Dipartimento di Economia Internazionale, delle Istituzioni e dello Sviluppo (DISEIS)

WORKING PAPER SERIES

WP 1703

June, 2017

Does democracy cause growth? A meta-analysis perspective

Marco Colagrossi, Domenico Rossignoli,

Mario A. Maggioni

Prior to being published, DISEIS working papers are subject to an anonymous refereeing process by members of the editorial board or by members of the department.

DISEIS Working Papers often represent preliminary work which are circulated for discussion and comment purposes. Citation and use of such a paper should account for its provisional character. A revised version may be available directly from the author.

DISEIS WP 1703

DISEIS
Dipartimento di Economia Internazionale, delle Istituzioni e dello Sviluppo
Università Cattolica del Sacro Cuore
Via Necchi 5
20123
Milano

Does democracy cause growth?
A meta-analysis perspective

Marco Colagrossi, Domenico Rossignoli, and
Mario A. Maggioni

CSCC and DISEIS Universitá Cattolica del Sacro Cuore di Milano

June, 2017

Abstract

The relationship between democracy and economic growth has long been investigated both in the political science and in the economic literature with inconclusive outcomes. By adopting a multi-level meta-analysis framework, we tried to shed lights on this conundrum. Our hierarchical sample includes 103 studies containing 942 point-estimates. Our random effects model suggests that the sign of this relationship, albeit positive, is statistically weak. We then address the high between-studies heterogeneity by adopting meta-regression analysis models. Results are striking: the effect sizes' variance is largely driven by spatial and temporal differences in the samples, indicating that the democracy and growth nexus is not homogeneous across world regions and time periods. Conversely, the large number of control variables included in the papers, do not impact the reported results. At the same time, models estimated by means of the *within estimator* have a significant, albeit negative, impact on economic growth. This seems to suggest that scholars have not yet found the appropriate control variables - or their suitable proxies - to explain such widely debated relationship.

Corresponding author: domenico.rossignoli@unicatt.it
Keywords: Democracy and Growth; Meta-Analysis; Meta-Regression; Multilevel Models

Acknowledgements: The authors are grateful to the participants to the 7th Annual General Conference of the EPSA (European Political Science Association) for insightful comments and feedbacks on a more recent version of the paper. The authors also thank S. Beretta for comments on the preliminary versions of the work. Of course, any errors are our own.

1 Introduction

Growth economists have long being studying, through quantitative methods, the determinants of cross-country differences in GDP levels and rates of change. While the first attempts of empirical investigations date back to the mid-1980s, it is only in the early 1990s (see among others Barro, 1991; Barro and Sala-i Martin, 1992) that regression methods have been extensively applied to the subject. In their simplest form, growth regressions involve a measure of economic growth as dependent variable and its alleged determinant(s) as independent variable(s), so that:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \tag{1}$$

Where y is the growth rate of the economy, $x_1 \cdots x_n$ its explanatory variables and ϵ the error term. Unfortunately, as Sala-i Martin (1997) notes, economic growth theories are usually not explicit in stating which are the variables that matter most for a country to prosper. This, in turn, brought an accumulation of evidences that was, at best, disorderly.

This has been (and still is) particularly true for social scientists who focused on the institutional determinants of economic growth. Among them, several engaged in a debate - not always independent of ideological burdens - regarding the impact of democratic institutions on the living standards of the population. Despite several attempts, consensus has been far from being reached. As Sirowy and Inkeles noted "many of the central questions pertaining to the developmental consequences of political democracy remain, by and large, unresolved"; furthermore, "the relevant quantitative, cross-national research continues to be plagued by conflicting findings" (Sirowy and Inkeles, 1990, p. 127). Few years later, similar conclusions have been reached by Przeworski and Limongi: when it comes to the nexus between democratic institutions and economic development, "social scientists know surprisingly little" (Przeworski and Limongi, 1993, p. 51).

In fact, despite the democracy and growth *conundrum* benefited from renewed attention after the influential contribution by Barro (1996) and the introduction of more sophisticated econometric techniques (such as System GMM), the key question remained largely undressed (Knutsen, 2012).

A single paper is unlikely to solve this puzzle. Within a classical hypothesis testing framework, a null hypothesis is posed against an alternative; whether the null hypothesis is rejected or accepted, depends upon a single test statistics exceeding or not an arbitrary critical value. Therefore, as the *American Statistical Association* explains, despite "researchers often wish to turn a p-value into a statement about the truth of a null hypothesis", the p-value is just "a statement about data in relation to a specified hypothetical explanation" (Wasserstein and Lazar, 2016, p. 131). Scientific conclusions should then not be based only on whether a single test statistics passes a specific threshold. As De Long and Lang noted, "since no individual test is definitive, it is somewhat surprising that the rhetoric of article writing suggests that a single test or series of tests in the individual article is conclusive" (De Long and Lang, 1992, p. 1258).

A more statistically appropriate methodology to draw a more robust, albeit not definitive, inference about an alleged relation is the one followed by Doucouliagos and

Ulubaşoğlu (2008). By applying a meta-regression analysis to a sample of 483 estimates included in 84 studies, these authors investigated the causes behind the ambivalent results obtained in the literature on the relation between democracy and economic growth.

Using modern meta-analysis techniques and a new, larger and up-to-date sample of studies, we build upon the research by Doucouliagos and Ulubaşoğlu (2008) to better unravel the democracy and growth conundrum.

In their meta-analysis, Doucouliagos and Ulubaşoğlu choose not to include working papers and unpublished literature as a way to control for the quality of the effect sizes included in the sample. Yet, if *publication bias* occurs in this literature, then it might results in a *prejudice against the null hypothesis*, hence an over-representation of statistical significant results within the sample (Rothstein et al., 2006a). We therefore also included 23 unpublished paper in our sample. Moreover, we decided to collect all point estimates included in each study, to avoid introducing a further bias in the selection process. By adopting such an approach however our observations are unlikely to be independent from each other, since effect sizes collected from the same study are likely contingent on each other, an issue sometimes called *hierarchical dependence* (Stevens and Taylor, 2009). To deal with it, we resorted to multilevel models.

The remaining of the papers is organized as follows: section 2 briefly reviews the literature on growth determinants, especially focusing on the role of democracy. Section 3 outlines the preliminary steps that we undertook to perform our meta-analysis, explains how we computed our effect size measures and introduces the multilevel model that we will adopt. Finally, section 5 presents and discusses the results we obtained, before drawing our conclusions, presented in section 6.

2 Modern economic growth theories; a review

The theoretical framework in which the empirical analyses upon the determinants of economic growth and the disparities in growth rates across countries can be traced back to Solow (1956). In describing the dynamics of economic growth, such model assumes that the increase in the aggregate production of goods and services is determined by the growth in the stock of capital (K), labor force (or population, L) and the level of technology (A), which lately became mostly known as the Total Factor Productivity (TFP).

TFP accounts for all the omitted factors - such as institutions, resource endowments, climate, and the literacy and skills of the workforce, and is said exogenous. This characteristic, alongside the *diminishing marginal returns* assumption, leads the model to predict steady-state convergence.

The mathematical versatility of the Solow model resulted in a large success, thereby offering favorable foundation for different extensions[1]. Yet, a mounting discontent with the inability of the original neoclassical paradigm to justify increasingly divergent growth patterns led to the development of alternative growth models. Starting from the

[1] Cass (1965) and Koopmans et al. (1965) integrated the analysis of consumer optimization into the original textbook Solow model; as a result, the savings rate, which were exogenous in the Solow (1956) model, became endogenous.

mid-1980s, a new group of theories tried to offer a *endogenous* explanation of economic growth. This new breed of theories challenged the firsts modern growth models in postulating that physical and labor capital are the only two observable inputs in the production function. Instead, emphasis was given to *human capital* and other forms of intangible capital like knowledge, considering it as the result of intentional investment in education and formation (Romer, 1990; Lucas, 1988)[2].

The idea of *absolute convergence* across different economies became a rather unrealistic assumption. In the augmented growth models, the convergence assumed by Solow (1956) became *conditional* upon other parameters including propensity to save and investment rate, fertility choices and population growth, state technology, trade agreements, education policies, and legal frameworks. Many of the latter can be linked to one encompassing concept: institutions (e.g. Gwartney et al. 1999; (e.g. Acemoglu et al., 2014; Gwartney et al., 2006; De Haan et al., 2006; Acemoglu et al., 2001; Rodrik et al., 2004).

2.1 The role of institutions and political regimes

Despite an initial fierce debate on the role of institutions in determining individual economic behavior and, above all, aggregate performance, starting in the late 1990s economists widely acknowledged that "institutions matter" (Acemoglu et al., 2001, p. 1370).

