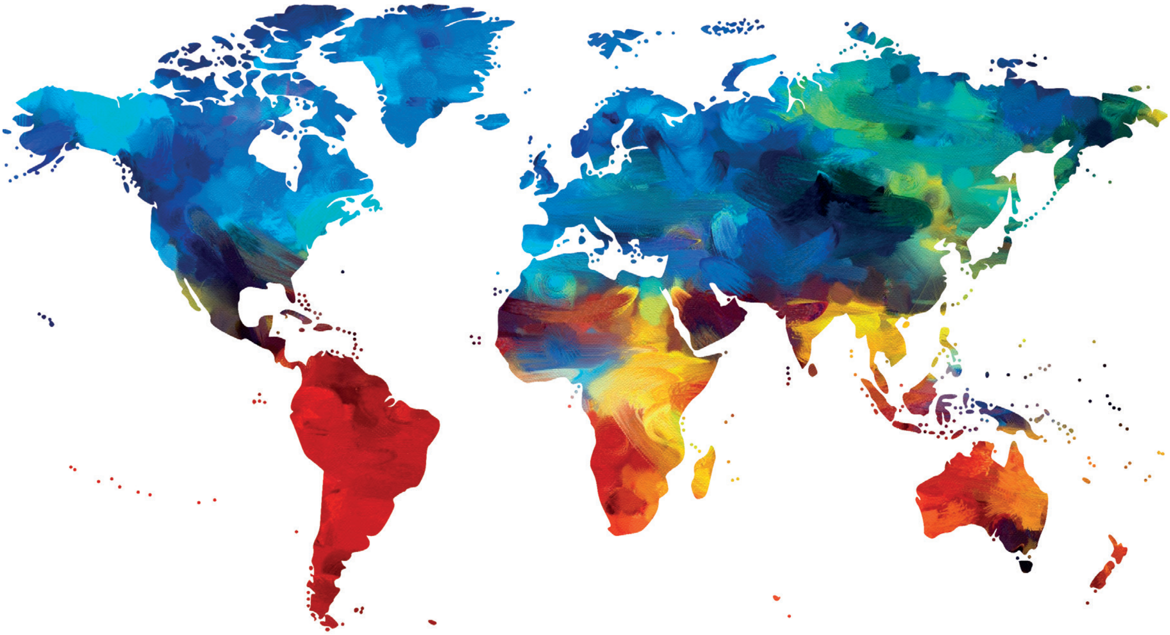
Klaus Gründler

A Contribution to the Empirics of Economic Development

The Role of Technology, Inequality, and the State



Würzburg University Press

Klaus Gründler

A Contribution to the Empirics of Economic Development

Klaus Gründler

A Contribution to the Empirics of Economic Development

The Role of Technology, Inequality, and the State

Würzburg University Press

Dissertation, Julius-Maximilians-Universität Würzburg
Wirtschaftswissenschaftliche Fakultät, 2017
Gutachter:
Prof. Dr. Norbert Berthold (University of Würzburg) und
Prof. Dr. Ansgar Belke (University of Duisburg-Essen)

Impressum

Julius-Maximilians-Universität Würzburg
Würzburg University Press
Universitätsbibliothek Würzburg
Am Hubland
D-97074 Würzburg
www.wup.uni-wuerzburg.de

Print on Demand

Coverdesign: Julia Bauer
Weltkarte: Felix_Broennimann/pixabay.com, CC0 1.0 Universell Public Domain Dedication

ISBN 978-3-95826-072-6 (print)
ISBN 978-3-95826-073-3 (online)
URN urn:nbn:de:bvb:20-opus-141520

Acknowledgments

Die vorliegende Dissertation entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Volkswirtschaftslehre, insbesondere Wirtschaftsordnung und Sozialpolitik an der Julius-Maximilians-Universität Würzburg. Mein erster Dank gilt daher meinem Doktorvater, Herrn Professor Dr. Norbert Berthold, der mir in zahllosen Diskussionen kritische und wertvolle Hinweise gegeben hat. Herr Berthold hat mir stets großen Freiraum für die Forschung gewährt und mich in allen wissenschaftlichen Belangen unterstützt. Bedanken möchte ich mich überdies bei Herrn Professor Dr. Ansgar Belke von der Universität Duisburg-Essen, der so freundlich war, mich als Zweitgutachter zu unterstützen.

Ein besonderes Dankeschön gebührt auch meinen Kollegen Alexander Brunner, Mustafa Çoban, Christian Erthle, Markus Fraaß, Sebastian Köllner, Nadya Kosturkova, Jörg Rieger und Sarah Sauerhammer für die stets sehr gute Zusammenarbeit und die vielen anregenden wissenschaftlichen Dispute. Des Weiteren danke ich Tommy Krieger (Universität Konstanz), Sebastian Köllner, Sarah Sauerhammer und Philipp Scheuermeyer (alle Universität Würzburg) für die stets überaus produktive und inspirierende Zusammenarbeit an verschiedenen Forschungsprojekten. Auch die akribische Korrekturarbeit von Herrn Michael Labate und Frau Margaret Sloop hat wesentlich zur sprachlichen Qualität der Arbeit beigetragen. Für ihren unermüdlichen Einsatz bei organisatorischen Dingen möchte ich mich bei Frau Silke Amelang bedanken, die mich in Lehrstuhlfragen stets unterstützt hat.

Mein größter Dank gilt meiner Familie und meiner Freundin Jennifer, denen ich diese Arbeit widme. Ohne die vorbehaltlose moralische wie finanzielle Unterstützung meiner Familie wäre weder mein Studium noch diese Dissertation möglich gewesen. Ihnen verdanke ich außerdem meine Neugierde und die Freude daran, Dinge zu hinterfragen und verstehen zu wollen. Und hier gibt es noch einiges zu tun. Oder wie es der großartige Richard Feynman ausgedrückt hat:

> *Der Spaß fängt erst dann an, wenn man die Regeln kennt. Im Universum aber sind wir momentan noch dabei, die Spielanleitung zu lesen.*
> (Richard P. Feynman)

Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Why study economic growth?

"Long-run economic growth is the single most important determinant of the economic well-being of a nation's citizens. Everything else that macroeconomists study—unemployment, inflation, trade deficits, and so on—pales in comparison."
(N. Gregory Mankiw)

"I do not see how one can look at figures like these without seeing them representing possibilities. Is there some action a government of India could take that would lead the Indian economy to grow like Indonesia's or Egypt's? If so, what exactly? If not, what is it about the "nature of India" that makes it so? The consequences for human welfare involved in questions like these are simply staggering: once one starts to think about them, it is hard to think about anything else."
(Robert Lucas, Jr.)

Economists are fascinated by the study of the factors underlying economic development. Why are some nations rich, and others poor? Why are growth rates of some countries extraordinarily high, while progress in other economies still lags behind? And—most important from a politico-economic point of view—are there policy actions that can help to influence the development process? The importance of these questions becomes apparent when we look at empirical data on the distribution of incomes in world. These figures reveal a substantial degree of inequality across countries. According to the data collected from the World Bank (2017), the list of the richest countries in terms of average real per capita GDP is headed by Qatar, Luxembourg, Singapore, Kuwait, and Brunei. Incomes in these countries span from 134,182 USD in Qatar to 67,912 USD in Brunei. Close behind this group are the advanced economies of Norway (64,004 USD), Switzerland (55,260 USD), the United States (52,117 USD), and—at some distance—Germany (43,602 USD). At the other end of the spectrum, living standards lag substantially behind. The group of the poorest countries in the world primarily consists of Sub-Saharan countries, including the Central African Republic (567 USD), the Democratic Republic of the Congo (712 USD), Burundi (734 USD), Malawi (784 USD), and Liberia (802 USD). To break the differences in wealthiness across countries down to one number, Lakner and

Milanovic (2015) conclude that the global Gini coefficient of income inequality was roughly 70 percent in 2008, and has only slightly declined since the early 1980s. Or, put differently, average household incomes in Qatar in 2016 were about 237 times as much as those in the Central African Republic.

The striking disparity in wealthiness across countries has its origin in substantial differences in economic development during the past centuries. The drivers of these development processes are deeply rooted in history. But even today, we observe strong cross-country disparities in economic growth: Whereas per capita incomes on average grew by 2.87% in the period between 2010 and 2014, living standards in some countries increased considerably faster, while the income level of other nations revealed a tendency to stagnate or even decline. The most substantial increases in per capita GDP took place in Turkmenistan (10.56 percent), Mongolia (10.34), China (9.23), Ethiopia (9.50), Argentina (8.07), and Singapore (7.21). Compared with these numbers, growth rates in the United States (1.21 percent), Azerbaijan (1.24), France (1.11), Mali (1.14), the Netherlands (0.98), and the United Kingdom (0.66) were considerably lower. Even more striking, the data implies that some countries experienced a *decrease* in the standard of living in the post-2010 period, most notably Yemen (-4.30 percent), Côte d'Ivoire (-2.85), Iceland (-2.01), and Syria (-1.03).

At first glance, there seems to be no distinct pattern which explains why some countries grow fast, while the pace of income increases in other nations is slower. The study of economic growth therefore requires systematic research to identify the determinants of long-run development. Given the huge disparity in living standards across households on the globe, it is comprehensible that most economists agree with Robert Lucas and Gregory Mankiw about the importance of the analysis of long-run growth. Mankiw further emphasizes that *"economists know quite a lot about the forces that govern economic growth"* (see Mankiw, 2009). In fact, over the last 250 years—and in particular, during the last six decades—the growth literature brought forward a number of theoretical explanations about the origins of economic well-being and growth. This literature emphasizes the role of saving, population growth, technological progress, and human capital in determining the level and growth of a nation's living standards. In more recent years, the growth literature has moved from the study of proximate factors to determinants that are deeply rooted in the history and geography of countries. Understanding these determinants is key to increase the wealth of individuals and to lessen poverty in the world. While the exploration of the factors of growth long concentrated on the theoretical mechanisms, growth research during the 1990s started to pay close attention to empirical regularities and to the relationship between data and theory. This applied perspective initially focused on empirical investigation of the neoclassical theory and was later expanded to application of the more recent endogenous growth models during the early 2000s. The empirical evidence contributed considerably to the understanding of historical growth patterns and led to a rich set of policy implications that helped enhance the standard of living in both advanced and developing countries.

However, while the last two decades saw tremendous progress in poverty reduction in large parts of Asia and Latin America, living standards in Sub-Saharan African countries still lag alarmingly behind. A similarly disconcerting observation is that growth rates in the advanced OECD countries declined significantly since the turn of the millennium. These phenomena underscore that there is still a lot of uncharted territory in the area of economic growth research that requires forceful empirical exploration. Moreover, recent studies in the field have revealed a striking number of puzzling relationships that are still poorly understood and analyzed. For instance, there is a widespread belief among citizens around the world that democracy contributes to an increase in wealthiness (see World Value Survey, 2014). However, the empirical literature investigating the effect of democracy on growth could not be more divided. Another example involves the impact of the financial sector. While the growth models of the 1990s and the early 2000s stress the importance of financial intermediaries to realize long-run increases in incomes, recent empirical surveys suggest that both variables are negatively associated. Finally, the substantial increase in inequality of both income and wealth observable in many parts of the world has encouraged a large number of researchers to examine the effect of income disparities on economic well-being. Although the literature on this topic is growing rapidly, the results are still far from being consistent. Even more importantly, there is very little evidence for the growth-effects of policy actions that seek to equalize household incomes.

