arrow". *Journal of Politics,* (59): 148-64.
Burkhart, R., E. & Lewis-Beck, M. (1994): "Comparative democracy: The economic development thesis". *American Political Science Review* 4(88): 903-10.
Callot, L. & Paldam, M. (2010): Natural funnel asymmetries: A simulation analysis of the three basic tools of meta analysis *Economics Working Paper.* School of Economics and Management: Aarhus University
Capoccia, G., R., & Kelemen, D. (2007): "The study of critical junctures: Theory, narrative, and counterfactuals in historical institutionalism". *World Politics,* 59 (3): 341-369.
Card, N., A. (2012): *Applied meta-analysis for social science research.* New York: Guilford Press.
Che, Y., Lu, Y., Tao, Z. & Wang, P. (2013): "The impact of income on democracy revisited". *Journal of Comparative Economics,* 41 (1): 159-169.
Clark, T. S. & Linzer, D. A. (2014): "Should I use fixed or random effects?". *Political Science Research and Methods,* FirstView: 1-10.
Cohen, J. (1992): "Statistical power analysis". *Current Directions in Psychological Science,* 1 (3): 98-101.
Colaresi, M. & Thompson, W. R. (2003): "The economic development-democratization relationship: Does the outside world matter?". *Comparative Political Studies,* 36 (4): 381-403.
Coleman, J., S. & Almond, G. (1960): *Conclusion: The political systems of the developing areas.* Princeton, NJ:: Princeton University Press.
Collier, R. & Collier, D. (1991): *Shaping the political arena.* Princeton: Princeton University Press.
Cooper, H. M. (2010): *Research synthesis and meta-analysis: A step-by-step approach.* Los Angeles: Sage.
Coppedge, M. (2003): "Book review: Przeworski et al. 2000,". *Studies in Comparative International Developmen,* 38 (1): 123-126.

Coppedge, M. (2012): *Research methods and democratization.* New York: Cambridge Univ. press.
Coppedge, M. & Midlarsky, M. I. (1997): "Modernization and thresholds of democracy: Evidence for a common path and process". *Inequality, democracy, and economic development*: 177-201.
Cornell, A. (2012): "Does regime type matter for the impact of democracy aid on democracy?". *Democratization,,* 20 (4): 642-67.
Costa-Font, J., De-Albuquerque, F. & Doucouliagos, H. (2014): "Do jurisdictions compete on taxes? A meta-regression analysis". *Public Choice,* 161 (3-4): 451-470.
Cutright, P. (1963): "National political development: Measurement and analysis". *American Sociological Review,* 28 (2): 253-264.
Dahl, R. A. (1971): *Polyarchy: Participation and opposition.* New Haven: Yale University Press.
Dahl, R. A. (1989): *Democracy and its critics.* New Haven, CT: Yale Univ. Press.
Diamond, L. (1992): "Edonomic development and democracy reconsidered". *American Behavioral Scientist* 35 (4): 450-499.
Doucouliagos, H. & Laroche, P. (2003): "What do unions do to productivity? A meta-analysis". *Industrial Relations: A Journal of Economy and Society,* 42 (4): 650-691.
Doucouliagos, H., Laroche, P. & Stanley, T. D. (2005): "Publication bias in union-productivity research?". *Relations Industrielles-Industrial Relations,* (60): 320-47.
Doucouliagos, H. & Paldam, M. (2008): "Aid effectiveness on growth: A meta study". *European Journal of Political Economy,* 24 (1): 1-24.
Doucouliagos, H. & Paldam, M. (2013): "The robust result in meta-analysis of aid effectiveness: A response to mekasha and tarp". *The Journal of Development Studies,* 49 (4): 584-587.
Doucouliagos, H. & Stanley, T. D. (2008): Publication selection bias in minimum-wage research? A meta-regression analysis IUniversity, D. (red.) *Faculty of Business and law, School working paper.* Deakin University.
Doucouliagos, H. & Stanley, T. D. (2013): "Are all economic facts greatly exaggerated? Theory competition and selectivity". *Journal of Economic Surveys,* 27 (2): 316-339.
Doucouliagos, H. & Ulbasoglu, M. A. (2008): "Democracy and economic growth: A meta-analysis". *American Journal of Political Science,* 52 (1): 61-83.
Efendic, A., Pugh, G. & Adnett, N. (2011): "Institutions and economic performance: A meta-regression analysis". *European Journal of Political Economy,* 27 (3): 586-599.
Egger, M. S., Schneider, M. & Minder, C. (1997): "Bias in the meta-analysis detected by a simple, graphical test". *British Medical Journal,* (315): 629-34.
Epstein, D. L., Bates, R., Goldstone, J., Kristensen, I. & O'Halloran, S. (2006): "Democratic transitions". *American Journal of Political Science,* 50 (3): 551-569.
Faria, H. J., Montesinos-Yufa, H., M. & Morales, D. R. (2014): "Should the modernization hypothesis survive acemoglu, johnson, robinson, and yared? Some more evidence". *Econ Jorunal Watch,* 11 (1): 17-36.
Feld, L. P. & Heckemeyer, J. H. (2011): "Fdi and taxation: A meta-study". *Journal of Economic Surveys,* 25 (2): 233-272.
Gerber, A. & Malhotra, N. (2008): "Do statistical reporting standards affect what is published? Publication bias in two leading political science journals". *Quarterly Journal of Political Science,* (3): 313-26.
Giles, M. W. & Garand, J. C. (2007): "Ranking political science journals: Reputational and citational approaches". *PS: Political Science & Politics,* (40): 741-51.
Glaeser, E. L., Giacomo, A. M. P. & Shleifer, A. (2007): "Why does democracy need education?". *Journal of economic growth,* 12.