In simple terms, institutions represent "systems of established and embedded social rules that structure social interactions" (Hodgson, 2006, p. 13). Therefore, institutions can influence the incentive structure by affecting the underlying economic context, potentially curbing or fostering economic activity, as pointed out by North (1990).

Much of the literature, by adopting already existing indices of political regimes, focused on the impact of democracy on economic growth. Such tradition dates back to the seminal comparative study of Lipset (1959) and gained a further momentum in the aftermath of the global economic and financial crisis. Notwithstanding the vast amount of empirical research spurred on the topic, results were mostly contrasting and inconclusive[3]. The first meta-analysis on the issue (Doucouliagos and Ulubaşoğlu, 2008) found that a positive coefficient was found in 27% of the cases, while in 37% it was positive but not significant and in 36% it was either negative or negative and not

[2] Romer (1990) pointed out that private firms have an incentive to invest in research and development to differentiate their production and gain market power. In this context, the productivity of human capital is higher when it is combined with the previous stock of knowledge, through a self-reinforcing process of accumulation (Romer, 1990). This can explain why countries might end up in different steady-state growth paths. Lucas (1988), on the other end, tried to solve the dilemma of divergent national productivity rates by taking into account the rigidity in the international patterns of migration and wage levels. Since the human capital cannot freely flow from a country to another, the differences in initial stocks of human capital are translated in differences into productivity - that is, lower levels of human capital in poor countries determine their lower productivity.

[3] As an example, Przeworski (2000) shows that though the direct relation is quite inconclusive, dictatorship tend to allocate capital more efficiently , while democracies (in high income countries) produce more output per capita. Studies focusing on levels of democracy (rather than regime transitions) identified either negative but not significant (Helliwell, 1994) or negative and slightly significant (Barro, 1996) or inverse-U shaped (Barro et al., 2003) relations, with low or no statistical significance

significant (Doucouliagos and Ulubaşoğlu, 2008, p. 62). In the same study, the authors suggest that other factor, as human capital, possibly act as channels of transmission.

Several scholars focused on such possible channels. One of the arguments in favor of the beneficial effects of democracy on economic growth focuses on democratic institutions as a mean to guarantee property rights (Przeworski and Limongi, 1993). Another argument considers the ability of democratic institutions to assist production and maximize the total output by guaranteeing private activity or stimulating it directly by supplying inputs; supposedly, autocratic rulers have less incentives to do so (Barro, 1991; Przeworski and Limongi, 1993; Olson, 1993). Minier (1998) explored the possibility of a productivity enhancing role of democracy, driven by a more efficient allocation of production factors compared to autocratic regimes. In a different study, Rodrik and Wacziarg found that the process of democratization exerts a positive though heterogeneous effect on growth, where heterogeneity depends on the consolidation of democracy itself (Rodrik and Wacziarg, 2005).

Conversely, a consistent body of scholars perceived the relation between development and political democracy as conflicting or even incompatible (Huntington, 1987). Democracy and economic growth are considered to be competing concerns and trade off in the political sphere are necessary to achieve growth (Sirowy and Inkeles, 1990). Following Przeworski and Limongi (1993), the essential dynamic through which democracy is believed to hamper growth are political pressures for immediate consumption, which reduce the level of investments. Furthermore, whereas there is agreement on the benefits of securing property rights, it is controversial whether dictatorships can better secure these rights (Przeworski and Limongi, 1993). Finally, authoritarian regimes can be more effective in timely implementing the kinds of policies reputed necessary to boost growth (Sirowy and Inkeles, 1990).

The presence of a large number of contrasting scientific results configures democracy and growth as one of the best playing fields to perform meta-analysis. In the next section (3), we introduce this framework and present the model adopted in the paper.

3 Meta-analysis and meta-regression analysis; fixed and random effect(s) models

The term meta-analysis was first introduced by Glass, who defines it as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" of a "rapidly expanding research literature" (Glass, 1976, p. 3)[4]. Meta-analysis is designed to avoid possible biases that might arise in the context of qualitative reviews, such as the choices, made by the reviews' authors, of which papers to include or not include, or how to weight them, based on their own prior beliefs on the topic (Stanley and Jarrell, 1989). In its simplest application, meta-analysis concerns the collection of effect sizes shared by empirical studies which are then aggregated using a weighted average. The weights are usually a measure of the precision of the primary studies, such as their sample size or their standard errors. Theoretically, such aggregation increases statistical power - i.e. decreases the probability of *type II* errors -

[4]However, as Hedges note, "physicists had been doing meta-analysis for 50 years before the term had been coined; they just called it reviewing [...]" (Hedges, 2015, p. 284)

and allows practitioners to draw more robust conclusions regarding whether an effect exists or not, as opposed to a less precise measure derived from a single study.

Meta-regression analysis involves the regression of the observed effect sizes on one or multiple study characteristics. As such, meta-regression analysis could be seen as a body of statistical tools to find and account for systematic differences in the size of the effect or outcome that is being meta-analyzed (Stanley, 2001).

Meta-analysis and meta-regression analysis can prove particularly useful when the empirical evidences on a topic are mixed or, as in the case of the literature on democratic institutions and economic growth, conflicting. Specifically, meta-analysis could help detecting whether democracies have an impact upon growth, and, if they do, what is the direction and the strength of such influence. Meta-regression analyses, on the other hand, could help unravel what Stanley and Jarrell (1989) call *the specification problem* - the fact that different samples, estimators and covariates gave rise to disparate results. By introducing such specifications as regressors, meta-regressions help assessing whether they have an impact upon the results obtained by the scholars engaged in the democracy and growth debate.

3.1 Fixed and random effect(s) models in meta-analysis

Meta-analysis models are usually divided between fixed and random effects models (Borenstein et al., 2010)[5]. The main intuition behind the former is that the difference in the effect sizes collected is given by sampling variation. In other words, if the sample in each primary analysis converges towards infinity, each study will document the same effect size. Analytically, being $i = 1, \dots, n$ the studies in the meta-analytic sample, the fixed effect model can be written as:

$$y_i = \theta + \varepsilon_i, \quad \varepsilon_i \sim N(0, v_i^2) \tag{2}$$

where y_i is any effect size measure (raw coefficients, partial correlation, t-statistics ...), θ is the underlying (true) effect and ε_i is the error term that is assumed to be normally distributed with mean 0 and variance v_i^2 - the sampling error. Erroneously, few authors confuse the variance components of the fixed effect model with an *homogeneity assumption*; indeed, as mere examples of a spread misunderstanding of the model, Rose and Stanley (2005, p. 351) and De Dominicis et al. (2008, p. 662), after correctly stating that *[in fixed effect models] the different magnitude of the estimates is solely due to sampling variation*, argue that such models *[assume that] there is no heterogeneity across studies*.

Such *homogeneity assumption* is not, at best, entirely correct. As Hedges and Vevea (1998) note, fixed effect models are capable of perfectly valid inferences even under heterogeneity conditions as long as we limit these inferences to the sample of studies collected in the meta-analysis itself. If the practitioners interest lies in inferring only from the set of studies included in the meta-sample, even in presence of *substantial*

[5]Unfortunately, as Gelman and Hill (2006, p. 245) note, within applied statistics there are at least five different, and sometimes conflicting, definitions of what *random* and *fixed* effect(s) models are. This issue is exacerbated by the fact that meta-analysis methods embrace different disciplines, each one of them with its own peculiar *jargon*. To further ravel the problem, the rationales behind meta-analysis *random* and *fixed* effects models are not always flawlessly understood.

heterogeneity among the underlying effects, a fixed effect approach is more appropriate than a random effect one (Hedges and Vevea, 1998).

In other words, the discrimination between the fixed and random effect(s) models should not be based on the potential *heterogeneity* or *homogeneity* of the sample, but rather on the type of inference that we are ultimately interested in performing (Viechtbauer et al., 2010). If researchers objective is to summarize the results of their meta-sample only - what is known as *conditional inference* - then they should consider employing *fixed effect* models. Conversely, when the interest lies in drawing a more general conclusion regarding an empirical relation not bounded to the set of studies included in the meta-analysis - an *unconditional inference* - *random effect(s)* models are suggested.