This book provides an attempt to explain some of the new questions arising from both the observable stylized facts of the last decade and the implications of recent research projects. The focus of the analyses is on the empirical examination of these research questions. However, to place the empirical studies on a solid foundation, it is worthwhile to briefly overview the growth theory that has been developed during the past 250 years.

1.2 A brief history of growth theory

> *"Always remember: Your focus determines your reality."*
> (Qui-Gon Jinn)

Every good empirical analysis requires a sound theoretical framework which provides scholars with hypotheses that can be tested with data. In many respects, it is the theory that should determine the empiricists' focus of inquiry, rather than the allure of unprecedented opportunities to model empirical relationships.[1] To put the focus in the right direction, this section gives a brief overview of the history of growth theory and its important predictions. While the later chapters of this book provide a much more detailed explanation of the particular mechanisms implied by the theoretical work, this section features a short summary of the basic concepts and an illustration of how the different strands of theory are linked historically. It does so with a particular emphasis on the neoclassical theory, which will not be discussed explicitly in the later chapters, but whose general ideas still lie at the heart of the overwhelming majority of growth models to date.

[1] A famous example in the field of empirical growth research that at first glance entirely contradicts this assessment is the article of Sala-i Martín (1997), which he named—somewhat playfully—*"I Just Ran 2 Million Regressions"*.

Classical growth theory

In the late 18th century, Smith (1776) conducted the first conclusive investigation on the sources of national wealth. Today, his famous book *"An Inquiry into the Nature and Causes of the Wealth of Nations"* is widely recognized as the birthplace of growth theory. While Smith (1776) did not propose a theory on long-run development as such, he introduced many concepts that are now part of the fundamental toolbox of modern growth models. Among the most influential of these notions are the emphasis on the importance of labor productivity and savings as well as the theory concerning the "stationary state", the latter being a situation where capital accumulation and population growth have reached their boundaries, with the result that the economy may not progress any further. Moreover, Smith (1776) underscores the importance of technological progress for long-run growth, as well as the detrimental effects emanating from high rates of population growth.

Both concepts also constitute the main body of the growth theory of David Ricardos *"On the Principles of Political Economy and Taxation"*. In a two-sector model consisting of a manufacturing and an agricultural sector, Ricardo (1817) illustrated how capital owners—in Ricardos theory referred to as the "productive class" of the society—devote their profits to further capital accumulation. Ricardos theory went on to emphasize that population growth prevents this process from being sustainable in the long-run, with the result that the economy converges into the stationary state. Technological progress—although in the end not sufficient to prevent the situation of a stationary state from eventually occurring—can postpone this development.

A considerable time after Ricardo published his ideas, Schumpeter (1911, 1942) introduced his theory on long-run development, emphasizing the role of innovations and technical change. In short, the entrepreneur, widely acknowledged as a pioneer in the field of new technologies, innovates via a process of *"creative destruction"*. This innovation process eventually triggers long-run increases in living standards.

As one of the first theorists, Harrod (1939) focused on the *rate* of growth, laying the foundation for the necessary conditions for long-run equilibrium growth models. Likewise, Domar (1946) stressed the importance of a dynamic analysis for understanding long-run growth patterns. Being developed independently, the main growth processes formulated in Harrod (1939) and Domar (1946) display some considerable similarities, which is why the underlying theory is often referred to as the *Harrod-Domar model*. The model builds on the Keynesian short-term analysis to examine the long-run development of the economy, a strategy which is accompanied with a number of conceptual problems that Barro and Sala-i-Martín (2004) refer to as "implausible assumptions" yielding "undesirable outcomes". Today, the Harrod-Domar model is appreciated as the link between classical and neoclassical theory.

Neoclassical growth theory

The main problem of the Harrod-Domar model is the tendency toward instability, which is a particularly unfavorable property of any approach whose aim is to describe long-run patterns. Solow (1956) criticizes the equilibrium of the model as "balancing on a knife's edge".[2] Using neoclassical production functions that allow for varying shares of labor and capital inputs, the models of Solow (1956) and Swan (1956) form the foundation of neoclassical growth theory and mark the starting point for most studies of economic growth to date. Similar to the *Harrod-Domar* case, Robert Solow and Trevor Swan developed their models independently, but due to the large number of shared features, the model is now known as the *Solow-Swan* model of growth.

Consider an economy that produces one commodity, output as a whole. The quantity of this commodity reflects the community's real income, with the rate of its production being denoted by $Y(t)$. In each instant, part of the total production is saved and invested, while the other part is consumed. Individuals in the economy save a constant fraction s of their incomes, which yields a saving rate $sY(t)$. The capital stock of the economy takes the form of an accumulation of the composite commodity and is denoted with $K(t)$. It follows that an increase in $K(t)$ is the result of net investment, that is $\dot{K} \equiv dK/dt$. This yields a basic identity at every instant of time given as

$$\dot{K} = sY. \tag{1.1}$$

The output is produced with the help of two input factors. These are physical capital and labor, with the latter's rate of input denoted by $L(t)$. The potential output implied by the input factors depends on the technological possibilities, which are reflected by a neoclassical production function

$$Y = F(K, L), \tag{1.2}$$

which is homogeneous of first degree. In this assumption, the model deviates from the Ricardian growth theory that incorporates scarce nonaugmentable resources like land, which would imply decreasing returns to scale. Combining Equations (1.1) and (1.2) yields an equation with two unknowns

$$\dot{K} = sF(K, L). \tag{1.3}$$

To close the system, the model of Solow (1956) makes use of some ideas initially proposed by Harrod (1939). Due to exogenous population growth, the labor force increases at a constant rate n. In the simple version of the model, there is no technological change, which implies that n is Harrod's neutral growth rate. Hence, it follows that

$$L(t) = L_0 e^{nt}, \tag{1.4}$$

2 In the introduction of his 1956 seminal paper, Robert Solow writes: *"All theory depends on assumptions which are not quite true. That is what makes it theory. The art of successful theorizing is to make the inevitable simplifying assumptions in such a way that the final results are not very sensitive. A 'crucial' assumption is one on which the conclusions do depend sensitively, and it is important that crucial assumptions be reasonably realistic. When the results of a theory seem to flow specifically from a special crucial assumption, then if the assumption is dubious, the results are suspect. I wish to argue that something like this is true of the Harrod-Domar model of economic growth."*

which implies perpetually maintained full employment in the economy.[3] Inserting Equation (1.4) in Equation (1.3) gives

$$\dot{K} = sF(K, L_0 e^{nt}). \tag{1.5}$$

This equation determines the path of capital accumulation that must be followed, given that all labor is employed. Equation (1.5) is a differential equation in the single variable $K(t)$. Computation of a solution for Equation (1.5) gives the only time profile of $K(t)$ possible, subject to the side condition of full employment. If that time path is known, the corresponding time path of the real output can be calculated with the help of the modeled production function. Without knowing the exact functional form of the production function, however, it is naturally impossible to obtain an exact solution. To counter this disadvantage, Solow (1956) derives a number of broad properties, which he isolates in a surprisingly simple manner.

To illustrate his procedure, introduce the ratio of capital to labor, denoted with $\rho = K/L$. It follows that

$$K = \rho L = \rho L_0 e^{nt}. \tag{1.6}$$

Differentiating Equation (1.6) with respect to time gives

$$\dot{K} = L_0 e^{nt} \dot{\rho} + n\rho L_0 e^{nt},$$

which implies by substitution in Equation (1.5) that

$$(\dot{\rho} + n\rho) L_0 e^{nt} = sF(K, L_0 e^{nt}).$$

The renunciation of the Ricardian scarce-land case proves to be of great advantage in this part of the model, as homogeneity of degree one allows for division of both variables in $F(\cdot)$ by $L_0 e^{nt}$, provided that $F(\cdot)$ is multiplied by the same factor. Thus, it follows that

$$(\dot{\rho} + n\rho) L_0 e^{nt} = sL_0 e^{nt} F\left(\frac{K}{L_0 e^{nt}}, 1\right).$$

The last step to obtain a differential equation that is only composed of the capital-labor ratio requires dividing out the common factor, i.e.

$$\dot{\rho} = sF(\rho, 1) - n\rho. \tag{1.7}$$

This fundamental equation is easy to interpret. It provides computation of the output per worker as a function of capital per worker, stating that the rate of change of the capital-labor ratio is simply the difference of the increase in capital and the increment of labor.

What are the implications for long-run growth provided by this model? When $\dot{\rho} = 0$, the capital-labor ratio is constant, while the capital stock expands at rate

[3] The reason is that Equation (1.3) models L to reflect total employment, whereas it denotes the available supply of labor in Equation (1.4).

n. This case yields a stable equilibrium rate ρ^* that is referred to as the "steady state'. Once this condition is established, it will be maintained, implying—by constant returns to scale—that output per head of the labor force will be constant. An important conclusion drawn by this growth mechanism is convergence. If the initial capital stock is below its equilibrium ratio, the economy will grow at faster pace until ρ^* is reached. This, on the other hand, implies that less developed economies must grow at a faster rate than advanced countries, with the results that poor economies catch-up with the industrial nations.

Although Solow (1956) also presents cases in which economies possess multiple steady states, the neoclassical growth model implies that the achievement of long-run growth based on capital accumulation alone is an impossible task. For this reason, the model also provides the extension of a neutral technological change. Consider a form of technological progress that simply enhances the production function in Equation (1.2) by an increasing scale factor $\Psi(t)$. This yields

$$Y = \Psi(t)F(K, L), \tag{1.8}$$

suggesting that the isoquant maps are unchanged, but their attached output number is multiplied by $\Psi(t)$. In this case, the capital-labor ratio in equilibrium is "blown-up", resulting in an ever-changing number of ρ^*. This mechanism can be illustrated very simply in the case of the Cobb-Douglas function.

For $\Psi(t) = e^{gt}$, the basic differential equation becomes

$$\dot{K} = se^{gt}K^{\alpha}(L_0 e^{nt})^{1-\alpha} = sK^{\alpha}L_0^{1-\alpha}e^{(n(1-\alpha)+g)t},$$

with the solution

$$K(t) = \left\{K_0^{\xi} - \frac{\xi s}{n\xi + g}L_0^{\xi} + \frac{\xi s}{n\xi + g}L_0^{\xi}e^{(n\xi+g)t}\right\}^{\frac{1}{\xi}}, \tag{1.9}$$

where $\xi = 1 - \alpha$ is introduced to keep the equation well-arranged. In the long-run, the capital stock now increases at the relative rate $n + g/\xi$, resulting in a real output growth rate of $n + \alpha g/\xi$. In per capita terms, the introduction of technological progress implies that the equilibrium is no longer stable, but eventually increases at the rate g/ξ. In the case of a high capital labor ratio, it may fall initially, but will eventually behave asymptotically as described.