Gleditsch, K. S. & Ward, M. D. (2006): "Diffusion and the international context of democratization". *International Organization,* 60.
Goertz, G. (2006): *Social science concepts: A user's guide.* Princeton, N.J.: Princeton University Press.
Goldberg, E., Wibbels, E. & Mvukiyehe, E. (2008): "Lessons from strange cases: Democracy, development, and the resource curse in the u.S. States". *Comparative Political Studies,* 41 (4-5): 477-514.
Gonick, L., S., & Robert M., R. (1988): "The structural constraints of the world-economy on national political development". *Comparative Political Studies,* (21): 171-99.
Greenberg, D., Michalopoulos, C. & Robins, P. (2003): "A meta-analysis of government - sponsored training programs". *Industrial Labor Relations Review* 57 (1).
Gujarati, D. N. & Porter, D. C. (2010): *Essentials of econometrics.* Boston: McGraw-Hill.
Gundlach, E. & Paldam, M. (2012): "The democratic transition: Short-run and long-run causality between income and the Gastil Index". *European Journal of Development Research,* 24: 144–168.
Hadenius, A. (1992): *Democracy and development.* Cambridge: Cambridge University Press.
Hadenius, A. & Teorell, J. (2005a): Assessing alternative indicies of democracy. *Concepts & methods Working Papers.* IPSA.
Hadenius, A. & Teorell, J. (2005b): "Cultural and economic prerequisites of democracy: Reassessing recent evidence". *Studies in Comparative International Development,* 39 (4): 87-106.
Hariri, J. G. (2015): "A contribution to the understanding of middle eastern and muslim exceptionalism". *Journal of Politics*.
Heid, B., Langer, J. & Larch, M. (2012): "Income and democracy: Evidence from system GMM estimates". *Economics Letters,* 116.
Herb, M. (2004): "No representation without taxation? Rents, development, and democracy". *Comparative Politics January,* 37 (297-319).
Houle, C. (2009): "Inequality and democracy why inequality harms consolidation but does not affect democratization". *World Politics,* 61 (4).
Hunter, J. E. & Schmidt, F. L. (2004): *Methods of meta-analysis: Correcting error and bias in research findings.* Thousand Oaks, Calif.: Sage.
Huntington, S. P. (1968): *Political order in changing societies.* New Haven, Conn.: Yale University Press.
Huntington, S. P. (1991): *The third wave: Democratization in the late twentieth century.* Norman: University of Oklahoma Press.
Högström, J. (2013): "Does the choice of democracy measure matter? Comparisons between the two leading democracy indices, Freedom House and Polity IV". *Government and Opposition,,* 48.
Jackman, R. W. (1973): "On the relation of economic development and democratic performance". *American Journal of Political Science,* 3 (17): 611-21.
Jaggers, K. & Gurr, T. R. (1995): "Tracking democracy's third wave with the Polity III data". *Journal of Peace Research,* 32 (4): 469-82.
Kallager, P. K. R. (2014): *Globalisation and the welfare state – a meta-regression analysis.* Univ. of Bergen.
Kelsen, H. (1945): *General theory of law and state.* Cambridge: Harvard University Press.
Knutsen, C., H. & Wig, T. (2015): "Government turnover and the effects of regime type: How requiring alternation in power biases against the estimated economic benefits of democracy". *Comparative Political Studies,* 33 (1).
Knutsen, C. H. (2010): "Measuring effective democracy". *International Political Science Review,* 31 (2): 109-128.