Unconditional inference is obtained by explicitly assuming that the effect sizes included in the meta-analysis are a random sample of all the effect sizes available or, in other words, that the papers included in the meta-analysis database are a random sample of the population of studies on the topic. Analytically, with $i = 1, \dots, n$ being the studies in the sample, this is given by:

$$y_i = \mu + \eta_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, v_i^2), \quad \eta_i \sim N(0, \tau^2) \tag{3}$$

where y_i is the effect size measure, μ is the underlying average (true) effect, ε_i is the variance components that is assumed to be normally distributed with mean 0 and variance v_i^2 (the sampling error) and η_i is the random effect introduced to model variability among true effects. Equation 3 can be seen a special case of a multilevel/hierarchical model (Konstantopoulos, 2011). It is indeed possible to partition the model into two levels:

$$y_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, v_i^2) \tag{4}$$

equation 4 represents the within-study component of the random effect model. It differs from the fixed effect equation 2 because it introduces a random effect, θ_i, which is defined at the second level - the between-studies part - of the model:

$$\theta_i = \mu + \eta_i, \quad \eta_i \sim N(0, \tau^2) \tag{5}$$

equation 5 explicitly model the variability introduced at the first level among the true effects treating it as purely random. It is easy to show that if the effects sizes are actually homogeneous then $\theta_1 = \theta_2 = \cdots = \theta_i \equiv \theta$, that is, the fixed and random effect models are equivalent. There are however no unbiased instruments to test whether the true effects are really homogeneous or not (Viechtbauer et al., 2010), yet another reason to not choose which model to adopt based on the alleged homogeneity of the sample but, as above-mentioned, on the desired inference[6].

[6]Fixed effect model can be fitted adopting a weighted or unweighted least squares. Conversely, random effect(S) models are estimated through a two-steps approach. First, the residual heterogeneity τ^2 is estimated through one of the several estimators proposed in the literature. Next, the model parameters are estimated via weighted least squares with weights equal to $w_i = 1/(v_i + \hat{\tau}^2)$, with $\hat{\tau}^2$ being the estimate of τ^2. We will perform such estimation by adopting the *Restricted (Residual) Maximum Likelihood Estimator* (*REML*). As Viechtbauer observes, such estimator “strikes a good balance between unbiasedness and efficiency” (Viechtbauer, 2005, p. 291).

3.1.1 Multilevel random effects model

Our sample is composed by 942 effect sizes collected from 103 papers. While there are no reasons to believe that between-studies observations are not independent, it is possible to argue that, within-study, observations are contingent upon each other. Point estimates collected from the same research are in fact likely to use similar samples, specifications and estimation techniques. Such particular data structure is known as *hierarchical dependence* (Stevens and Taylor, 2009); if unaccounted, models' estimation fails in delivering a correct inference (De Dominicis et al., 2008)[7].

To handle *hierarchical dependence* we apply multilevel models, which can naturally handle the hierarchy of our sample Gelman and Hill (2006). Here we will show how they can be applied to meta-analysis random effect models. We begin by assuming $i = 1, \ldots, n$ studies (our level-3 units) and $j = 1, \ldots, k$ observations, the end points collected. The first level of the model, the within-observations part, is composed by:

$$y_{ij} = \theta_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, v_{ij}^2) \tag{7}$$

where y_{ij} are the effect sizes collected from all the different studies, θ_{ij} are the underlying (true) effects and v_{ij}^2 is the variance of the sampling error. The second level (the between-observations within-study part) is then given by:

$$\theta_{ij} = \mu_j + \eta_{ij}, \quad \eta_{ij} \sim N(0, \sigma_W^2) \tag{8}$$

where μ_j is the average (true) outcome at study level and σ_W^2 is the between-observations within-study variance. Finally, at the third and last level, the between-studies part of the model is detailed as:

$$\mu_j = \gamma_{00} + \upsilon_{0j}, \quad \upsilon_{0j} \sim N(0, \sigma_B^2) \tag{9}$$

where γ_{00} is the average (true) outcome and σ_B^2 is the between-studies variance. As all multilevel/hierarchical models, we can write our model on a single level notation, such as:

$$y_{ij} = \gamma_{00} + \upsilon_{0j} + \eta_{ij} + \varepsilon_{ij} \tag{10}$$

where γ_{00} is the average (true outcome), υ_{0j} is the random effect that allows for heterogeneity between-studies, η_{ij} is the random effect that allows for heterogeneity within-study and ε_{ij} is the sampling error.

Equation 10 can be easily extended to accommodate a random effects meta-regression analysis. Since our interest lies in tackling the *specification problem* - that is, finding and

[7] We assess such possibility by computing the *Intra-class Correlation Coefficient* (ICC), which is given by:

$$ICC = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} \tag{6}$$

where σ_B^2 is the between-studies variance component while σ_W^2 is the between-observations within-study variance component (Viechtbauer et al., 2010). We document an ICC value of 0.773, indicating that point estimates collected from the same study are, on average, highly correlated.

estimating the sources of heterogeneity at between-studies level - we add q predictors at the third level of the model. We can then rewrite equation 9 - the between-studies level of the multilevel random effects meta-analysis - as:

$$\mu_j = \gamma_{00} + \gamma_{01}\omega_{1g} + \ldots + \gamma_{0g}\omega_{qg} + \upsilon_{0j}, \quad \upsilon_{0j} \sim N(0, \sigma_R^2) \tag{11}$$

where $\omega_{1g}, \ldots, \omega_{qg}$ are level-3 unit specific predictors and $\gamma_{1g}, \ldots, \gamma_{0g}$ their coefficients. The variance of υ_{0j} is now written as σ_R^2, since it now indicates the residual variance between-studies (Konstantopoulos, 2011). On a single level notation, such three-level random effects meta-regression model can be written as:

$$y_{ij} = \gamma_{00} + \gamma_{01}\omega_{1g} + \ldots + \gamma_{0g}\omega_{qg} + \upsilon_{0j} + \eta_{ij} + \varepsilon_{ij}, \quad \upsilon_{0j} \sim N(0, \sigma_R^2) \tag{12}$$

Since the predictors are added at the third level of the model, the number of observations used to compute the degrees of freedom correspond to the number of level-three units – that is, the number of studies in the meta-sample - and not the number of end-points. Adding too many predictors could therefore (rightfully) raise over-fitting concerns - and possibly multicollinearity issues. While there is no best practice on such an issue yet, we find reasonable the rule of thumb proposed by Van Houwelingen et al. (2002), which argue that practitioners should not exceed one explanatory variable every 5 to 10 level-three units.

4 The meta-sample

The selection process has been conducted following the *reporting guidelines for meta-regression analyses in economics* (Stanley et al., 2013). Such guidelines require meta-analysis practitioners to explicitly state the ex-ante inclusion (and exclusion) criteria adopted while searching the literature of interest.

Explanatory and control variables We began by collecting point estimates from *all* regressions outputs containing a proxy for economic growth as dependent variable and a measure of democratization as independent variable. While such criteria might seem obvious, it is appropriate to issue a caveat here. Econometric models often divide the $\mathbf{X}$ matrix of independent variables into a vector $\mathbf{x}_e$, containing the explanatory variable, and a $\mathbf{X}_c$ matrix, containing the *control variables*. Such model specification is designed to allow readers to focus on the relationship between the dependent variable and the explanatory variable by removing the influence of the control variables. Statistically, however, such partition is only conceptual; $\mathbf{x}_e$ and $\mathbf{X}_c$ are input in the model in the same way. Therefore, when a primary study presents the independent variable in which we are interested in as control variable, we coded it regardless of the aforementioned distinction.

Published and unpublished literature Doucouliagos and Ulubaşoğlu (2008) choose not to include working papers and unpublished articles to control for the research quality. We however agree with Stanley, who argues that "differences in quality, data or methods do not provide a valid justification for omitting studies" (Stanley, 2001, p.

135), but rather that such differences provide rationale for performing a meta-analysis and a meta-regression analysis in the first place. Furthermore, publication status is at least a controversial way to asses the quality of an article, there are other reasons why a paper might not have been published beside its dubious *virtues*.

There are indeed reasons to believe that the published literature might be affected by *publication biases* and the so-called *prejudice against the null hypothesis*, the preference by editors and reviewers of academic journal for studies showing statistical significant results over studies unable to reject the null hypothesis (Rothstein et al., 2006b). Retrieving only published studies might potentially lead to an over-representation of statistically significant results, challenging the random sampling assumptions underlying meta and meta-regression analyses random effects models. We therefore included both published and unpublished papers. A failure in doing so might severely threatens the unbiasedness of the results.

Multiple end-points For the same above-mentioned reasons, we decided not to discriminate between what authors indicate as *best specification* and the other outcomes contained in the primary researches, hence collecting (and computing) all available estimates per study.