In many respects, the model of Solow (1956) and Swan (1956) led to a revolution in growth theory, providing for the first time a conclusive framework that enables the explanation of the origins of long-run development. However, as the model left a number of questions unresolved, it was not surprising that the new framework inspired a number of theorists to propose improvements, among which the most notable attempts were made by Koopmans (1965) and Cass (1965). This theory has been consolidated into what is now known as the *"Ramsey-Cass-Koopmans"* model. The main difference from the Solow-Swan model is that the choice of consumption is explicitly microfounded, allowing for endogenization of the savings rate.

One of the most controversial implications of the neoclassical growth model is the prediction that individuals have a strong incentive to invest in economies where the rate of return is highest, suggesting strong capital flows from advanced economies to the developing world. Whereas these flows indeed exist and have increased significantly since the early 1990s, the vast majority of investment still takes place in the origin country (see Sardadvar, 2011). The critical question of why the neoclassical growth model fails to accurately predict the quantity of capital flows has become known as the *"Lucas paradox"*. Lucas (1990) provides two explanations for the occurrence of this paradox, one dealing with capital market imperfection, and the other introducing the concept of different levels of human capital across advanced and developed economies. The latter is a crucial part of endogenous growth theories which arouse around the time when Lucas first brought forward his conceptualization of the capital flow puzzle. In attempt to augment the standard neoclassical model, Mankiw et al. (1992) introduce human capital accumulation into the neoclassical model of Solow (1956). While the resulting long-run growth rate resembles the quantitative conclusions of the Solow model, the incorporation of human capital has a substantial impact on the analysis. The main deviation from the standard neoclassical growth model made by the Mankiw et al. (1992) framework is that it also allows for "saving" in human capital. The model then illustrates how saving either in physical capital, human capital, or both results in shift-effects in long-run output levels.

Endogenous growth models

Although Mankiw et al. (1992) and others succeeded in solving some of the drawbacks of the Solow-Swan model, a crucial disadvantage persisted: the exogeneity assumption of technological progress. Once reaching its equilibrium, the per capita growth rate of the economy in the neoclassical framework is entirely determined by technological change. The Solow-Swan model assumes that the rate of technological progress is primarily influenced by a scientific process that is independent of, and separate from, economic forces. Therein lies a serious pitfall from a politico-economic perspective, as the neoclassical theory implicitly assumes the long-run growth rate to be determined from outside the economic system, with little impact of political action. In the mid-1980s, growth theorists became increasingly dissatisfied with this prediction, desiring to explain long-run patterns as being influenced by economic factors, particularly—but not exclusively—by describing endogenous incentives and opportunities to create new (technological) knowledge. In light of this goal, this strand of theory is called *"endogenous growth theory"*. What began with a small group of theorists in the 1980s has now evolved into a field covering numerous theoretical models and explanations. Summarizing all of the influential models is therefore an impossible task. Hence, the following brief overview concentrates on some of the most trailblazing theories, which provide the groundwork of the models employed in the chapters of this book.

The first version of an endogenous growth theory was the AK model. An early variant of this model was introduced by Frankel (1962), while Romer (1986) provided a similar examination with a more general production structure. The basic idea is to dissolve the distinction between capital accumulation and technological progress, consolidating physical capital, human capital and "intellectual capital"—the latter being accumulated whenever innovations occur—into a single concept of capital. The cornerstone of this model, initially introduced by Frankel (1962), is that the aggregate production function resulting from such a broader definition of capital is not necessarily shaped by decreasing marginal returns. The reason is that parts of the aggregated capital stock contain intellectual elements reflecting technical progress, which offsets the tendency of the marginal product of capital to diminish. Whereas the model theoretically also enables increasing marginal returns of capital, the special case of constant marginal products is examined in the AK framework. In this case, aggregate output is proportional to the capital stock, i.e.

$$Y(t) = \Psi K(t). \tag{1.10}$$

As the positive constant Ψ is often denoted by $A(t)$, it is from here that the model receives its name.[4] Similar to the Solow-Swan model, capital accumulation depends on the economy's saving rate. If a fixed fraction of output is saved and a fixed fraction of depreciation occurs, then the rate of aggregate net investment is

$$\frac{dK}{dt} = sY - \delta K. \tag{1.11}$$

Equations (1.10) and (1.11) together imply that the growth rate γ of the economy is given by

$$\gamma \equiv \frac{1}{Y}\frac{dY}{dt} = \frac{1}{K}\frac{dK}{dt} = s\Psi - \delta. \tag{1.12}$$

Unlike in the neoclassical growth model, Equation (1.12) suggests that an increase in the saving rate results in a permanent increase in the growth rate. In the augmented Ramsey version of the model developed by Romer (1986), saving is generated by intertemporal utility maximization rather than by the assumption of a fixed rate.

A similar analysis is conducted by Lucas (1988), who extends the AK model to capture the effect of human capital accumulation. The model builds on the assumption of Uzawa (1965), who introduced the idea that human capital and technical knowledge are identical. In a manner similar to the AK model, the Uzawa-Lucas framework shows how human capital accumulation contributes to long-run growth.

The AK theory was followed by a number of endogenous growth models which reject the idea that innovation activity and human as well as physical capital are identical. The motivation for this new wave of theory was the inability of the AK model to provide a convincing explanation for convergence, and the oversimplification in lumping together various different types of capital. In contrast, the "innovation-based" growth theory emphasizes that the stock of physical and human capital accumulates through saving and education, while growth in intellectual capital has its origin in innovation activity. Innovation-based growth theories can be distinguished into two

[4] In order to guarantee comparability with the models in the subsequent chapters, however, the term Ψ is used rather than A.

parallel branches. The first branch, originated by Romer (1990), builds on product varieties, while the second branch grew out of modern industrial organization theory and is concerned with quality-improving innovations that render the vintage variants obsolete (see, for instance, the models of Grossman and Helpman, 1991 and Aghion and Howitt, 1992). As these models are built on forces that Schumpeter (1942) called "creative destruction", this branch of innovation-based growth models is often referred to as "Schumpeterian growth theory".

The innovation-based growth mechanism modeled by Romer (1990) specifies the capital stock as consisting of a number of intermediate goods. The framework illustrates how the invention of new products causes productivity growth. The basis of this analysis is the production function developed by Dixit and Stiglitz (1977) and Ethier (1982), in which production of the final output is achieved via two input factors, namely labor and a continuum of intermediate capital variants. This function is given by

$$Y = L^{1-\alpha} \int_0^N x_i^{\alpha} di, \; 0 < \alpha < 1, \tag{1.13}$$

where x_i is the flow input of intermediate product i, and N is the number of different intermediate goods available in the economy. Due to $0 < \alpha < 1$, each i features decreasing marginal returns, which eventually leads to a convergence effect of the economy towards its steady state. However, the creation of new i via innovation activity opens up a source of long-run growth, as for each element of the capital stock, the convergence process is initiated separately. As long as new capital variants can be invented, the per capita growth rate of the economy is greater than zero.

In the Schumpeterian variant of innovation-based models, aggregate output is again produced by a continuum of intermediate goods. However, this branch further captures the effect of quality-improvements, which is accomplished as follows (see Aghion and Howitt, 2009 for a detailed discussion)

$$Y = L^{1-\alpha} \int_0^1 \Psi_i^{1-\alpha} x_i^{\alpha} di, \; 0 < \alpha < 1. \tag{1.14}$$

This case introduces a fixed measure of product variety, normalized to unity. In addition, the function features a separate productivity parameter Ψ_i for each element i. Assume that each sector i is shaped by a monopolist that produces its product with a constant marginal cost of unity and faces a demand curve that is implied by the marginal product in the final sector, i.e. $\alpha \left\{\Psi_i L / x_i\right\}^{1-\alpha}$. To arrive at the monopolist's profit-maximizing intermediate output, it suffices to equate marginal revenue—that is, the marginal product multiplied by α—to the marginal cost

$$x_i = \zeta L \Psi_i,$$

where $\zeta = \alpha^{2/(1-\alpha)}$. Substituting this identity for each x_i in Equation (1.14) gives the aggregate production function

$$Y = \lambda \Psi L, \tag{1.15}$$

where $\Psi \equiv \int_0^1 \Psi_i di$ is the average productivity parameter and $\lambda = \zeta^{\alpha}$. The key element of this version of innovation-based growth theory is that innovation yields improved versions of old products. Such an improvement in i features a productivity parameter Ψ_i that exceeds that of its precursor by a fixed factor $\eta > 1$. If the probability of occurrence of an innovation in i over any interval of length di is $\mu \times dt$, then the growth rate of Ψ_i is

$$\frac{d\Psi_i}{\Psi_i} \times \frac{1}{dt} = \begin{cases} (\eta - 1) \times \frac{1}{dt} & \text{with probability } \mu \times dt \\ 0 & \text{with probability } 1 - \mu \times dt, \end{cases}$$

which implies that the expected growth rate of Ψ_i is $\mathrm{E}(\gamma) = \mu(\eta - 1)$. The probability μ in these models is proportional to the productivity-adjusted R&D expenditures, i.e.

$$\mu = \theta R/\Psi. \tag{1.16}$$

In Equation (1.16), R is the amount of output spent on research activity, and the division by Ψ highlights the influence of an increase in complexity of new knowledge as technology advances. Equation (1.15) shows that the growth rate of the economy follows the growth rate of average productivity. If a law of large numbers can be applied, then γ equals the expected growth rate $\mathrm{E}(\gamma) = \mu(\eta - 1)$. From this and Equation (1.16), it is easy to obtain

$$\gamma = (\eta - 1)\theta R/\Psi.$$

In accordance with Equation (1.15), it follows that the growth rate is contingent upon the fraction of GDP spent on research activity. Let $z = R/Y$, then the growth rate of the economy is

$$\gamma = (\eta - 1)\theta\eta L z. \tag{1.17}$$

This equation yields different implications for economic policy in comparison to the neoclassical growth theory. As Equation (1.17) highlights, it is not aggregate saving that leads to the most rapidly increases in growth, but rather the expenditure on research activity. More advanced innovation-based growth models also focus on the description how research activity is influenced by various policies and how technological advancements affect the income distribution.

Unified Growth Theory

More recently, a separate strand of growth theories has been developed, which has become known as the *"Unified Growth Theory"* (UGT). Pioneered by Galor and Weil (1999, 2000), Galor and Moav (2002, 2004), and Galor (2011), the Unified Growth Theory seeks to provide a fundamental framework to analyze the evolution of economies over the entire course of human history. This includes the epoch of Malthusian stagnation that characterizes most of human history, the contemporary era of sustained growth, and the driving forces that led from stagnation to long-run development, including—in particular—the democratic and the demographic transition.