Kontopantelis, E. & Reeves, D. (2010): "Metaan: Random-effects meta-analysis". *Stata Journal,* 10 (3): 395-407.
Krieckhaus, J. (2004): "The regime debate revisted: A sensitivity analysis of democracy's economic effect". *British Journal of Political Science,* 34 (4): 635-55.
Levine, R. & Renelt, D. (1991): Cross-country studies of growth and policy methodological, conceptual, and statistical problem. *Macroeconomic Adjustment and Growth.* World Bank: World Bank.
Lipset, S. M. (1959): "Some social requisites of democracy: Economic development and political legitimacy". *The American Political Science Review,* 53 (1): 69-105.
Lipset, S. M. (1960): *Political man: The social bases of politics.* New York: Doubleday.
Ludvigsen, S. S. (2010): *Post-mortem of the VP-function?: Meta-regression analyses of economic voting in the United Kingdom; an anthology.* Århus: Politica.
Luke, D. A. (2004): *Multilevel modeling.* Thousand Oaks, Calif: Sage Publications.
Maddison, A. (2003): *The world economy: Historical statistics.* OECD Publishing.
Mahoney, J. (2001): "Path dependent explanations of regime change". *Studies in Comparative and International Development,* 36 (1): 114.
Mainwaring, S. & Pérez-Liñán, A. (2003): "Level of development and democracy Latin American exceptionalism, 1945-1996". *Comparative Political Studies,* 36 (9): 1031-1067.
Midtbø, T. (2012): *Stata: En entusiastisk innføring.* Oslo: Universitetsforlaget
Montesquieu, C., Derathé, R. & Casabianca, D. (2011): *L'esprit des lois.* Paris: Classiques Garnier.
Moore, B. (1966): *Social origins of dictatorship and democracy.* Beacon Press.
Moral-Benito, E. & Bartolucci, C. (2012): "Income and democracy: Revisiting the evidence". *Economics Letters,* 117.
Munck, G. & Verkuilen, J. (2002): "Conceptualizing and measuring democracy: Evaluating alternative indices". *Comparative Political Studies,* Vol. 35 (No. 1).
Murtin, F. & Wacziarg, R. (2014): "The democratic transition". *Journal of Economic Growth,* 19 (2): 141-181.
Nelson, J. & Kennedy, P. (2009): "The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment". *Environmental and Resource Economics,* 42 (3): 345-377.
Nickell, S. (1981): "Biases in dynamic models with fixed effects',". *Econometrica,* 49: 1417-1426.
O'Donnell, G. (1963): *Modernization and bureaucratic-authoritarianism.* Berkely: University of California.
O'Donnell, G. (2001): "Democratic theory and comparative politics". *Studies in Comparative and International Development,* 36 (1).
Oskarsson, S. & Ottosen, E. (2010): "Does oil still hinder democracy?". *The Journal of Development Studies* 46 (6): 1067-1083.
Persson, T. & Tabellini, G. (2009): "Democratic capital: The nexus of political and economic change". *American Economic Journal,* 1 (2): 88-126.
Pierson, P. (2004): *Politics in time.* Princeton: Princeton University Press.
Popper, K. R. (1959): *The logic of scientific discovery.* London: Hutchinson.
Przeworski, A. (2010): *Democracy and the limits of self-government.* Cambridge: Cambridge University Press.
Przeworski, A., Alvarez, M., Cheibub, A. & Limongi, F. (2000): *Democracy and development: Political institutions and well-being in the world, 1950-1990.* Cambridge University Press.