Granger causality test Furthermore, while Doucouliagos and Ulubaşoğlu (2008) include in their sample few studies containing *Granger causality test* only, we decided not to do so. Leaving aside the debate over the effectiveness of such methods, we believe the interpretation of its point estimates to be different than the one collected from linear (and non-linear) regressions. Finally, we investigated only articles written in English in which data are collected only from the 1950s onwards. Articles including historical data-sets were excluded from our sample.

Starting on 19 January 2016 we began the literature search process to build the meta-sample; such task was divided in four steps:

1. we started by querying *SSRN*, *RePEc*, *Web of Science*, *Econlit*, *Jstor* and *Science Direct* with the following research string: `"democracy" AND "growth"` (or equivalent). After a preliminary leaf through, 109 papers were added into our sample;

2. the second phase involved the analyses of the bibliographies of the qualitative reviews investigating the democracy and growth literature. This yielded further 19 papers to be added to the initial 109;

3. we then moved towards a *snowballing* sampling approach by using the *Google Scholar* citation system. We identified some of the most cited papers investigating the democracy and growth conundrum and after performing a reverse search upon them we added 12 additional articles to our sample;

4. finally, we performed an inspection of the sample collected by Doucouliagos and Ulubaşoğlu (2008), founding 49 supplementary papers.

We concluded the literature search process by rigorously scrutinizing the sample obtained. Starting on 21 March 2016, we coded the articles that conformed to the ex-ante criteria previously highlighted; studies that failed to comply were excluded from the sample. Overall, the process yielded 942 effect sizes nested around 103 papers comprising 924460 primary observations.

4.1 Effect size measures

We computed two different effect size measures; a Fisher's z-transform of the partial correlation and a discrete indicator that classifies documented point estimates into positive, negative, and non-significant ones at the 10% significance level, allowing a direct comparison with the results of Doucouliagos and Ulubaşoğlu (2008). The partial correlation was computed following Aloe and Thompson (2013):

$$R_p = \frac{t_f}{\sqrt{t_f^2 + df}} \tag{13}$$

where t_f is the t statistic of the regression coefficient β_f while df is the degrees of freedom $(n-p-1)$ in which p is the number of regressors and n is the number of observations. The sample variance of the partial correlation, again following Aloe and Thompson (2013), was instead given by:

$$var(r_p) = \frac{(1-r_p^2)^2}{n-p-1} \tag{14}$$

the values obtained were then normalized accordingly to the *Fisher's z-score* transformation. The motivation behind such choice is twofold; meta-analytic methods usually assume that the sampling distribution of the observed outcomes is (at least approximately) normal. When ρ (the population correlation) in a particular study is far away from 0 and the sample size is small, the sampling distribution of the raw (partial) correlation becomes very skewed and is not well approximated by a normal distribution. *Fisher's z-score* corrects for skewness. Moreover, to calculate the variance, the unknown value of ρ should be estimated. This can be done, as shown in equation 14, by plugging the observed (partial) correlation (i.e. R_p) into the equation. This will provide an estimate of the sampling variance, but this happens to be a rather inaccurate estimate, especially in small samples. On the other hand, the sampling variance of an *Fisher's z-score* transformed correlation is (approximately) equal to:

$$var(z) = \frac{1}{n-3} \tag{15}$$

which no longer depends on unknown quantities. Ultimately, a positive Fisher's z-score indicates a positive relation between the strength of democratic institutions and economic growth, while a negative Fisher's z-score indicates a negative relation [8].

[8]Social scientists have developed several indexes to (try to) quantify the qualitative concept of democracy. However, not all the indexes have the same direction. Therefore, before performing our analysis, we transformed all our effect sizes on a positive scale - i.e. the lowest the values the lowest is the level of democratization, the highest the value the highest is the *democratic score*.

4.2 Descriptive statistics

With regards to the dependent variables, the large majority of the effect sizes included in our sample (i.e. 793 versus 149) are computed by adopting GDP growth rates rather than GDP levels.

According to our discrete indicator, 34 papers within the meta-sample exhibit, on average, a positive and statistically significant relation between democracy and growth, 19 a negative and significant one, while the vast majority (50) reports non-significant coefficients.

By analyzing the 942 end-points collected, we observe that 360 are statistically insignificant at the 10% while 201 present a negative direction and 380 a positive one. Results however scatter widely depending on the democracy index used.

	Negative	Non-significant	Positive	Total
Study average	19	50	34	103
Effect sizes				
Polity Index	73	106	116	295
Freedom House	61	142	132	335
Dummy Variable	34	49	76	159
Other	33	64	56	153
Total	201	361	380	942

Table 1: Effect size discrete indicator at 10% significance level

At a glance, effect size estimated in regressions including as proxy for democracy the *Freedom House Indexes* are more likely to point towards a positive impact. Conversely, if democratic institutions are captured by dummy variables or by the *Polity Index*, outcomes are more scattered. These findings seem to suggest that the independent variable adopted might help explaining the heterogeneous results obtained in the literature. Yet, our sample is unbalanced, with some studies contributing with more effect sizes than others. Five studies detailed more than 50 regression coefficients while 33 documented three or less effect sizes, six of which only one. With regards to the year of publication, which impact on the sample time span, the oldest article within our sample has been published in 1985 while the newest one in 2016. There are 4 researches from the 1980s, 30 from the 1980s, 40 from the 2000s and 29 reporting a date after 2010. Therefore, we advise caution in drawing preliminary conclusions from the discrete indicator.

We then defined articles as *published* if their publication outlet is included in the *Research Papers in Economics* (RePEc) database. Overall, 689 point estimates are collected from 80 published articles while 253 from 23 unpublished papers. 12 papers appears on political science journal series while 89 belong to economics-focused publications. Moreover, 75 out of 103 studies have been conducted by at least one author affiliated with an institution located in the UK or in the US at the year of the publication.

The estimation techniques scatter widely. While the majority of the effect sizes are estimated via *OLS* (328), 199 end-points are the result of a *(system) GMM* approach and 76 estimations are conducted by means of a *within estimator*. 301 effect sizes are results of techniques – such as the 2SLS – that tries to account for a potential endogenous relation between democracy and growth while 641 treat the above-mentioned relation as exogenous. Additionally, there are 87 estimations that account for a potential non-linear effect in the democracy and growth nexus.

Covariate	Included	Excluded	Inclusion ratio
Convergence	825	117	0.87
Human capital	490	452	0.52
Investments	399	543	0.42
Population size	365	577	0.39
Trade openness	348	594	0.37
Government Size	255	687	0.27
Instability	190	752	0.20
Inflation	177	765	0.19
Rule of law	169	773	0.18
Property rights	127	815	0.13
Corruption	115	827	0.12
Ethnicity	58	884	0.06
Inequality	53	859	0.06
Physical capital	52	890	0.06
Economic freedom	43	899	0.05
Stateness	20	922	0.02
Religion	8	934	0.01
Culture	8	934	0.01

Table 2: List of covariates collected and their inclusion and exclusion rate

Finally, specifications are, as expected, widely heterogeneous. As shown in table 2 we tried to account for some of the most common control variables. Out of the 942 effect sizes collected the vast majority (825) include a variable to capture the convergence process, 490 a proxy for human capital and 399 a measure of investments. Conversely, only 53 effect sizes include measures of economic inequality and 43 a proxy for economic freedom. As well as for inequality and economic freedom, cultural variables such as religion and ethnicity are scarcely adopted by the scholars recorded in our sample with, respectively, only 8 and 58 end-points[9].

5 Results

According to the fixed effect model shown in table 3, the impact of democratic institutions on economic development, while weak in its magnitude, it is positive and significant (1%). This seems to confirm the intuition firstly expressed by North (1990): governments are tempted to appropriate the economic gains generated by individuals and intermediate groups and "only democratic institutions can constrain them to act in general interest" (Przeworski and Limongi, 1993, p. 144).

We must however note that, as explained in section 3, the inference obtained from the fixed effect model cannot be generalized. This inference is bounded to the 103 papers that compose our sample.

To overcome the limits of such *conditional inference* we implement the random effects model detailed in section 3.1.1. By explicitly assuming that the studies in our sample are just an unknown proportion of the population of the researches investigating

[9]Further descriptives and the meta-sample composition are available in section A in table 6.

Multilevel Meta-Analysis Model (y = 929; method: FE)					
Variance Components: none					
Model Results:					
estimate	se	zval	pval	ci.lb	ci.ub
0.0358***	0.0010	34.3106	<.0001	0.0337	0.0378

Table 3: Fixed effect multilevel meta-analysis. *, ** and *** denotes significance at the 10%, 5% and 1% levels; results reported are, in order, the estimated coefficient, standard errors, p and z values and lower and upper bounds of the confidence interval.

the relationship between democracy and growth, we can extend our conclusions and perform an *unconditional inference*.