With this goal in mind, the critique exercised against the neoclassical model and the endogenous theories is not the inability of these theories to describe recent growth patterns, but rather the counterintuitive results that occur when trying to link these models to the Malthusian epoch. In addition, neither the neoclassical nor the endogenous framework is able to describe the great divergence in incomes across the world that occurred during the industrial revolution. The UGT aims at providing a compelling analysis of these phenomena, but this strand of the literature is still in its infancy, with only little empirical evidence at hand.

1.3 A short journey through this book

Based on the fundamental ideas outlined in the previous section, the following chapters study the determinants of economic development in greater detail. The main focus will be on the "modern growth regime", the period of sustained growth which was initiated by the Industrial Revolution during the second half of the 18th century and that brought about an increase in wealthiness unparalleled in the entire human history. The emphasis of this book is placed on the empirical investigation of the origins of long-run growth, and will be enriched by contributions to the theory of economic growth and methodological advances. With respect to theory, Chapters (2) and (3) develop an endogenous growth model, which includes culture, entrepreneurship, and the financial sector into the innovation-based framework presented in Section 1.2 of this chapter. On the methodological side, Chapter (6) advances mathematical algorithms of machine learning and pattern recognition for practical applications and illustrates how these techniques can be used for economic purposes. In this book, machine learning and techniques from the field of artificial intelligence are employed to derive a new technique for computation of composite measures. This technique resolves existing pitfalls in the construction of indices and classifications, particularly with regard to the aggregation function of the underlying attributes. Machine-learning-based measurements substantially enhance our ability to understand empirical patterns, and consequently help in both the study of empirical growth patterns and the derivation of better targeted economic policy measures.

This book is divided into seven chapters. These chapters aim to contribute to a better understanding on why disparities in living standards across the world are substantial and persistent. A wealth of theoretical and empirical studies have investigated the growth processes occurring during previous decades and centuries that have caused these cross-national differences in incomes and wealthiness, stressing that among the diverse factors that influence long-run economic well-being, a particularly prominent role is played by technological progress (see the "brief history of growth theory" in Section 1.2 of this chapter). More recently, the empirical growth literature has moved from the analysis of proximate factors to the study of more fundamental determinants rooted in long-term history, including geography, institutions, and the population (Spolaore and Wacziarg, 2013). Chapter (2) links these two branches of the literature and raises a very fundamental question, asking whether there are deep-rooted factors that distinguish countries in terms of their tendency and ability to spur innovation.

In fact, a striking empirical observation is that the diffusion of technology progresses differently across countries (Caselli and Coleman, 2001), and that these disparities are substantial even among countries with similar levels of income and education (Vandenbussche et al., 2006). This suggests that there are other, non-income-related factors which motivate individuals to pursue innovation activities and which, consequently, lead to cross-national differences in technological progress. Chapter (2) argues that cultural socialization accounts for large parts of these non-monetary factors, thereby adding to the burgeoning literature that foregrounds the deep roots of economic development. Although a variety of authors have underscored the fundamental role of culture for socio-economic transformations (Weber, 1905) and economic development (Landes, 1998; Guiso et al., 2006), there have been relatively few attempts to systematically study the empirical growth effect of cultural legacy. While culture consists of a rich palette of different facets, Chapter (2) argues that the propensity of the population to engage in entrepreneurial activity is a key element through which culture is transmitted to growth.

The importance of entrepreneurship for economic growth has been stressed in early essays of the economic profession (see, e.g., Schumpeter, 1934, 1942 and the *"brief history of growth theory"* in the previous chapter) and also in more recent growth models (Mokyr, 2016; Doepke and Zilibotti, 2014). While these models broadly suggest a positive relationship between cultural attitudes towards entrepreneurship and the growth rate, documentation of entrepreneurship's empirical relevance has progressed much more slowly (Doepke and Zilibotti, 2014; Glaeser et al., 2015), and the results are thus far rather ambiguous, particularly in the cross-country context. Chapter (2) demonstrates that the inconclusiveness of recent studies has its roots in three methodological pitfalls, namely i) insufficient consideration of endogeneity and reverse causation, ii) sample selection biases, and iii) inadequate identification of entrepreneurial activity.

To tackle the first of these issues, Chapter (2) uses a number of instrumental variables to identify a causal effect running from culture to growth. These instruments include genetic distances measured by the frequency of blood-types (Gorodnichenko and Roland, 2017), a broad set of genetic polymorphisms (Spolaore and Wacziarg, 2009), two specific genetic characteristics (Chiao and Blizinsky, 2010 and Way and Lieberman, 2010), historical prevalence of pathogens (Murray and Schaller, 2010), and linguistic differences (Licht, 2007). The rationale for this strategy is that both genes and culture are transmitted simultaneously from parents to their offspring, hence there is a strong correlation between both variables (in biological studies, this phenomenon is referred to as "co-evolution", see Chiao and Blizinsky, 2010).

To account for potential sample selection biases, the analysis in Chapter (2) exploits information from a large panel of countries and from historical data covering the time span from 1500 to present. Finally and most importantly, the chapter shows that the greatest source of ambiguity in recent studies is the inadequate measurement of entrepreneurship. While empirical studies routinely measure entrepreneurship using the self-employment rate, a striking empirical pattern is that self-employment is much more prevalent in the developing world. As a consequence, the parameter estimates in models that assume a linear relationship between self-employment (or similar proxies of entrepreneurship) and economic growth may be biased. To overcome this bias, the analysis estimates the effect of the self-employment rate conditional on the initial development level. The results provide strong evidence for a significantly positive effect of entrepreneurship and its associated cultural traits on economic growth, which mainly works via facilitation of technological progress.

The analysis in Chapter (2) demonstrates how investment projects carried out by entrepreneurs stimulate long-run development. However, the wealth of entrepreneurs may often not be sufficient to bear the cost of the projects themselves. For this reason, the financial sector may exert positive growth stimuli as it enables funding of projects, thereby facilitating investment and innovation, which eventually translates to an increase in the growth rate. Additionally, the growth literature emphasizes potential growth effects of financial development (FD) on education and health. To study these effects, Chapter (3) is concerned with the empirical analysis of FD's role in the growth process. Investigation of this role is also motivated by the growing skepticism about the contribution of the financial sector to the development of the real economy and the "Too Big To Fail" doctrine, which were renewed by the Financial Crisis of 2007–2008 and supported by the more recent literature (see, e.g., Arcand et al., 2015; Schularick and Taylor, 2012).

The analysis starts with the introduction of the financial sector into the model framework of entrepreneurship and growth introduced in Chapter (2). The model implies a number of potentially positive growth effects, but also points to circumstances in which FD may be an impediment to growth. The latter is in accordance with some of the more recent studies conducted to explore the finance-growth nexus, which find a negative effect of the financial system on income increases. This paradox result is referred to as the *"vanishing effect of finance"* (see, e.g., Arcand et al., 2015). At first glance, this effect seems to be puzzling, and the underlying reasons for why such a negative effect may occur are still poorly understood.

Chapter (3) provides an attempt to empirically explain the negative influence of financial institutions on economic well-being. The key argument advanced in this chapter is that an understanding of the growth effect of FD requires examination of the transmission mechanisms through which it is channeled to growth. While almost all cross-sectional or panel data studies focus on a direct link running from financial development to growth, very few investigations take into account the channels by which financial deepening is translated into income increases (Levine, 2005). These examinations typically model a linear relationship between FD and potential transmission channels, reaching a tentative consensus that the financial system fostered fixed investment and productivity in the post-1960s period (King and Levine, 1993a; Beck et al., 2000; Benhabib and Spiegel, 2000) and over the past 140 years (Madsen and Ang, 2016). However, what is common to all of these studies is that they neglect potential conditionalities in the transmission mechanism. Such conditionalities directly follow from the Unified Growth Theory outlined in Chapter

(1.2), which stresses that the primary engine of economic growth changes throughout the development process. In brief, this theory suggests that investment in physical capital and education are prime sources of growth at earlier stages of development, but their key role is increasingly superseded by that of factor productivity gains as economies become wealthier and credit constraints become less binding. As a result, we may expect changes in the way financial development transmits to growth in the course of the development process and—given the substantial increase in incomes during the past decades—over time.

This chapter employs two strategies to identify such changes, utilizing country-year observations drawn from a panel of 129 countries over the past 50 years as well as those from a historical dataset featuring observations from 21 OECD countries between 1870 and 2009. The results from both datasets provide strong support for the conditionality hypothesis, suggesting that financial development was conducive to growth in earlier periods and that this effect has largely disappeared over time and space. The explanation for this "vanishing effect of finance" is that FD contributes to an increase in physical capital investment and education, and to a reduction in the fertility rate in poorer economies, but these stimulating effects disappear during the development process. In advanced economies, the financial sector primarily fosters productivity gains, leading to a dependency of the financial sector's growth effect on new ideas and potentials for innovation projects.

In fact, productivity gains have declined considerably since the turn of the millennium in the overwhelming majority of advanced economies, resulting in disproportionately low growth rates compared to the historical development of incomes during the last 70 years. This observation led to a renewed interest in the theory of *"secular stagnation"*, a phenomenon originally described in the late 1930s by Hansen (1939). As the concept of secular stagnation may be subject to a close entanglement with the negative effect of finance on growth highlighted in Chapter (3), the analysis in Chapter (4) is concerned with an examination intended to unravel the causes of the reduction in growth observed in the advanced economies. It does so with a particular focus on the German economy. Like most Central-European countries, Germany has realized tremendous growth rates in the aftermath of World War II. Since the early 1970s, however, growth rates have declined and had settled down at a more or less constant rate of 2 percent per year, only to experience a renewed negative trend around the early 2000s. Notably, the decline in growth initiated around the turn of the millennium can be observed in each of the advanced economies. The analysis shows that this decrease is first and foremost the result of global trends that affect most developed countries equally, and is accompanied by country-specific circumstances. The chapter provides evidence for a supply-side explanation of secular stagnation due to a lack of radically new ideas.

Subsequently, Chapter (5) introduces income disparities among households into the empirical analysis. Interpreted in the context of the growth mechanisms discussed in Chapters (2)–(4), inequality of incomes implies that some individuals have the ability to realize their investment projects, while insufficient wealth of others prevents some individuals from exploiting their full intellectual and entrepreneurial potential. The result is a decline in education and innovation activity, which translates to a lower rate of economic growth. Chapter (3) demonstrates that under some circumstances, functioning capital markets can help to mitigate this problem. However, particularly in the case of developing economies, capital markets are often imperfect (see Galor and Zeira, 1993 and Galor and Moav, 2004). In such cases, high levels of inequality are likely to impede economic development.

To prevent the negative impact of inequality on economic growth in the presence of capital market imperfections, public redistribution may seem to be an adequate policy measure to equalize the investment opportunities among individuals. This, however, only holds if the effect of redistribution itself is neutral in the growth process. Chapter (5) analyzes the effect of income inequality and its direct remedy—public redistribution—empirically on the basis of a broad international panel of countries. Research in this field has been strongly limited due to a restriction of reliable and comparable measures of redistribution. Fortunately, the recent advancements in data availability made by the SWIID version 6.1 of Solt (2016) allows for an analysis that is based on a substantially higher number of country-years.