Przeworski, A. & Limongi, F. (1997): "Modernization: Theories and facts ". *World Politics,* 49 (2): 155-183.
Rabe-Hesketh, S. & Skrondal, A. (2012): *Multilevel and longitudinal modeling using stata.* College Station, Tex.: Stata Press.
Robinson, J. A. (2006): "Economic development and democracy". *Annu. Rev. Polit. Sci,* 9: 503-27.
Rodrik, D. (2005): Why we learn nothing from regressing economic growth on policies. Cambridge, MA: Harvard University.
Roodman, D. (2009): "Practitioners' corner: A note on the theme of too many instruments". *Oxford Bulletin of Economics and Statistics,* 71 (1).
Ross, M. (2001): "Does oil hinder democracy?". *World Politics,* 53.
Ross, M. (2009): "Oil and democracy revisited". *University of California–Los Angeles. Mimeo*.
Rudra, N. (2005): "Globalization and the strengthening of democracy in the developing world". *American Journal of Political Science,* 49 (4): 704-730.
Rueschemeyer, D., Stephens, E., H., & Stephens, J., D., (1992): *Capitalist development and democracy.* Chicago: University of Chicago Press.
Schumpeter, J. A. (2006 [1942]): *Capitalism, socialism and democracy.* London: Routledge.
Seawright, J. (2007): "Democracy and growth: A case study in failed causal inference" s. i Munck, G. (red.), *Regimes and democracy in latin america: Theories and metods.* Oxford: Oxford University Press.
Sekhon, J. (2009): "Opiates for the matches: Matching methods for causal inference". *Annual Review of Political Science* 12: 487-508.
Shafiq, M. N. (2010): "Do education and income affect support for democracy in muslim countries? Evidence from the pew global attitudes project". *Economics of Education Review,* 29 (3): 461-469.
Skog, O., J. (2007): *Å forklare sosiale fenomener: En regresjonsbasert tilnærming.* Oslo: Gyldendal akademisk.
Smets, K. & van Ham, C. (2013): "The embarrassment of riches? A meta-analysis of individual-level research on voter turnout". *Electoral Studies,* 32 (2): 344-359.
Snijders, T., A., B., & Bosker, R. J. (2012): *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.
Stanley, T. D. (2001): "Wheat from chaff: Meta-analysis as quantitative literature review". *Journal of Economic Perspectives,* 15 (3): 131-150.
Stanley, T. D. (2004): "Does unemployment hysteresis falsify the natural rate hypothesis? A meta-regression analysis". *Journal of Economic Surveys,* 18 (4): 589-612.
Stanley, T. D. (2005): "Beyond publication bias". *Journal of Economic Surveys,* 19 (3): 309-345.
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.
Stanley, T. D. & Doucouliagos, H. (2012): *Meta-regression analysis in economics and business.* Florence, KY, USA: Routledge.
Stanley, T. D. & Doucouliagos, H. (2013): "Are all economic facts greatly exaggerated? Theory competition and selectivity". *Journal of Economic Surveys,* 27 (316-39).
Stanley, T. D. & Doucouliagos, H. (2013a): Better than random: Weighted least squares meta-regression analysis. *Economics Series.* Faculty of Business and Law, School of Accounting, Economics and Finance: Deakin University.

Stanley, T. D. & Doucouliagos, H. (2013b): "Neither fixed nor random: Weighted least squares meta-analysis". *Economic series Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.,* 1.

Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., Nelson, J. P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R. S. & Rost, K. (2013): "Meta-analysis of economis: Research reporting guidelines". *Journal of Economic Surveys,* 27 (2): 390-394.

Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., Nelson, J. P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R. S. & Rost, K. (2015): *Meta-analysis of economics research network*. https://www.hendrix.edu/maer-network/. [02.02. 2015].

Stanley, T. D. & Jarrell, S. B. (1998): "Gender wage discrimination bias? A meta-regression analysis". *Journal of Human Resources,* 33 (4): 947-973.

Stanley, T. D., Jarrell, S. B. (1989): "Meta-regression analysis: A quantitiative method of literature surveys". *Journal of Economic Surveys,* 19 (3).

Steenbergen, M. R. & Jones, B. S. (2002): "Modeling multilevel data structures". *American Journal of Political Science,* 46 (1): 218-237.

Summers, L., H. (1991): "He scientific illusion in empirical macroeconomics". *Scandinavian Journal of Economics,* 93 (2): 129-48.

Teorell, J. (2010): Determinants of democratization: Explaining regime change in the world, 1972-2006. Cambridge: Cambridge University Press.

Treisman, D. (2011): Income, democracy and the cunning of reason. *NBER Working Paper*

Ulfelder, J. (2007): "Natural-resource wealth and the survival of autocracy". *Comparative Political Studies*.

Vanhanen, T. (2000): "A new dataset for measuring democracy, 1810-1998". *Journal of Peace Research,* 37 (2): 251-265.

Werger, C. (2009): Effect of oil and diamonds on democracy: Is there really a resource curse? *OxCarre Research Pape.* Oxford: Oxford.

Wilson, D. B. (2001): "Meta-analytic methods for criminology". *The ANNALS of the American Academy of Political and Social Science,* 578.