The outcome shown in table 4 confirms the substantial inconclusiveness of the literature on democracy and growth; while the coefficient is still positive and of a similar magnitude as the one obtained in the fixed effect model, the multilevel random effects model shows a weak statistically significant impact of democratic institutions on economic growth (10%). Furthermore, the lower bound is now within the negative region. Overall, the democracy and growth nexus is still a *conundrum*.

Multilevel Meta-Analysis Model (y = 929; method: REML)					
Variance Components:					
	estim	sqrt	levels	fixed	factor
σ_1^2	0.0301	0.1735	102	no	study
σ_2^2	0.0110	0.1047	929	no	study/case
Model Results:					
estimate	se	zval	pval	ci.lb	ci.ub
0.0324*	0.0187	1.7373	0.0823	-0.0042	0.0690

Table 4: Random effects multilevel meta-analysis. *, ** and *** denotes significance at the 10%, 5% and 1% levels; results reported are, in order, the estimated coefficient, standard errors, p and z values and lower and upper bounds of the confidence interval.

To identify the drivers behind these results, we start by investigating the between-studies heterogeneity of our sample. Both of our models report a chi-squared statistic (also known as *Cochran's Q*) of 7037. Such statistical test is computed by summing the squared deviations of each study estimate from the overall meta-analytic estimate, using the same weights applied to meta-analysis models.

To interpret the meaning of the *Cochran's Q*, we resort to the approach proposed by Higgins et al. (2003) - the I^2, that measures the between-studies consistency in a meta-analysis as:

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100 \qquad (16)$$

where df are the degrees of freedom. I^2 yields a value of 87%, which, accordingly to the rule of thumb proposed by Higgins and Thompson (2002), indicates a *considerable heterogeneity*. We investigate such high between-studies variance through the multilevel regression model detailed in equation 12. Results and variables definition are outlined in table 5. Results are robust to alternative model specifications.

	Multilevel meta-regression analysis						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
intercept	0.022	0.096**	0.032	0.035	0.056**	-0.013	-0.064
	(-0.001)	0.041	(0.048)	(0.034)	(0.0.26)	(0.040)	(0.042)
Proxy for democracy							
Polity	-0.001	-0.002	0.004	-0.001			0.0021
	(0.020)	(0.020)	(0.017)	(0.020)			(0.020)
FH	0.043**	0.045**	0.0477***	0.047**			0.038**
	(0.019)	(0.019)	(0.016)	(0.020)			(0.019)
Dummy	-0.050**	-0.056**	-0.041**	-0.051**			-0.041*
	(0.023)	(0.023)	(0.020)	(0.024)			(0.023)
Proxy for growth							
GDP gr.		-0.084**	-0.079**				
		(0.038)	(0.034)				
World Bank regions							
Africa			0.251***		0.254***	0.255***	
			(0.027)		(0.027)	(0.028)	
E. Asia & P.			0.027		0.055	0.035	
			(0.059)		(0.059)	(0.060)	
Europe & C. Asia			0.259***		0.273***	0.265***	
			(0.042)		(0.049)	(0.043)	
High Inc.			0.020		0.020	0.017	
			(0.021)		(0.021)	(0.022)	
Lat. Am. & Carib.			-0.161***		-0.163***	-0.162***	
			(0.031)		(0.030)	(0.031)	
MENA			-0.061*		-0.074**	-0.061*	
			(0.035)		(0.035)	(0.036)	
S. Asia			-0.274***		-0.297***	-0.286***	
			(0.051)		(0.051)	(0.052)	
Estimation techniques							
OLS Est.					-0.029		
					(0.018)		
Within Est.					-0.061***		
					(0.018)		
GMM Est.					0.036**		
					(0.016)		
Endogenous					-0.025*		
					(0.015)		
Non-linear					-0.116***		
					(0.029)		
Covariates							
Inflation				0.027		0.025	
				(0.024)		(0.022)	
Population				-0.053**		-0.044**	
				(0.025)		(0.022)	
Corruption				-0.044		-0.039	
				(0.053)		(0.052)	
Rule of law				-0.016		-0.007	
				(0.041)		(0.039)	
Eco. free				-0.066		-0.049	
				(0.048)		(0.045)	
Prop. rights				0.011		0.012	
				(0.051)		(0.048)	
Instability				-0.027		-0.022	
				(0.021)		(0.018)	
Human c.				-0.024		-0.037	
				(0.025)		(0.023)	
Physical c.				0.034		0.074	
				(0.046)		(0.044)	
Converg.				0.039		0.037	
				(0.025)		(0.025)	
Opennes				-0.010		-0.012	
				(0.017)		(0.014)	
Gov. size				-0.014		-0.002	
				(0.023)		(0.021)	
Investment				0.010		0.007	
				(0.020)		(0.018)	

Religion	0.027	0.002	
	(0.068)	(0.061)	
Ethnicity	-0.005	-0.002	
	(0.052)	(0.048)	
Time periods			
1950s			0.022
			(0.035)
1960s			-0.089***
			(0.025)
1970s			0.056**
			(0.024)
1980s			0.114***
			(0.025)
1990s			0.033
			(0.029)
2000s			-0.078***
			(0.028)

Table 5: *, ** and *** denotes significance at the 10%, 5% and 1% levels; standard errors in parenthesis. Variables definition; Proxy for democracy identifies the independent variable adopted. Baseline is *Other*, a dummy = 1 that captures all the effect sizes that adopt neither the Polity Index, nor the Freedom House Index, nor Dummy variables to gauge democracy. Proxy for growth is GDP growth, a dummy = 1 if the effect size dependent variable is a measure of GDP (both per capita and aggregate) growth. Baseline is GDP level, which captures GDP (both per capita and aggregate) levels. Spatial characteristics is based on the *World Bank* regions' definitions and each variable assumes the value of 1 if at least one country in the sample adopted in the primary research is in that specific region. Estimation techniques: *OLS est.*, *Within est.* and *GMM est.* are dummies that captures, respectively, OLS, within estimation and GMM estimation techniques. *Endogenous* is equal 1 if the study uses models that tries to account for the potential endogeneity between democracy and economic growth. *Non-linear* is equal 1 if the model includes a non-linear specification for democracy. Covariates indicates, through dummies, variables listed as inputs in the primary studies. Time characteristics is a set of 6 dummies related to the decades included in the primary studies.

Proxy for democracy The first panel of table 5 shows that the choice of alternative proxies for measuring democracy weakly affects the strength and direction of the relationship. In particular, the choice of the popular Freedom House Indices significantly increases the probability of finding a positive relationship between democracy and growth, while the opposite is true when democracy is proxied by a dichotomous variable simply discriminating between democracy and autocracy. A similar result is obtaining by using the Polity Index, although this coefficient is very small in size and not significant. These first results suggest that when democracy is measured according to procedural definitions, its effect on growth is more likely to be neglectable or even slightly negative. Conversely, when a more substantial definition is chosen, as in the case of the Freedom House indicators, it is more likely to capture a positive relationship: however, in this case it is not clear whether the effect is truly driven by the regime type or by other institutional components included in such broader definitions of democracy.

Proxy for growth The second panel simply show that the probability of obtaining a negative relation is higher when the dependent variable is proxied by GDP growth rates than GDP levels, although the magnitude of the coefficient is small.

World Bank regions This panel refers to the group of countries included in the primary studies. As table 5 shows, results are highly driven by the countries and regions that researchers choose to investigate. The size of these effects is the largest among those that we estimate.

Following the regional classification used by the *World Bank*, we discover that democratic institutions have a strong and positive impact (1%) on economic growth in (Sub-Sahara) African and (Eastern) European countries. If such regions are included within the samples, scholars are more likely to report the *bright side* of democracy. It is likely that in these countries democratic institutions intertwine with other institutional settings that support total output generation (Barro, 1991; Przeworski and Limongi, 1993).

The opposite is true for Latin America (1%) and South Asia (1%), where democracy has a detrimental effect on growth. The same applies to Middle East and North Africa, albeit with a smaller magnitude and significant level (5 to 10%, depending on model specification). In these countries, the theories of Huntington (1968) and Huntington and Dominguez (1975) seems to prevail. Democracy, through an increased demands for current consumption, reduces investment and hinders growth (Alesina et al., 1992; Persson and Tabellini, 2003): here, authoritarian regimes can overcome such issue and help countries to reach their growth potential[10].