The analysis reveals a robust negative effect of inequality on economic growth, which is traced back to the transmission channels through which income disparities exert their influence on income increases. By showing that less equal societies tend to have a less educated population and higher fertility rates—in particular when credit availability is low—the chapter supports the credit market imperfections mechanism and the endogenous fertility channel. There is also a correlation between inequality and physical capital investment, but this relationship is less pronounced. In line with the political economy mechanism of the endogenous fiscal policy channel, Chapter (5) also shows that a higher level of market inequality predicts more public redistribution. Moreover, redistribution by taxes and transfers turns out to directly harm economic growth when net inequality is held constant. This effect is due to an impairment to physical capital investment and an increase in the fertility rate triggered by high redistributional efforts by the government.

When estimating the aggregate growth effect of redistribution—that is, its *direct* negative effect combined with its *indirect* positive effect resulting from lower net inequality—the results suggest that both effects are offsetting. Thus, at a given level of market inequality, redistribution at first glance seems to be a free lunch. However, a more in-depth analysis shows that the growth effects of inequality and redistribution vary with the development level, which is in line with the unified growth theory outlined in Section (1.2). In developing and middle-income countries, income inequality exerts a significantly negative effect on economic development, this being a result of capital market imperfections and the insufficient provision of public goods. When examining the sample of advanced economies, however, the picture changes. In richer economies with—on average—higher equality of opportunities, the negative relationship between inequality and growth disappears. In accordance with the changing effect of inequality across different levels of development, the analysis highlights that redistribution by taxes and transfers is beneficial for growth in poor countries, but harmful in rich economies.

While Chapter (5) examines the role of the state in the growth process by focusing on governmental interventions, Chapter (6) enlarges the analysis by exploring the effect of the political framework on long-run growth. The previous chapters provide clear evidence that individual investment decisions—in physical and human capital, as well as in innovation activity—contribute to economic development. Chapter (6) argues that this mechanism can only take effect if individuals have the ability to carry out their desired investment. Put differently, in order for the growth drivers illustrated in Chapters (2)–(5) to operate, the political framework of the country requires a form of government that provides individual and economic freedom, freedom of contract, property rights, and legal liability. This framework is particularly likely to come into being in democratic countries. Although the definitions of democracy in political science vary, there is a broad consensus that it consists at least of four key elements: (1) A political system for choosing and replacing the administration via free and fair election, (2) the active participation of citizens in the political process, (3) the guarantee of human rights for all inhabitants, and (4) a rule of law which applies equally to all citizens (see Diamond, 2008).

Despite these convincing theoretical mechanisms, the hitherto existing empirical results concerning the effect of democracy are far from being distinct. While some studies find tentative evidence for a positive relationship of political rights and income increases, others stress that the link between these variables is either insignificant or even negative. Chapter (6) argues that the ambiguous effect of democracy is the result of its imprecise measurement. In fact, the most challenging methodological hurdle to compute democracy measures—or composite measures in general—is derivation of an appropriate rule to aggregate the underlying data series into an indicator of democratic institutions. As the latent variable “democracy” cannot be *directly* observed, traditional strategies are based on arbitrary assumptions about the functional relationship between the input variables and the degree of democratization. Chapter (6) demonstrates how this essential problem can be resolved by utilizing mathematical algorithms for machine learning and pattern recognition. These algorithms can be used to place the crucial question of how to aggregate information from different variables into the context of a nonlinear optimization problem, thereby obtaining much more consistent and plausible results.

Specifically, the underlying method uses Support Vector Machines (SVM) to compute a democracy indicator that is continuously on the (0,1) interval, enabling a very detailed and sensitive measurement of democracy for 188 countries in the period between 1981 and 2011, called *“Support Vector Machine Democracy Indicator”* (SVMDI). SVM-based techniques have generated a number of promising results in classification applications in various branches of science; however, little effort has been made to apply this method to economic problems. By maximizing comparability for the broadest possible sample of countries, the SVMDI algorithm facilitates empirical investigations of democracy. A direct result of this methodical progress is a substantial increase in the level of detail in comparison to established approaches. Its application for classification of democracy may be seen as a first introduction of the technique in the field of economics. Yet the unprecedented potential of machine learning enables researchers to make highly accurate classifications, which is why the proposed technique may also yield very promising results for economic problems beyond its utilization for measuring democracy.

In the second step, Chapter (6) analyzes the effect of democracy on growth based on the new measurement of democracy. This examination highlights a robust positive influence of democracy on growth. The results imply that the ambiguity in recent studies stems from two main sources. First, in light of the diversity of political institutions across countries, the lack of a sufficient reaction of traditional democracy indicators to political events and regime changes only allows for a rough classification of democracy, which leads to biases in empirical analyses. Second, when using empirical models that rely on the within-country variation—as is the case in the overwhelming majority of recent studies—, the problem of inadequate and insensitive measurement of democracy becomes particularly severe.

A more in-depth analysis of the democracy-growth nexus provides little indication of a nonlinear relationship between the variables. The analysis of the transmission channels through which democracy exerts its influence on growth illustrates why: whereas democratic countries typically have more educated populations, higher investment shares and lower fertility rates, there is little evidence for a redistribution-enhancing effect of democratization. According to the findings of Chapter (5), the latter would be detrimental to growth in advanced economies.

Finally, Chapter (7) summarizes the main results found in this book and discusses its implications for economic policy. The book concludes with a brief outlook on the long-run determinants upon which future growth potentials may depend.

To complement the analyses conducted in Chapters (2)–(6), the mathematical appendix (Chapter A) provides a very detailed description of the applied mathematical concepts, with a particular focus on the mathematical theory underlying machine learning, pattern recognition algorithms, and Support Vector Machines. This theory is crucial for construction of the SVM-based classification procedure of democracy introduced in Chapter (6). Machine learning approaches are widely used in various branches of science, for instance in physics, computer science, and mathematics. However, as machine learning algorithms have received little attention in the field of economics, the appendix provides a broad introduction in the fundamentals of machine learning, SVM-based methods, the referring dual optimization problems and quadratic programs, the utilization of kernels (also known as the *"kernel trick"*), risk functionals, pattern recognition algorithms, and procedures for estimating optimal hyperplanes in higher-dimensional space. In doing so, the chapter is divided into two parts: Section (A.2) explains how Support Vector machines can be used for estimation of real-valued functions, while Section (A.3) describes their application for estimation of indicator functions and classification problems.

Chapter 2
Entrepreneurship, Culture and Growth

Background With the pioneering work of Solow (1956, 1957) and the extensions made by Barro and Sala-i-Martín (2004), the neoclassical growth theory suggests that technological progress is the main driver of long-run growth in advanced economies. The endogenous growth theories highlight how these technical advancements are developed and implemented (see, for instance, Romer, 1990 and Aghion and Howitt, 1992, 2009). Among those theories, a particularly influential line of reasoning called the *"Schumpeterian"* models—named after the work of Schumpeter (1911, 1942)—focuses on quality-improving innovations that render old products obsolete. Schumpeter (1942) provides an illustration of this process in his theory of *"creative destruction"*. However, these theories pay little attention to cross-country differences in the tendency of individuals to carry-out innovation projects. This tendency is shaped by entrepreneurial behavior and differs considerably across nations.

What is the reason for these systematic cross-country differences in in the propensity to innovate? This chapter shows that large parts of these differences can be traced back to cultural socialization. By influencing attitudes towards risk-aversion, active self-assertion to achieve individual goals, daring, and success, culture substantially influences entrepreneurial behavior, and consequently contributes to differences in innovation activity and growth across nations.

2.1 Introduction

One of the most fundamental questions in the economics of growth is why disparities in living standards across the world are substantial and persistent. A wealth of theoretical and empirical studies have investigated the growth processes occurring during previous decades and centuries that have caused these cross-national differences in incomes and wealthiness, stressing that among the diverse factors that influence long-run economic well-being, a particularly prominent role is played by technological progress (Romer, 1990; Acemoglu et al., 2006).[5] More recently, the

[5] The earliest attempt to study the empirical effect of technological progress relative to other factors dates back to Solow (1957), who found that from 1900 to 1949, 88% of output growth per hour worked in the United States can be attributed to factor productivity growth. Subsequent studies broadly support these findings for other nations and panels of countries (see, e.g., Mankiw et al., 1992; Jones, 1997; Baier et al., 2006).

empirical growth literature has moved from the analysis of proximate factors to the study of more fundamental determinants rooted in long-term history, including geography, institutions, and the population (Spolaore and Wacziarg, 2013). This chapter links the two branches of the literature, raising the question of whether there are deep-rooted factors that distinguish countries in terms of their tendency and ability to spur innovation.

While monetary rewards resulting from patents, market power, and the acquisition of new markets provide strong and universal incentives for individuals to engage in innovation activities, a striking empirical observation is that the diffusion of technology progresses differently across countries (Caselli and Coleman, 2001), and that these disparities are substantial even among countries with similar levels of income and education (Vandenbussche et al., 2006). This suggests that there are other, non-income-related factors which motivate individuals to pursue innovation activities and which, consequently, lead to cross-national differences in technological progress. This chapter argues that cultural socialization accounts for large parts of these non-monetary factors, thereby adding to the burgeoning literature that foregrounds the deep roots of economic development. Although a variety of authors have underscored the fundamental role of culture for socio-economic transformations (Weber, 1905) and economic development (Landes, 1998; Guiso et al., 2006), there have been relatively few attempts to systematically study the empirical growth effect of cultural legacy. Due to the frequent usage of the term "culture" in common parlance and the vagueness which accompanies this ambiguity, it is difficult to design refutable hypotheses based on this concept. To define culture in a sufficiently narrow sense, this analysis assumes it to be *the set of values and beliefs shared by a given group of people, as well as the norms and behaviors derived from that framework, which is passed from one generation to the next*.[6]

Drawing on a similar definition of culture, Gorodnichenko and Roland (2017) demonstrate that by awarding social status to personal accomplishments, countries which are shaped by individualistic values achieve higher levels of real per capita GDP. Similarly, Tabellini (2010) shows that cultural differences with regard to trust and respect are substantially responsible for the differences in GDP growth across 69 European regions in 8 different countries between 1995 and 2000. While there are certainly solid theoretical reasons to believe that individualism, trust, and respect may be growth-enhancing, this study focuses on a cultural trait that is more narrowly connected to innovation and knowledge production: the propensity of the population to engage in entrepreneurial activity.