# 7. Appendix

## 7.1 Income and Democracy

### Figure 7.1: Income and democracy, 1990s

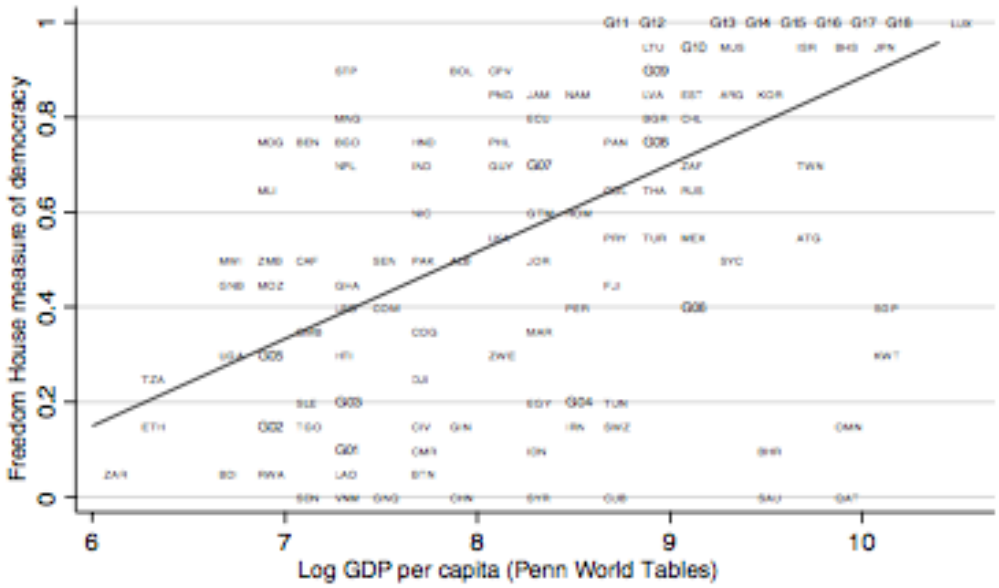



Figure 7.1 replicated from (Acemoglu et al. 2008: 809)

### Figure 7.2: Impact of income on contestation by year: Cross sectional estimates with 95% confidence intervals

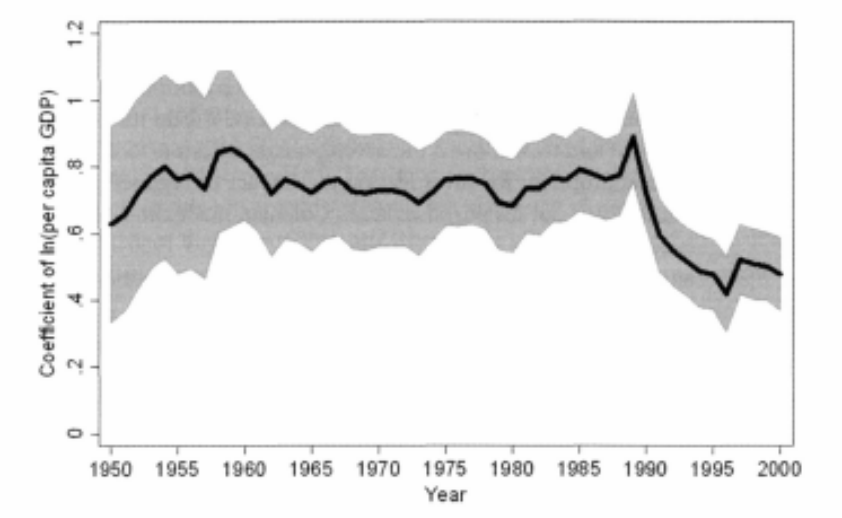



Figure 7.2 replicated from (Coppedge 2012: 272)