Our findings are mostly consistent with the theoretical and empirical literature on the subject. In Latin America, where income inequality has historically been endemically high, claims for redistributions by lower income groups drive democratic institutions to promote populist policies, which in turn contribute to bad economic performances (Sachs, 1989; Acemoglu and Robinson, 2001). Similarly, in South Asia, the lobbying power of some labor and industrial groups can lead to an inefficient investment allocation in democratic regimes, by promoting rent-seeking related inefficiencies. Against this background, authoritarian political elites can have the autonomy needed to promote economic growth without being restrained by rent-seekers' pressures (Kriechkhaus, 2006).

Conversely, democracy enhances growth opportunities in Africa, where clientelism has historically been regarded as the region's main political economy feature (Wantchekon, 2003; Vicente and Wantchekon, 2009) and the first cause behind the continent economic fragilities (Sandbrook and Barker, 1985). Indeed, authoritarian regimes, which protect clientelistic interests, more likely allow corrupted politicians to plunder economic gains (Krieckhaus, 2006).

Finally, no matter the specification we adopt, we find that the presence of East Asian countries in the sample does not seem to matter for the results that scholars have obtained. This is in contrast with the literature focusing on the positive impact of autocratic institutions on economic growth in East Asia (Krieckhaus, 2006). Accordingly to

[10]This interpretation seems particularly fitting the cases of some of the *Asian Tigers* and of patrimonialistic states in the MENA region.

such literature, political elites, without the restraints imposed by democratic institutions, can commit to promote rapid industrialization over all others social objectives, hence fostering GDP growth (Cumings, 1984).

Overall, our results indicate that the democracy and growth nexus is not ubiquitous. The impact of democratic institutions on economic growth is largely driven by regional specific aspects, suggesting that some countries or regions might enjoy economic development under autocracies, while others prosper under democratic institutions. Further cross-country and cross-region analysis should therefore be drafted accordingly.

Estimation techniques The next panel in table 5 shows that alternative estimation techniques tend to drive different outcomes. In particular, researchers adopting within estimators are more likely to report the *dark side* of the relationship between democracy and growth. The same applies for those researches that try to model non-linearities. Considering that *Fisher's Z* corrected partial correlation between democracy and growth indicates that a non-linear effect might exist, future researches on the topic should then take it into account.

Effect sizes estimated by means of System GMM estimation techniques report a positive and significant (5%) result, albeit small in magnitude. Probably this result suggests that accounting for the dynamics of the democratization and growth processes lead to find a positive nexus between democracy and growth. However, the size of such nexus is indeed very small (almost one-tenth) when compared to most regional dummies.

Moreover, controlling for *endogeneity* seems to have a small and weakly significant effect on the outcomes. Yet, there is a vast literature on the reverse causality relation between democratic institutions and growth (e.g. Lipset, 1959; Barro, 1999). However, it must be considered that instrumental variable(s) (IV) estimators (the most common technique used to addressed this issue) have larger standard errors than OLS and fixed effect (FE) estimators. Therefore, it is more difficult to find a statistically significant effect while accounting for endogeneity. This issue could explain why the size and significance of our *Endogenous* control variable are smaller than expected.

Control variables The largest panel in table 5 reports the results relating to most of the control variables that scholars have used over the years to model the relation between democracy and growth. Our meta regression models reports some strong and statistically significant coefficients, as shown above. However, the fact the almost all control variables-related coefficients are not significant is probably the most striking outcome of our analysis: *specifications of the augmented production function do not matter*. With the only exception of the growth rate of population, all the covariates included in the models we analyzed do not affect the results.

Such outcome, which is robust to different specifications is remarkable considering the attention that researchers have dedicated at growth regressions' covariates over the last three decades (Sala-i Martin, 1997).

Time periods Finally, the last panel reports a further important result of our analysis: not only regions, but also *time matters*. The coefficients of the dichotomous variables included in the bottom panel of table 5 highlight that the studies' time span does matter for the sign of the relationship between democracy and growth. In particular, including the 1960s increases the probability of observing a negative relationship. This result is consistent with the fact that during the 1960s a relevant subset of democratizing countries were experiencing the de-colonization phase. Thus, despite a formal increase in their democracy levels, they were also experiencing economic turmoils, hence low (or even negative) growth rates. Conversely, including 1970s and 1980s increases the probability of obtaining a positive relationship. The progressive stabilization of the de-colonization processes, and the begin of the downturn of the Soviet block could be interpreted as a golden age of the democracy and growth relationship. Finally, the 2000s crises drive instead the negative and significant coefficient of this dummy.

6 Conclusion

The relationship between democracy and economic growth has long been investigated both in the political science and in the economic literature with inconclusive outcomes. By adopting the meta-analysis framework, we tried to shed lights on this conundrum.

We began by creating a hierarchical sample of 103 studies containing 942 observations - the largest sample of effect sizes dealing with such an issue. The *conditional inference* indicates that democracy has a positive effect on economic growth; however, when we extend such inference to the population of potentially existing studies by adopting a random effects model - to perform an *unconditional inference* - we find that the impact of democracy, albeit weakly positive, loses much of its statistical strength.

Our meta-analysis documents a high degree of between-studies heterogeneity ($I^2 =$ 87%) that we investigated by adopting meta-regression techniques. Results are striking: the effect sizes collected are largely driven by spatial and time differences in the sample, indicating that the democracy and growth nexus is not ubiquitous across world's regions and over time, but largely depend on the characteristics of the regions and periods themselves.

Which characteristics are relevant is however less clear. The (several) control variables included in the papers to account for potential confounding effects, do not impact the reported point estimates. At the same time, models estimated by means of the *within estimator* have a significant, albeit negative, impact on economic growth.

Considering that the latter have been used to control for omitted variables - or better, time-constant unobserved heterogeneity - there are number of possibilities to explain such result.

First: despite their best efforts, scholars have not yet found the control variable(s) that matter(s) to explain such much-debated relationship.

Second: had the covariates, controlling for the democracy and growth nexus, been correctly identified, the way to measure them is still an *unknown unknown* - e.g. how to measure property rights or economic freedom is still matter of debate.

Third, and consequently: we cannot rule out the possibility that there is no such thing as a control variable, measured by a unique proxy, able to explain the relation between democracy and growth always and everywhere.

Our findings suggest that further research is needed to better assess which explanation is more fitted. As of today, the drivers of the relation between democracy and growth are still largely a conundrum. Standing from our results, the debate is unlikely to end soon.