The importance of entrepreneurship for economic growth has been stressed in early essays of the economic profession (see, e.g., Schumpeter, 1934, 1942) and also in more recent growth theories (Mokyr, 2016; Doepke and Zilibotti, 2014). In their model linking growth to innovation activity carried out by entrepreneurs, Doepke and Zilibotti (2014) show that individuals are faced with a key occupational choice between being a worker and being an entrepreneur. This decision crucially hinges on the degree of risk tolerance, where individuals engaging in entrepreneurial activity

---

[6] A detailed discussion on how to make the concept of culture feasible for systematic research is provided by Hofstede (2001), who applies a very similar definition.

typically possess a greater readiness to take risks, encouraging them to invest in new technologies and products whose outcome is uncertain.[7] As with any dimension of of culture, risk-tolerance along with other entrepreneurship-related traits—such as active self-assertion to achieve individual goals, daring, and success—are passed from one generation to the next, dependent upon a society's understanding of how best to respond to the occupational choice.

While the theoretical models broadly suggest a positive relationship between cultural attitudes towards entrepreneurship and the growth rate, documentation of entrepreneurship's empirical relevance has progressed much more slowly (Doepke and Zilibotti, 2014; Glaeser et al., 2015), and the results are thus far rather ambiguous, particularly in the cross-country context.[8] The empirical literature can be divided into two groups. The first group concludes that entrepreneurship is generally beneficial to the economy (see, e.g., Acs et al., 2004, 2006, Wong et al., 2005, and Audretsch et al., 2008), while a second group of studies finds no robust effects (see, e.g., Salgado-Banda, 2007 and Carree et al., 2007). The goal of the present chapter is to resolve this ambiguity by demonstrating that the inconclusiveness of recent studies has its roots in three methodological pitfalls, namely i) insufficient consideration of endogeneity and reverse causation, ii) sample selection biases, and iii) inadequate identification of entrepreneurial activity.

To tackle the first of these issues, this chapter uses a number of instrumental variables to identify a causal effect running from culture to growth. These instruments include genetic distances measured by the frequency of blood-types (Gorodnichenko and Roland, 2017), a broad set of genetic polymorphisms (Spolaore and Wacziarg, 2009), two specific genetic characteristics (Chiao and Blizinsky, 2010 and Way and Lieberman, 2010), historical prevalence of pathogens (Murray and Schaller, 2010), and linguistic differences (Licht, 2007). To be clear, the argument is not that there may be a causal link running from genetic distance to cultural distance, but rather that both genes and culture are transmitted simultaneously from parents to their offspring, hence there is a strong correlation between both variables (in biological studies, this phenomenon is referred to as "co-evolution", see Chiao and Blizinsky, 2010). Owing to growing concerns about weak instrumentation in empirical growth regressions (Bazzi and Clemens, 2013), the results are reported along with an extensive set of weak-instrument diagnostics.

To account for potential sample selection biases, the empirical strategy exploits cross-sectional and panel data that covers a large sample of countries in the period between 1970 and 2015. The analysis additionally provides a historical perspective, studying the link between culture and growth from 1500 to present. Finally, and most importantly, the chapter shows that the greatest source of ambiguity in recent studies

[7] In fact, this argument is key in the theory of the entrepreneur. In an early essay, Schumpeter (1947) notes that entrepreneurs facilitate "the doing of new things or the doing of things that are already being done in a new way".

[8] Glaeser et al. (2015) emphasizes the lack of cross-country evidence on the effect of entrpreneurship by noting that *"it is quite striking that we now have several studies evaluating the causal links between entrepreneurial finance and industry or city growth [...], but we have very little evidence on entrepreneurship's role more generally."* For empirical studies in a single-country context, see Kortum and Lerner (2000) and Samila and Sorenson (2011).

is the inadequate measurement of entrepreneurship. While in empirical studies entrepreneurship is routinely measured using the self-employment rate, a striking empirical pattern is that self-employment is much more prevalent in the developing world: the average share of self-employed individuals in high-income countries (determined according to the World Bank's critical threshold of 12,236 USD) was 13.92% in the post-2010 period, but it was more than twice as large in low- and middle-income countries (30.91%). These differences stem from the fact that there are a considerably larger number of small firms operated in developing countries (Poschke, 2014) and thus are more reflective of disparaties in labor market conditions than of a tendency towards entrepreneurship as a result of cultural socialization. As a consequence, the parameter estimates in empirical specifications that model a linear relationship between self-employment (or similar proxies of entrepreneurship) and economic growth cause a bias which may be figuratively termed the "corner-shop-effect". To overcome this bias, the analysis estimates the effect of the self-employment rate conditional on the initial development level. These results are reported along with models based on the Total Entrepreneurial Activity (TEA) index of GEM (2017), as well as a more direct measure of the cultural traits shaping a society's propensity to become an entrepreneur of Schwartz (2006).

The analysis provides strong evidence for a significantly positive effect of entrepreneurship and its associated cultural traits on economic growth. This effect is robust to changes in the instrumentation strategy, the time period, and the variable to proxy entrepreneurship and its cultural roots. A similar growth-enhancing influence emerges when we change the focus to a historical perspective covering the period from 1500 to present, proving strong evidence that cultural differences that have their roots in long-run history still influence living standards of today. The outcomes further suggest that the positive effect on growth mainly works via facilitation of technological progress.

2.2 Recent studies and three empirical challenges

2.2.1 A brief overview of the recent empirical literature

From a theoretical perspective, the point of departure for many recent studies is the survey of Thurik and Wennekers (1999), which synthesizes the disparate strands of relevant literature to link entrepreneurship to economic growth. Discussing a variety of theoretical channels between the two concepts, the authors stress that the main conduit of entrepreneurship to growth is stimulation of innovation. In addition, Audretsch and Keilbach (2008) and Audretsch et al. (2008) put great emphasis on the "knowledge spillover theory of entrepreneurship". A prominent argument of Arrow (1962) stresses that knowledge is only partially excludable and creates externalities, which is why the innovation activity of any agent fosters economic growth by enhancing the productivity of other agents (Acs et al., 2004 and Acs et al., 2006) while enabling improvements and adjacent innovations. Alongside these effects, a further source of growth created by entrepreneurs is entrepreneurial capital (see, e.g., Audretsch and Keilbach, 2004a,b and Audretsch, 2007). Shaped by a broad spectrum of factors (policies, institutions, traditions, law, and finance), entrepreneurial capital reflects social acceptance and valuation of entrepreneurial behavior, as well as attitudes towards risks.

Whereas most of the theoretical literature focuses on innovation activities, a different branch is concerned with growth resulting from imitation. By closing the gap between the current technology and the technological frontier, imitative activity stimulates knowledge diffusion and contributes to income increases, particularly in less developed countries (Acemoglu et al., 2006 and Perla and Tonetti, 2014).

Empirical evidence that sheds light on the link between entrepreneurship and growth is quite sparse, a shortcoming that has been criticized by Doepke and Zilibotti (2014) and Glaeser et al. (2015). The recent empirical work can be divided into two separate branches, the first of which is involved with the direct examination of the growth effect of entrepreneurship, and the second dealing more generally with the role of culture in the development process. With respect to the former, several studies have found an empirical regularity in the form of a positive relationship between economic growth and various measures of entrepreneurial activity. Applying fixed effects estimations for 20 OECD countries in the period between 1981 and 2000, Acs et al. (2004) and Acs et al. (2006) report a positive effect of entrepreneurship on growth and find some evidence for the knowledge spillover theory. Further evidence in this vein is provided by Audretsch and Keilbach (2008) and Audretsch et al. (2008) based on regional data for Germany. Van Stel et al. (2005) and Wong et al. (2005) apply OLS estimations for the period 1999–2003 using a panel of 36 countries (Van Stel et al., 2005) and in a cross-section of 37 economies (Wong et al., 2005). These studies generally confirm the positive influence found by previous research, but these studies identify an effect that is much less pronounced and only partially significant.

A contradictory result is obtained by Salgado-Banda (2007), who employs data from 22 countries for the 1980–1995 period, finding a positive relationship between growth and entrepreneurship measured via frequency of patent applications but a *negative* relationship when utilizing the self-employment rate. In a similar manner, Carree et al. (2007) provide evidence that growth penalties occur if the extent of entrepreneurship deviates from its equilibrium rate.

Focusing on the broader link between cultural values and economic performance, the second branch of empirical studies has produced results that are much less controversial, emphasizing a substantial influence of cultural differences on income disparities across the globe. These studies highlight the role of individualistic attitudes in fostering technological progress (Gorodnichenko and Roland, 2017) and further point to a beneficial effect of trust and respect in the growth process (Tabellini, 2010). In addition, Mokyr (2016) stresses the influence of culture on the development of growth-friendly institutions.

2.2.2 Three challenges to identify the empirical effect of culture and entrepreneurship on growth

Given the discrepancy in the implications of the two branches of studies, a key question is how this ambiguity can be explained. When analyzing the econometric setting of recent attempts to assess the growth effect of entrepreneurship, three substantial challenges are identified that are not adequately addressed. These challenges involve i) insufficient consideration of endogeneity and reverse causation, ii) sample selection biases, and iii) inadequate identification of entrepreneurial activity.

There are strong arguments for a reverse causation in the entrepreneurship-growth relationship, which, however, only few studies take into account. While the bulk of empirical strategies rest on OLS estimations, the approach of Acs et al. (2006) uses age and unemployment to instrument the self-employment rate. Additionally, Salgado-Banda (2007) utilizes GDP in previous periods and legal origin to eliminate the endogenous components in the data. However, neither of these approaches reports any investigation of the exclusion restriction. From a theoretical viewpoint, it is highly unlikely that this restriction is met, as this would only be the case if unemployment, age, GDP, and legal origin affected growth exclusively via their influence on entrepreneurship.

The second empirical challenge concerns the sample used to identify the effect of entrepreneurship. Most analyses draw on information from a limited number of countries, ranging from around 20 included nations (Acs et al., 2006; Acs et al., 2004; Salgado-Banda, 2007) to slightly less than 40 economies (Van Stel et al., 2005; Wong et al., 2005). A similar data limitation arises with respect to the time dimension, as most studies employ cross-sectional regressions based on data from the early 1980s to the late 1990s and early 2000s. The differences in sample sizes and the number of included countries and periods impedes direct comparison of the results, which is why the inconclusiveness may, to some extent, simply be traced back to sample selection biases. This is particularly relevant as the compositions of the included countries and their development levels strongly deviate across studies.

The final challenge, however, features the most severe pitfall: commonly applied measurements of entrepreneurship are inadequate to proxy the theoretical channels through which entrepreneurial attitudes and cultural traits are assumed to benefit growth, particularly facilitation of innovation and technological progress via a high tendency to engage in risky projects. This argument is illustrated in Figure (2.1), which graphs the self-employment rate across country income groups and continents for the period between 2010 and 2015. The figure emphasizes that the self-employment rate is particularly high in countries possessing low-and-middle and middle incomes, while it is lowest in the group of advanced economies.