## 7.2 Assumption tests

**Figure 7.3: Histogram of dependent variable**

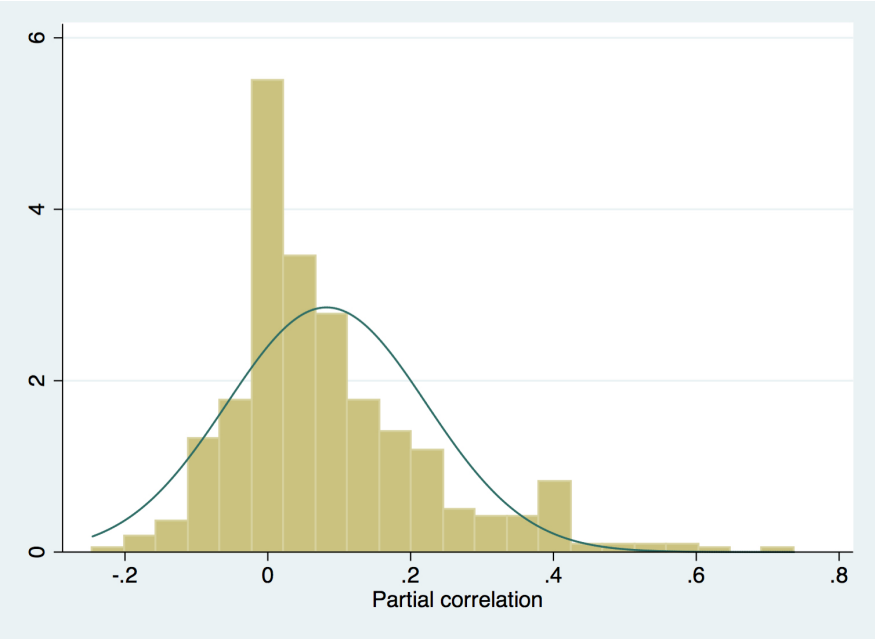



**Figure 7.4 Q-norm plot of partial correlation over inverse normal**

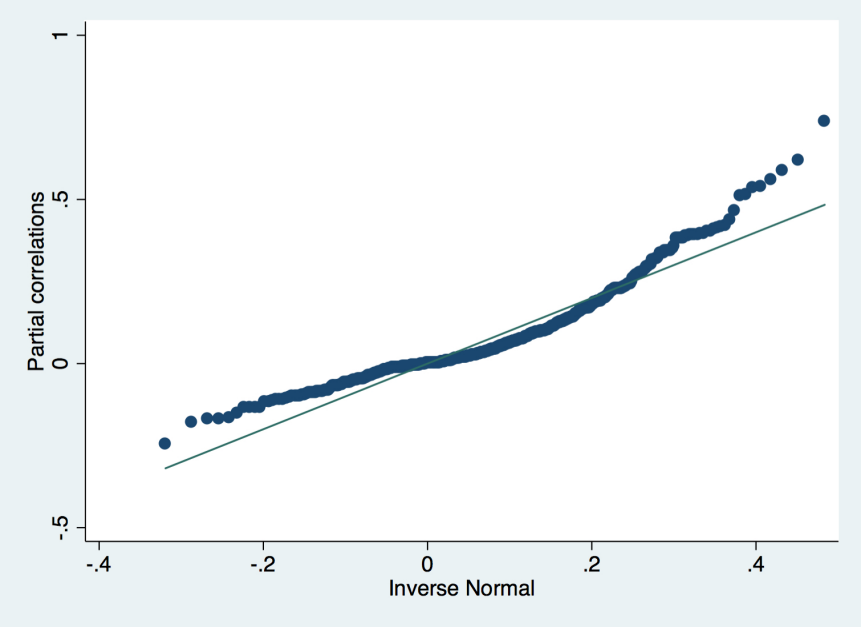


**Figure 7.5 Scatter plot of partial correlations linear regression fitted**

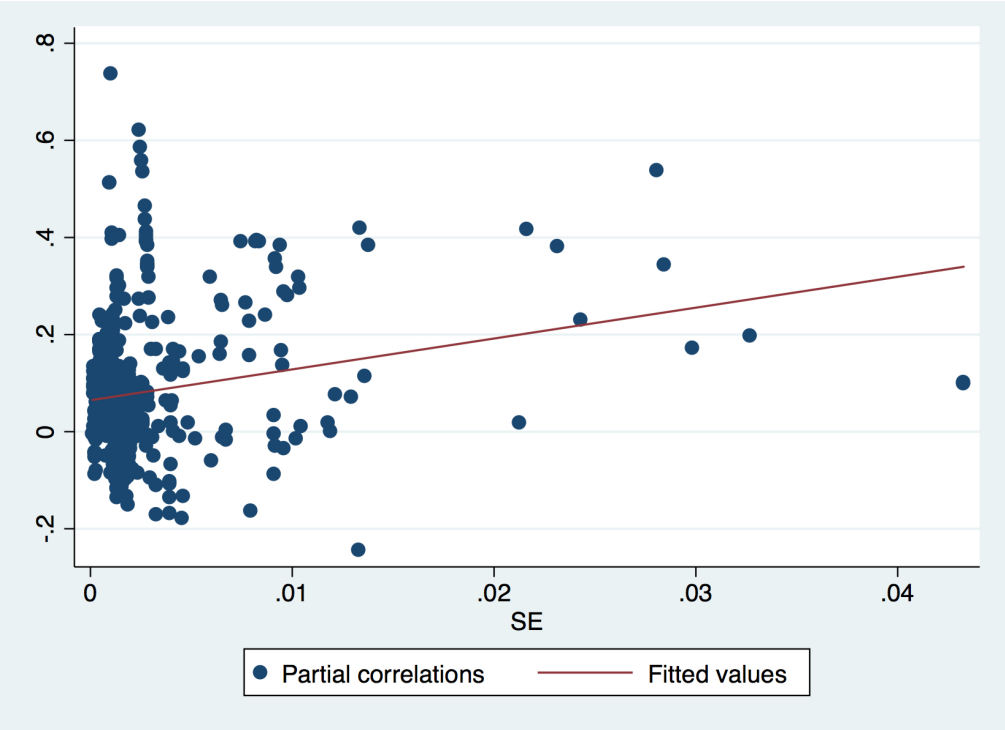



**Table 7.1 Test for multicollinearity from Model 1 - WLS R**

| **Variable** | **VIF** | **1/VIF** |
|---|---|---|
| Unpublished | 24.27 | 0.041198 |
| Penn | 22.23 | 0.044994 |
| Maddison | 19.77 | 0.050578 |
| Education | 13.79 | 0.072541 |
| Populations | 13.01 | 0.076854 |
| Endogeneity | 10.29 | 0.097213 |
| Income focus | 7.56 | 0.132346 |
| Five year | 6.63 | 0.150932 |
| Countries | 6.31 | 0.158525 |
| Pre 1960 | 6.06 | 0.164937 |
| Western | 5.98 | 0.167108 |
| FHI | 5.93 | 0.168607 |
| OLS FE | 5.64 | 0.177453 |
| Publication year | 4.98 | 0.200614 |
| SE | 4.62 | 0.216244 |
| Political scientist | 4.14 | 0.241588 |
| Annual | 3.68 | 0.271814 |
| No socialist | 2.68 | 0.373733 |
| Prior | 2.62 | 0.381083 |
| Oil | 1.76 | 0.568574 |
| **Mean VIF** | **8.60** | |