A Meta-sample: further descriptives and forest plot of effect sizes

Author(s)	pc.z.mean	pc.z.sd	pc.z.min	pc.z.max
Acemoglu et al. (2014)	0.036	0.014	0.013	0.092
Acemoglu et al. (2008)	0.097	0.008	0.082	0.102
Adams & Klobodu (2016)	-0.007	0.173	-0.195	0.147
Aghion et al. (2007)	0.081	0.070	-0.034	0.176
Aisen & Veiga(2013)	-0.044	0.048	-0.086	0.001
Albornoz & Dutta (2007)	-0.069	0.297	-0.438	0.288
Alesina & Rodrik (1994)	-0.009	0.024	-0.025	0.008
Alesina et al. (1996)	0.042	NA	0.042	0.042
Ali (2003)	-0.048	0.254	-0.432	0.330
Ali & Crain (2001)	-0.001	0.020	-0.015	0.014
Almeida & Ferreira (2002)	-0.013	0.057	-0.078	0.021
Antic (2004)	-0.020	0.025	-0.037	-0.002
Assane & Pourgerami(1994)	0.049	0.035	0.003	0.099
Assiotis & Sylwester (2014)	0.094	0.117	0.022	0.631
Assiotis & Sylwester (2015)	0.218	0.135	0.121	0.414
Baklouti & Boujelbene (2015)	-0.346	NA	-0.346	-0.346
Barro & Lee (1993)	0.001	0.309	-0.217	0.220
Barro (1996a)	0.200	0.100	0.052	0.328
Barro (1996b)	0.086	0.130	-0.081	0.231
Barro (2000)	0.122	0.004	0.119	0.125
Baum & Lake (2003)	-0.006	0.013	-0.016	0.003
Bleaney & Nishiyama (2002)	0.467	0.068	0.380	0.543
Bluedorn (2001)	-0.124	0.047	-0.185	-0.070
Butkiewicz & Yanikkaya (2007)	-0.057	0.046	-0.115	-0.017
Chatterji (1998)	0.217	0.034	0.184	0.264
Chen (2003)	0.598	0.234	0.248	0.741
Chousa et al. (2006)	0.159	NA	0.159	0.159
Collier (1998)	0.184	NA	0.184	0.184
Collier (2000)	0.091	0.074	-0.015	0.252
Comeau (2003)	-0.369	0.226	-0.548	0.226
Dasgupta et al. (2013)	0.008	0.008	0.002	0.013
Dawson (1998)	0.066	0.040	0.038	0.094
De Haan & Siermann (1995)	0.118	0.045	0.003	0.180
De Haan & Siermann (1996a)	0.022	0.070	-0.065	0.124
De Haan & Siermann (1996b)	0.005	0.402	-0.768	0.348
De la Croix & Delavallade (2011)	-0.055	0.016	-0.090	-0.044
De Luca et al. (2013)	0.088	NA	0.088	0.088
Deana & Gamba (2008)	-0.016	0.022	-0.032	-0.001
Decker & Lim (2008)	0.035	0.073	-0.120	0.313
Dias & Tebaldi (2012)	0.036	0.017	0.008	0.057
Diebolt et. al (2013)	-0.310	0.125	-0.422	-0.112
Doucouliagos & Ulubasoglu (2006)	0.004	0.045	-0.028	0.036
Drury et al. (2006)	-0.005	0.030	-0.025	0.030
Durham (1999)	0.010	0.056	-0.065	0.099
Esfahani & Ramirez (2003)	-0.151	0.090	-0.234	0.003
Fedderke & Klitgaard (1998)	0.093	0.186	-0.201	0.315
Feng (1996)	0.391	0.146	0.093	0.465
Feng (1997)	-0.233	0.112	-0.399	-0.159
Feng (1995)	0.093	0.182	-0.213	0.376
Fida & Zakaria (2011)	-0.531	0.254	-1.158	-0.260
Fidrmuc (2003)	0.147	0.288	-0.673	0.614
Flachaire et al. (2014)	0.022	0.076	-0.089	0.082
Fosu (2008)	-0.251	0.189	-0.477	0.079
Gasiorowski (2000)	-0.077	0.025	-0.105	-0.057
Gerring et al. (2013)	0.023	0.020	-0.022	0.042
Gerring et al. (2005)	0.084	0.074	-0.017	0.278
Ghosh et al. (2013)	0.064	0.174	-0.108	0.251
Glaeser et al. (2004)	0.230	NA	0.230	0.230
Goldsmith (1995)	0.359	0.096	0.292	0.427
Grier & Tullock (1989)	0.151	0.120	-0.024	0.270

Grindler & Krieger (2015)	0.075	0.043	-0.002	0.160
Gupta et al. (1998)	0.059	NA	0.059	0.059
Gupta (1988)	-0.239	0.157	-0.351	-0.128
Gwartney et al. (1999)	-0.208	0.203	-0.442	-0.084
Haggard & Tiede (2011)	-0.009	0.045	-0.071	0.108
Heckelman (2010)	0.812	0.793	-0.436	1.670
Helliwell (1994a)	-0.583	NA	-0.583	-0.583
Helliwell (1994b)	-0.048	0.101	-0.128	0.065
Henisz (2000)	0.170	0.100	0.086	0.307
Iqbal et al. (2012)	0.372	0.051	0.331	0.443
Jacob & Osang (2015)	-0.018	0.038	-0.097	0.096
Jalles (2010)	0.113	0.064	0.010	0.192
Jamali et al. (2007)	0.120	0.058	0.079	0.161
Jaunky (2013)	0.022	NA	0.022	0.022
Jetter (2014)	0.022	0.038	-0.046	0.102
Kagochi et al. (2007)	0.309	0.000	0.309	0.309
Kang et al. (2013)	-0.180	0.074	-0.240	-0.097
Ken Farr et al. (1998)	0.034	0.217	-0.120	0.187
Knack & Keefer (1995)	0.050	0.057	-0.020	0.119
Knutsen (2013)	0.099	0.039	0.041	0.168
Kormendi & Meguire (1985)	0.168	0.145	0.065	0.271
Krieckhaus (2006)	-0.062	0.771	-1.978	1.531
Krieckhaus (2004)	-0.064	0.259	-0.529	0.295
Kurzman et al. (2002)	-0.071	0.028	-0.103	-0.045
Landau (1986)	-0.107	0.044	-0.186	-0.037
Leblang (1997)	0.184	0.051	0.127	0.238
Leblang (1996)	0.105	0.030	0.084	0.127
Li & Zou (1998)	-0.157	0.261	-0.603	0.061
Lopes & de Jesus (2015)	0.150	0.122	-0.057	0.272
Lundberg & Squire (2003)	-0.462	NA	-0.462	-0.462
Masaki & Van de Walle (2014)	0.053	0.023	0.001	0.077
Masaki & Van de Walle (2014)	0.053	0.023	0.001	0.077
Mbaku & Kimenyi (1997)	0.470	0.194	0.167	0.829
Miguel et al. (2004)	0.000	NA	0.000	0.000
Minier (2003)	0.175	0.152	0.022	0.386
Minier (1998)	0.164	0.122	0.030	0.284
Mo (2000)	-0.243	0.098	-0.444	-0.074
Mo (2001)	-0.345	0.191	-0.722	-0.181
Mo (2015)	-0.194	0.086	-0.289	-0.061
Mobarak (2005)	-0.021	0.094	-0.198	0.146
Nelson & Singh (1998)	0.160	0.004	0.157	0.165
Oliva & Rivera-Batiz (2002)	0.183	0.057	0.096	0.297
Owen et al. (2009)	0.015	0.012	0.007	0.024
Zouhaier & Karim (2012)	-0.180	0.096	-0.248	-0.112

Table 6: Summary statistics of partial correlation transformed through *Fisher's Z score*) of estimations reported in primary studies. *pc.z.mean* is the average, *pc.z.sd* the standard deviation, *pc.z.min* and *pc.z.max* are respectively the lower and upper bounds. Studies with *pc.z.sd* = NA only include a single effect size and studies in which the *Fisher's Z score* partial correlation could not be computed.