What is the reason for this strong and counterintuitive deviation? The studies concerned with the role of firm establishment offer an explanation for the self-employment puzzle, stressing that the foundation of firms positively influences job creation (Reynolds, 1999) and helps to reduce unemployment (Picot et al., 1998). However, a substantial part of this unemployment-reducing effect is achieved by forcing individuals into self-employment (Evans and Leighton, 1989, 1990 and

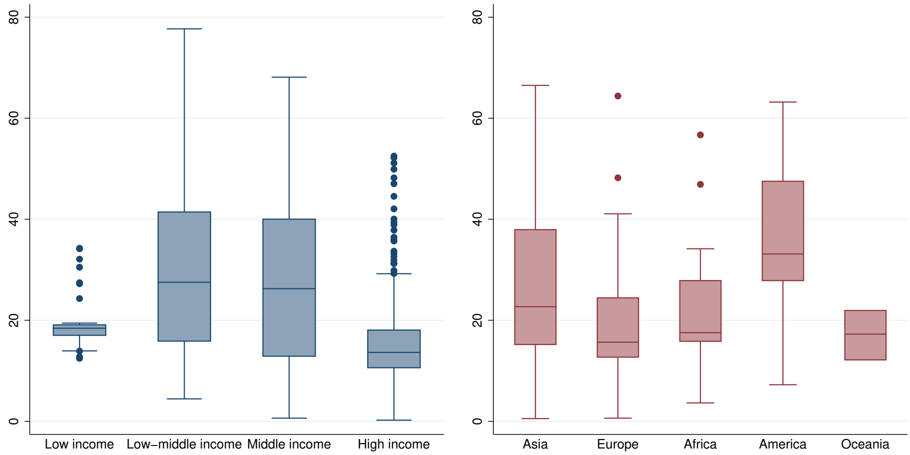


Figure 2.1 The distribution of self-employment among income groups (left panel) and continents (right panel), average of 2010–2015. Selection of income groups follows the World bank's country classification: low income (0–1,005 USD), low-and-middle-income (1,006 USD–3,955 USD), middle income (3,956 USD–12,235 USD), and high income (12,235 USD or more).

Reynolds, 1999). This may supply the formerly unemployed business owners with a—often temporary—source of income, but it contributes very little to economic growth (Van Stel and Storey, 2004). By showing that developing countries possess a considerably larger number of small firms, the results of Poschke (2014) imply that the "corner-shop effect" is much more prevalent in poorer economies. This provides an explanation for why the self-employment rate is higher in developing countries, and also explains why empirical studies modeling a linear relationship between entrepreneurship and growth find ambiguous results.

2.3 The theoretical framework: Linking entrepreneurship to economic growth

In order to profoundly study the empirical effect of entrepreneurship on long-run economic development, it is necessary to obtain refutable hypotheses on a potential link between the variables. For this reason, this section develops an endogenous growth model that illustrates the circumstances under which entrepreneurship—and, more importantly, its cultural origination—influences the growth rate.

2.3.1 The baseline model

As a starting point, consider a Romer (1987, 1990) framework with Dixit and Stiglitz (1977) preferences for varieties. The model enlarges the basic approach of the model class with an expanding variety of products as described by Barro and Sala-i-Martín (2004). An introduction to this class of models can be found in Section (1.2). Modeling a continuum of intermediate goods, the production function of firm i can be denoted as

$$y_i = \Psi L_i^{1-\alpha} \int_{\mathbb{R}^+} x_{ij}^{\alpha}, \tag{2.1}$$

where Ψ is factor productivity, L is labor, and x_{ij} gives the amount x of intermediate good $j \in \mathfrak{P}$ as used by the ith firm, $i = 1, ..., I$. Each j features diminishing marginal returns which yields $\alpha \in (0, 1)$. This condition also leads to constant returns to scale in the production function. Equation (2.1) illustrates the output potential if all entropy is exhausted, i.e. $N = |\mathfrak{P}| = |\mathbb{R}^+| = 2^{\aleph_0}$. Yet the nature of N implies two assumptions: first, when there is no war or natural disaster, N increases monotonically, $(\mathrm{d}N/\mathrm{d}t) \geq 0 \, \forall t$. Second, it follows that $N \subset |\mathbb{R}^+|$, which allows for reformulation of Equation (2.1) using a discrete and finite number of available specialized intermediate goods (Barro and Sala-i-Martín, 2004)

$$y_i = \Psi L_i^{1-\alpha} \sum_{j=1}^{N} x_{ij}^{\alpha} \,. \tag{2.2}$$

This formulation provides an intuitive economic interpretation and an illustration of the growth effect if N increases. Assuming that all j can be measured in consistent physical units and that $x_{ij} = x_i$, Equation (2.2) adjusts to

$$y_i = \Psi L_i^{1-\alpha} N x_i^{\alpha}. \tag{2.3}$$

It follows that each increase in N enhances the income level of the economy. Particularly, an increase in the term $N x_i^{\alpha}$ that emerges due to the utilization of a higher quantity of existing j encounters diminishing returns. In contrast, an increase in N does not result in diminishing returns.

According to Equation (2.2), the profit for i is

$$y_i - wL_i - \sum_{j=1}^{N} P_j x_{ij},$$

where w is wage and P_j is the price for j. As the marginal product of j is $(\partial y_i / \partial x_{ij}) = \Psi \alpha L_i^{1-\alpha} x_{ij}^{\alpha-1}$, the marginal product of P_j implies the quantity of j demanded as a function of price as $x_{ij} = L_i (\Psi \alpha / P_j)^{1/(1-\alpha)}$.

Households maximize utility over an infinite horizon and have CRRA preferences, i.e. $U = \int_0^{\infty} \left[\left(c^{1-\lambda} - 1 \right) / (1-\lambda) \right] \exp\{-\rho t\} \mathrm{d}t$. The common aggregate budget constraint is $(\mathrm{d}a/\mathrm{d}t) = wL + ra - C$, as households gain a rate of return r on assets a and receive wage w. The Euler equation then assumes the familiar form $\dot{C}/C = (1-\lambda)(r-\rho)$.

2.3.2 The role of entrepreneurship

What is the role of entrepreneurship in this framework? As Equation (2.3) illustrates, the main growth effect arises due to an increase in N, which raises the question of how such increases can be achieved. There are basically two ways to enhance N: either through national innovation activity or through the imitation or the import of existing goods. The latter, however, only enables temporary growth effects up until the point at which the world technological frontier is reached.

When residing in a closed economy, firms must innovate in order to acquire blueprints of new intermediate goods j^*. Following Barro and Sala-i-Martín (2004) and Romer (1990), the capital value $V(j)_t$ of any existing j at t is

$$V(j)_t = \int_t^\infty (P_j(v) - 1)x_j(v)\exp\{-\bar{r}(v,t)(v-t)\}\mathrm{d}v, \tag{2.4}$$

where $\bar{r}(v,t)$ denotes the average interest rate between t and v and $\pi_j(v) = (P_j(v) - 1)x_j(v)$ is the cashflow stream at time v. In order to calculate the profit of blueprint j^*, it is required to know the aggregate demand of any j from producers i. This is simply the sum of x_{ij} over all i, i.e.

$$x_j(v) = \sum x_{ij}(v) = L[\Psi\alpha/P_j(v)]^{1/(1-\alpha)},$$

where the latter term follows from the demanded quantity of x_{ij} as described previously. As new blueprints create a monopolistic position for the innovator, P_j can be chosen to maximize π_j. In order to set incentives for the creation of j^*, most countries have implemented patent protection laws, which ensure that the innovator remains in the monopolistic position, at least for a short period. In the absence of intertemporal elements in the demand function and any state variables on the production side, the maximization problem is simply

$$\max_{P_j(v)} \pi_j(v) = (P_j(v) - 1)L[\Psi\alpha/P_j(v)]^{1/(1-\alpha)} \tag{2.5}$$

and is solved for $P_j(v) = (1/\alpha) > 1$ (see Barro and Sala-i-Martín, 2004 for a similar case). For each new j^*, however, the innovating firm has to bear invention costs η. Using the above values for P_j and x_j in Equation (2.4) and bringing the constants in front of the integral gives the following condition for an investment in the invention of j^*

$$\eta \leq L\Psi^{1/(1-\alpha)}\frac{1-\alpha}{\alpha}\alpha^{2/(1-\alpha)}\int_t^\infty \exp\{-\bar{r}(v,t)(v-t)\}\mathrm{d}v\,. \tag{2.6}$$

Henceforth, define $\Omega \equiv \frac{1-\alpha}{\alpha}\alpha^{2/(1-\alpha)}$ for reasons of lucidity. Equation (2.6) illustrates the investment decision under certainty, that is for known values of $V(j^*)_t$. However, since the capital value relies on assumptions of the particular individual and can not be ex ante correctly anticipated, the decision to invest in j^* must include risk. The degree of risk tolerance introduces a cultural element in the model. The reason is that members of a particular society share a similar tendency to tolerate or refuse risks. This tendency distinguishes the society from other groups, depending on the collective mental programming that is passed from one generation to the next (see Hofstede, 2001). In the tradition of Schumpeter (1911), Knight (1921), and Kihlstrom and Laffont (1979), the national degree of entrepreneurship is reflected in a relatively low aversion against risk. Under risk, Equation (2.6) adjusts to

$$\eta < \mathrm{E}[V(j^*)_t],$$

where the expected value of $V(j^*)_t$ depends on the right-hand side of Equation (2.6) and the probability of success, so that $\mathrm{E}[V(j^*)_t] = V(j^*)_t - V(j^*)_t(1-p)$. The equation shows that an increasing probability of failure reduces the expected capital value and thus the likelihood that the innovation activity is carried out. As $p \in (0,1)$, its distribution can be modeled via $p \sim \mathrm{Beta}(a,b)$. If $a = b > 1$, the distribution is symmetric with $\mathrm{E}[p] = a/(a+b) = 0.5$.[9] Yet in most cases, the probability of failure cannot be correctly anticipated ex ante. Hence, the investment decision relies on the risk-averseness of i rather than on the “true” risk of j^*.

The best choice under complete uncertainty would be $a/(a+b)$ if empirical values of the shape parameters can be estimated. But even if for some reason i knows the exact value of p, the subjective assessment of that risk may easily differ between different choice makers and depends on the willingness or reluctance to take risks. Put differently, the subjective assessment of the risk associated with j^* varies between the individuals. For instance, one individual may hold a probability of failure of 50 percent as acceptable, whereas another individual may already consider a risk of 20 percent unreasonable. Hence, the effect of risk-averseness on innovation activity strongly resembles the effect of the probability of failure. Let $\theta_i \in (0,1)$ denote the extent of risk-averseness of an individual. The interpretation of that measure is the reluctance to take risks in percent (as measured relative to the most risk-averse individual). An individual with $\theta_i = 1$ would only make decisions under complete certainty, whereas the willingness to take risks increases as θ_i declines. Regarding the realization of an innovation, higher values of risk-averseness for given probabilities of failure have the same effect as higher probabilities of failure for given values of risk-averseness, so that $\theta_i \propto (1-p)$. As the analysis in this chapter focuses on cross-national differences in growth, the model henceforth turns to the more general case where $\theta = I^{-1}\sum_{i=1}^{I}\theta_i$ reflects the average risk-averseness shared by members of a particular country. The investment decision then becomes

$$\eta < \mathrm{E}[V(j^*)] = V(j^*) - \theta V(j^*) = (1-\theta)V(j^*)\,.$$

The above equation shows that a low level of risk-averseness enhances the probability to invest in j^* as it increases the capital value an individual expects from the research activity. The condition to invest in an innovation then adjusts to

9 For $a < b$, the Beta distribution is right-skewed, indicating that more innovations fail than become successful.

$$\frac{\eta}{1-\theta} < V(j^*)\,. \tag{2.7}$$

As Equation (2.7) illustrates, the extent of risk-averseness crucially influences the investment decision of inventing j^* under uncertain returns. The probability of investing in j^* rises as θ approaches zero, while increasing θ make investments in j^* more and more unlikely.