### 7.3 Selection and estimation procedures

**Figure 7.6 Flow-chart for construction of meta-data set**

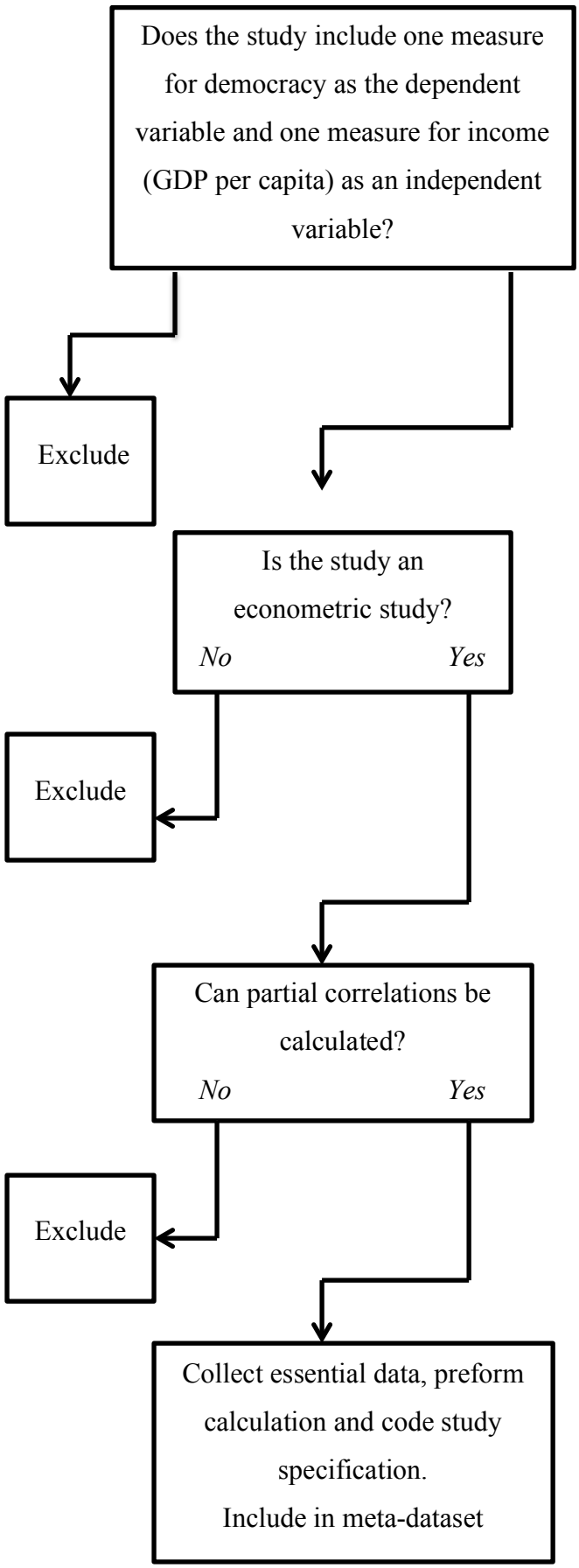


**Figure 7.7 Schema for investigating research heterogeneity**

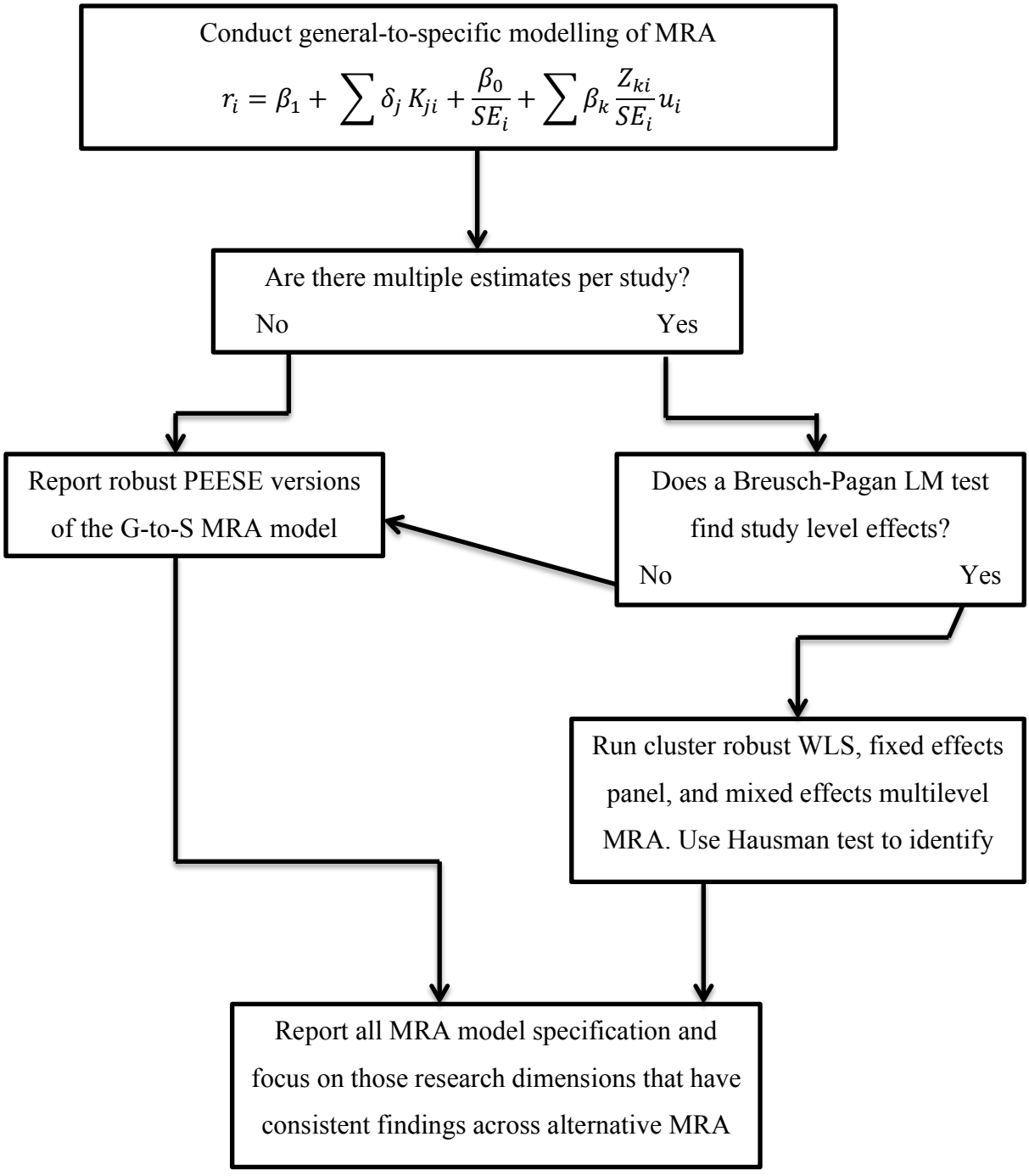



Figure 7.7 is based on Stanley & Doucouliagos (2012: 115).