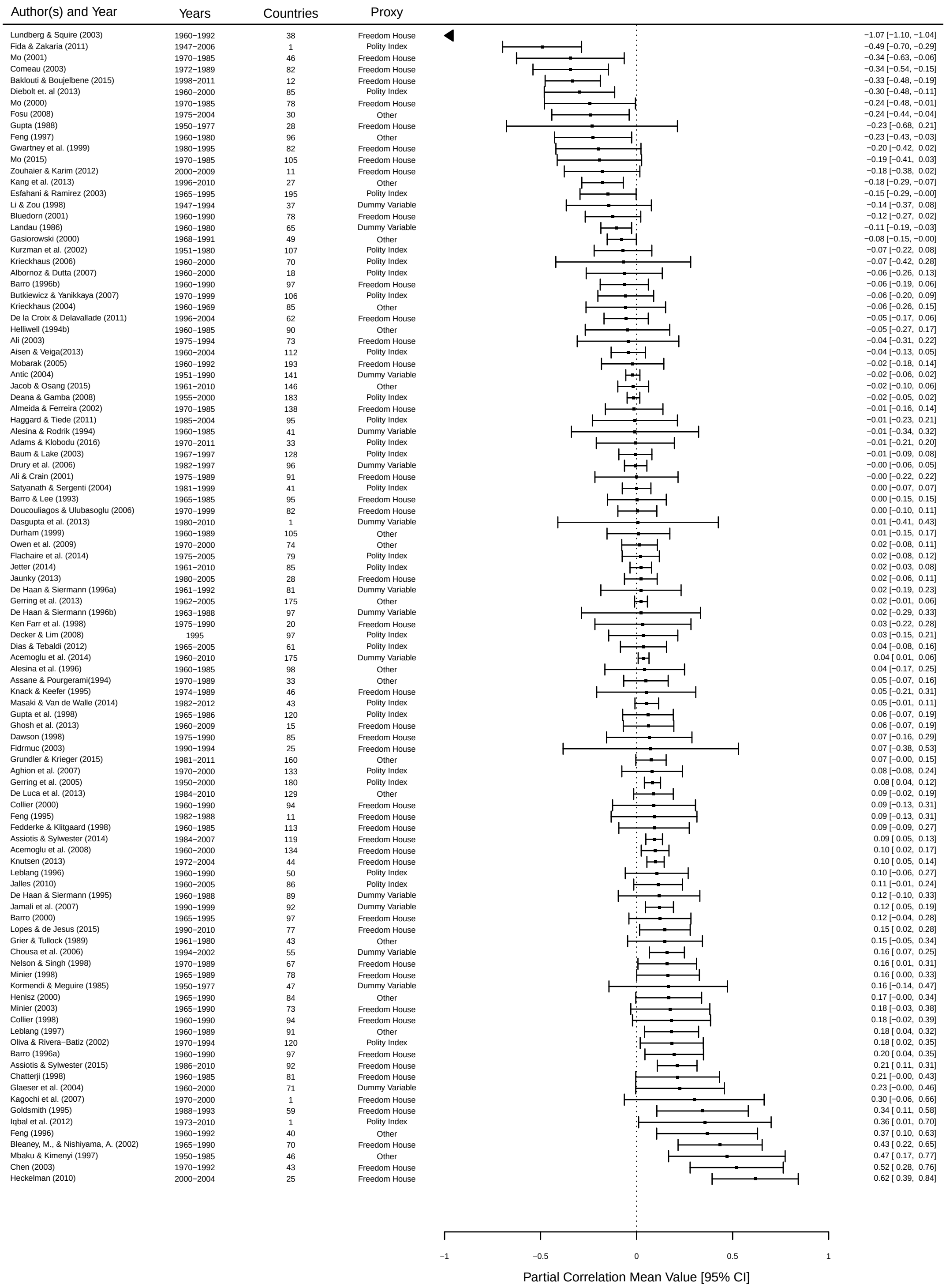

Figure 1: Forest plot of raw partial correlation effect sizes. *Years* is the average time-span of the primary analysis. *Countries number* is the average number of countries investigated within the study. *RePec 10 if* is the RePec 10 years discounted impact factor. 95% confidence interval.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review 91*(5), 1369–140l.

Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (2014). Democracy does cause growth. Technical report, National Bureau of Economic Research.

Acemoglu, D. and J. A. Robinson (2001). A theory of political transitions. *American Economic Review*, 938–963.

Alesina, A., G. D. Cohen, and N. Roubini (1992). Macroeconomic policy and elections in oecd democracies. *Economics & Politics 4*(1), 1–30.

Aloe, A. M. and C. G. Thompson (2013). The synthesis of partial effect sizes. *Journal of the Society for Social Work and Research 4*(4), 390–405.

Barro, R. J. (1991). Economic growth in a cross section of countries. *The quarterly journal of economics 106*(2), 407–443.

Barro, R. J. (1996). Democracy and growth. *Journal of economic growth 1*(1), 1–27.

Barro, R. J. (1999). Determinants of democracy. *Journal of Political economy 107*(S6), S158–S183.

Barro, R. J. et al. (2003). Determinants of economic growth in a panel of countries. *Annals of economics and finance 4*, 231–274.

Barro, R. J. and X. Sala-i Martin (1992). Convergence. *Journal of political Economy 100*(2), 223–251.

Borenstein, M., L. V. Hedges, J. Higgins, and H. R. Rothstein (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods 1*(2), 97–111.

Cass, D. (1965). Optimum growth in an aggregative model of capital accumulation. *The Review of economic studies 32*(3), 233–240.

Cumings, B. (1984). The origins and development of the northeast asian political economy: industrial sectors, product cycles, and political consequences. *International Organization 38*(01), 1–40.

De Dominicis, L., R. J. Florax, and H. L. De Groot (2008). A meta-analysis of the relationship between income inequality and economic growth. *Scottish Journal of Political Economy 55*(5), 654–682.

De Haan, J., S. Lundström, and J.-E. Sturm (2006). Market-oriented institutions and policies and economic growth: A critical survey. *Journal of Economic Surveys 20*(2), 157–191.

De Long, J. B. and K. Lang (1992). Are all economic hypotheses false? *Journal of Political Economy 100*(6), 1257–1272.

Doucouliagos, H. and M. A. Ulubaşoğlu (2008). Democracy and economic growth: a meta-analysis. *American Journal of Political Science 52*(1), 61–83.

Gelman, A. and J. Hill (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research 1. *Educational researcher 5*(10), 3–8.

Gwartney, J. D., R. G. Holcombe, and R. A. Lawson (2006). Institutions and the impact of investment on growth. *Kyklos 59*(2), 255–273.

Hedges, L. V. (2015). The early history of meta-analysis. *Research synthesis methods 6*(3), 284–286.

Hedges, L. V. and J. L. Vevea (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods 3*(4), 486.

Helliwell, J. F. (1994). Empirical linkages between democracy and economic growth. *British journal of political science 24*(02), 225–248.

Higgins, J. and S. G. Thompson (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine 21*(11), 1539–1558.

Higgins, J., S. G. Thompson, J. J. Deeks, D. G. Altman, et al. (2003). Measuring inconsistency in meta-analyses. *British medical journal 327*, 557–560.

Hodgson, G. M. (2006). What are institutions? *Journal of Economic Issues 40*(1), 1–24.

Huntington, S. P. (1968). *Political order in changing societies*. Yale University Press.

Huntington, S. P. and J. I. Dominguez (1975). Political development. *Handbook of political science 3*, 1–114.

Huntington, S. P. e. a. (1987). *Understanding political development: An analytic study*. Scott Foresman & Company.

Knutsen, C. H. (2012). Democracy and economic growth: A survey of arguments and results. *International Area Studies Review 15*(4), 393–415.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods 2*(1), 61–76.

Koopmans, T. C. et al. (1965). On the concept of optimal economic growth.

Krieckhaus, J. (2006). Democracy and economic growth: how regional context influences regime effects. *British Journal of Political Science 36*(02), 317–340.

Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *American political science review 53*(01), 69–105.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics 22*(1), 3–42.

Minier, J. A. (1998). Democracy and growth: Alternative approaches. *Journal of economic growth 3*(3), 241–266.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge university press.

Olson, M. (1993). Dictatorship, democracy, and development. *American Political Science Review 87*(03), 567–576.

Persson, T. and G. Tabellini (2003). Do electoral cycles differ across political systems? Technical report, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Przeworski, A. (2000). *Democracy and development: Political institutions and well-being in the world, 1950-1990*, Volume 3. Cambridge University Press.

Przeworski, A. and F. Limongi (1993). Political regimes and economic growth. *The journal of economic perspectives 7*(3), 51–69.

Rodrik, D., A. Subramanian, and F. Trebbi (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of economic growth 9*(2), 131–165.

Rodrik, D. and R. Wacziarg (2005). Do democratic transitions produce bad economic outcomes? *The American economic review 95*(2), 50–55.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy 98*(5, Part 2), S71–S102.

Rose, A. K. and T. D. Stanley (2005). A meta-analysis of the effect of common currencies on international trade. *Journal of economic surveys 19*(3), 347–365.

Rothstein, H. R., A. J. Sutton, and M. Borenstein (2006a). Publication bias in meta-analysis. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 21–27.

Rothstein, H. R., A. J. Sutton, and M. Borenstein (2006b). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.

Sachs, J. D. (1989). Social conflict and populist policies in latin america.

Sala-i Martin, X. X. (1997). I just ran two million regressions. *The American Economic Review*, 178–183.

Sandbrook, R. and J. Barker (1985). *The politics of Africa's economic stagnation*. Cambridge University Press.

Sirowy, L. and A. Inkeles (1990). The effects of democracy on economic growth and inequality: A review. *Studies in Comparative International Development 25*(1), 126–157.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics 70*(1), 65–94.

Stanley, T., H. Doucouliagos, M. Giles, J. H. Heckemeyer, R. J. Johnston, P. Laroche, J. P. Nelson, M. Paldam, J. Poot, G. Pugh, et al. (2013). Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys 27*(2), 390–394.

Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *The Journal of Economic Perspectives 15*(3), 131–150.

Stanley, T. D. and S. B. Jarrell (1989). Meta-regression analysis: A quantitative method of literature surveys. *Journal of economic surveys 3*(2), 161–170.

Stevens, J. R. and A. M. Taylor (2009). Hierarchical dependence in meta-analysis. *Journal of Educational and Behavioral Statistics 34*(1), 46–73.

Van Houwelingen, H. C., L. R. Arends, and T. Stijnen (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in medicine 21*(4), 589–624.

Vicente, P. C. and L. Wantchekon (2009). Clientelism and vote buying: lessons from field experiments in african elections. *Oxford Review of Economic Policy 25*(2), 292–305.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics 30*(3), 261–293.

Viechtbauer, W. et al. (2010). Conducting meta-analyses in r with the metafor package. *J Stat Softw 36*(3), 1–48.

Wantchekon, L. (2003). Clientelism and voting behavior: Evidence from a field experiment in benin. *World politics 55*(03), 399–422.

Wasserstein, R. L. and N. A. Lazar (2016). The asa's statement on p-values: context, process, and purpose. *Am Stat 70*(2), 129–133.