In equilibrium, it must hold that $\eta(1-\theta)^{-1} = V(j^*)_t$ under the free-entry condition. Differentiating this condition with respect to time, using $V(j^*)_t$ from Equations (2.4) and (2.6) and applying the condition $\bar{r}(t,v) = [1/(v-t)]\int_t^v r(\xi)\mathrm{d}\xi$, it follows that[10]

$$r(t) = \frac{\pi}{V(j^*)_t} + \frac{\dot{V}(j^*)_t}{V(j^*)_t}\,.$$

If η is a constant and θ is an inherent cultural factor that cannot change over time, it follows that $\dot{V}(j^*)_t = 0$. In this case, the interest rate simplifies to $r(t) = \pi/[\eta(1-\theta)^{-1}]$. Using π from the maximization problem illustrated in Equation (2.5), the interest rate becomes

$$r = \frac{L}{\eta(1-\theta)^{-1}}\Psi^{1/(1-\alpha)}\Omega\,.$$

Substituting the interest rate in the Euler equation gives the growth rate of the economy

$$\frac{\dot{y}}{y} = (1/\lambda)\left[\frac{L}{\eta(1-\theta)^{-1}}\Psi^{1/(1-\alpha)}\Omega - \rho\right]\,. \tag{2.8}$$

Entrepreneurship influences this growth rate through two channels: first, if individuals have a low aversion to risk, the growth rate increases. This is because θ is negatively correlated with the decision to invest in j^*. Second, each innovation creates an externality as it brings with it an increase in knowledge Ψ as a by-product of the innovation process. Thus, innovations from one firm provide a second firm with the possibility of investing in an innovation that perhaps would not have been profitable—or technologically conceivable—with the previously existing technology. This reflects the knowledge-enhancing effect of innovations via diffusion of that knowledge.

Recalling the relationship $\theta \propto (1-p)$, the growth rate reveals another interesting property: whenever innovation opportunities are good, the growth rate increases. Yet in periods where the best opportunities are exhausted—i.e. in which the probability of success is low—, the growth rate will decline. Thus, fundamental technological progress that enables a number of j^* with favorable prospects enhances incomes as p increases. Inversely, if the low-hanging fruits are picked, growth rates will drop until radically new inventions and technologies arise. Chapter (4) of this doctoral thesis

10 Following Leibniz's rule for differentiation under the integral sign.

attends to this issue and the resulting consequences for economic growth in greater detail.

Equation (2.8) illustrates that the long-run growth rate depends on entrepreneurship, measured via a country's average risk-aversion. As emphasized by Doepke and Zilibotti (2014), risk-aversion is part of cultural inheritance and is the result of learned behavior, which is passed from one generation to the next. This underlines that the essential characteristic of entrepreneurship for long-run growth can be traced back to culture.

Taken together, the testable implication from the theory derived in this section is that a higher average propensity of a nation's citizens to carry out entrepreneurial activity should correspond with a higher rate of economic growth. This propensity is influenced by cultural socialization, which is why we may expect that culture exerts a strong effect on long-run development.

2.4 Estimation technique and identification strategy

In the next step, we bring the theoretical arguments of Section (2.3) to the data. In doing so, it is crucial to account for the three challenges to assess the effects of entrepreneurship and its cultural dimensions on economic growth identified in section (2.2). To circumvent the first two pitfalls, the econometric specification employs a two-part strategy, employing instrumental variable strategies based on i) cross-sectional estimations that draw on current and historical data, and ii) panel data models that cover the period 1970 to 2015 for a large number of countries.

2.4.1 Cross-sectional estimations

The baseline econometric model studies the statistical effects of entrepreneurship and culture on growth via

$$y_{it} = \alpha + \gamma C_{it} + \boldsymbol{\theta}\mathbf{X}_{it-\tau} + \eta_j + v_{it}, \tag{2.9}$$

where y_i is the log of real per capita GDP in country i located on continent j, C_{it} denotes the applied measure of entrepreneurship and culture, and $\mathbf{X}_{it}$ is an array of control variables that includes education and investment, which are key determinants of economic well-being. Equation (2.9) is estimated for different 5-year periods t to rule out the possibility that the choice of the period influences the identified effect and to study changes in the modeled relationship over the short- and long-run history. Averaging the data is necessary to eliminate short-run fluctuations and to ensure that the empirical model captures the long-run impact of C_{it} on per capita income. The baseline estimates refer to the period 2005–2009, for which data availability is maximized, with control variables reflecting investment and education in the 1970–1974 period ($\tau = 7$). The lag structure ensures measurement of a causal effect rather than co-movements, which may be caused by a reverse causation.

Figure (2.1) has illustrated that there are strong disparities in self-employment across continents. To address the specific environments of different regions, Equation

(2.9) includes continent fixed effects η_j, implemented via a set of dummy variables $\sum_{j=1}^{J} \beta_j D_{ij}$. Data regarding real per capita income and investment is collected from Penn World Table version 9 (Feenstra et al., 2015) and—for the historical analysis—from the New Maddison Project Database (Bolt and van Zanden, 2014). Education is measured via average years of schooling and drawn from Barro and Lee (2013).[11]

To estimate Equation (2.9), we employ two different empirical strategies. The first strategy is OLS, which has been used in a number of recent studies dealing with the consequences of culture and entrepreneurship for economic outcomes (Gorodnichenko and Roland, 2017; Van Stel et al., 2005; Wong et al., 2005 and many more). While OLS enables direct comparison with previous studies on the entrepreneurship-growth nexus, it provides little information on causality. The second strategy therefore uses an instrumental variable approach to rule out reverse causation. The 2SLS version of Equation (2.9) is expressed by

$$y_{it} = \alpha_y + \gamma_y C_{it} + \boldsymbol{\theta}_y \mathbf{X}_{it-\tau} + \sum_{j=1}^{J} \beta_{yj} D_{ij} + v_{y,it} \tag{2.10}$$

$$C_{it} = \alpha_C + \gamma_C \Omega_{it} + \boldsymbol{\theta}_C \mathbf{X}_{it-\tau} + \sum_{j=1}^{J} \beta_{Cj} D_{ij} + v_{C,it}, \tag{2.11}$$

where Ω is the instrumental variable for entrepreneurship and its cultural traits.

2.4.2 Instruments used for the 2SLS regressions

When studying culture, a substantial challenge is to disentangle its effects from those of institutions. The two variables exhibit a symbiotic relationship (Hofstede, 2001; Tabellini, 2008) and complement each other (Alesina and Giuliano, 2015), but there is still the possibility of a causal link running from either variable to the other. Additionally, as the growth literature emphasizes that a rich set of factors influence the wealth of nations (Durlauf et al., 2005), omission of variables in Equation (2.9) may result in a correlation between C_{it} and the error term. To tackle these issues and to identify a causal relationship between culture and other variables, a commonly applied strategy is the epidemiological approach, linking the attitudes and behavior of immigrants to cultural traits prevalent in their countries of origin (Luttmer and Singhal, 2011; Fernández, 2011). However, this strategy does not entirely solve the problem of endogeneity, as different groups of immigrants may well encounter different informal institutional frameworks in their destination countries (Rauch and Trindade, 2002; Maseland, 2013). Glaeser et al. (2015) use geographical proximity to historical mining deposits to instrument firm size and start ups in the United States, which is, however, not a feasible approach in a cross-country setting.

[11] Table (A2-1) in the appendix provides descriptive statistics for the variables used in the empirical analysis, including their means, standard deviations, and the number of observations, as well as their minima and maxima.

This analysis follows a more recently developed branch of the empirical literature, making use of the observation that cultural differences are strongly correlated with biological and linguistic characteristics (Tabellini, 2008; Gorodnichenko and Roland, 2017). This correlation is the result of a co-evolution of genes and collective behavior, as both genetic information and social behavior are transmitted from parents to their children (Chiao and Blizinsky, 2010).

The empirical analysis brings this argument to the data by using several variables measuring the prevalence of certain genes and pathogens in a given population. The baseline specification of Equation (2.11) measures genetic differences between countries based on frequencies of blood types, as data on the frequency of alleles which determine blood types is more widely available than any other genetic information. For this reason, utilization of blood type frequency allows for construction of a measure of genetic distance that is available for a large set of countries. More specifically, the analysis follows Gorodnichenko and Roland (2017) in measuring genetic distance via the Mahalanobis distance between the frequency of blood types in the United Kingdom and their frequency in a given country. The Malahanobis distance between vectors p and q picked from distributions Φ is given by (see Malahanobis, 1936)

$$\mathrm{d}_M(p,q) = \{(p-q)^{'}\Sigma_{\Phi}^{-1}(p-q)\}^{\frac{1}{2}}, \tag{2.12}$$

where the covariance matrix Σ_Φ in terms of frequencies of blood type A ($f_{A,c}$) and B ($f_{B,c}$) in country c is given by $\Sigma_\Phi = \text{var}\left(\bar{f}_{A,c}\ \bar{f}_{B,c}\right)$. As information on blood types is available at the level of ethnic groups while that of economic outcomes and measures of culture is available at the country level, $\bar{f}_{A,c}$ denotes the country-level frequency of blood type A that is calculated based on the frequency of blood type A of ethnic group e in country c via

$$\bar{f}_{A,c} = \sum_e \psi_{e,c} f_{A,e,c},$$

where ψ denotes the ethnic shares of the population that are drawn from Fearon (2003). Instrumenting culture via Equation (2.12) is advantageous due to the fact that blood types are "neutral" genetic markers, i.e. they do not influence fitness or intelligence and therefore do not directly influence economic outcomes. Moreover, economic development is unlikely to influence genetic pools over the short time period that may be reconstructed with the available data. Blood type frequencies are taken from Gorodnichenko and Roland (2017), who collect data from sources predominately stemming from the 1950s and 1960s. When using this data to instrument economic outcomes of more recent periods, the temporal distance helps alleviate endogeneity concerns with respect to between-country migration. Moreover, Cavalli-Sforza et al. (1994) document that for the vast majority of countries, genetic variation was largely influenced during the Neolithic migration of early humans.

In order to rule out the possibility that the results are triggered by the chosen instrumentation strategy, the robustness of the findings is evaluated based on additional information on genetic distance. This information includes the frequency of the S-allele in the serotonin transporter gene 5HTTLPR (Chiao and Blizinsky, 2010),

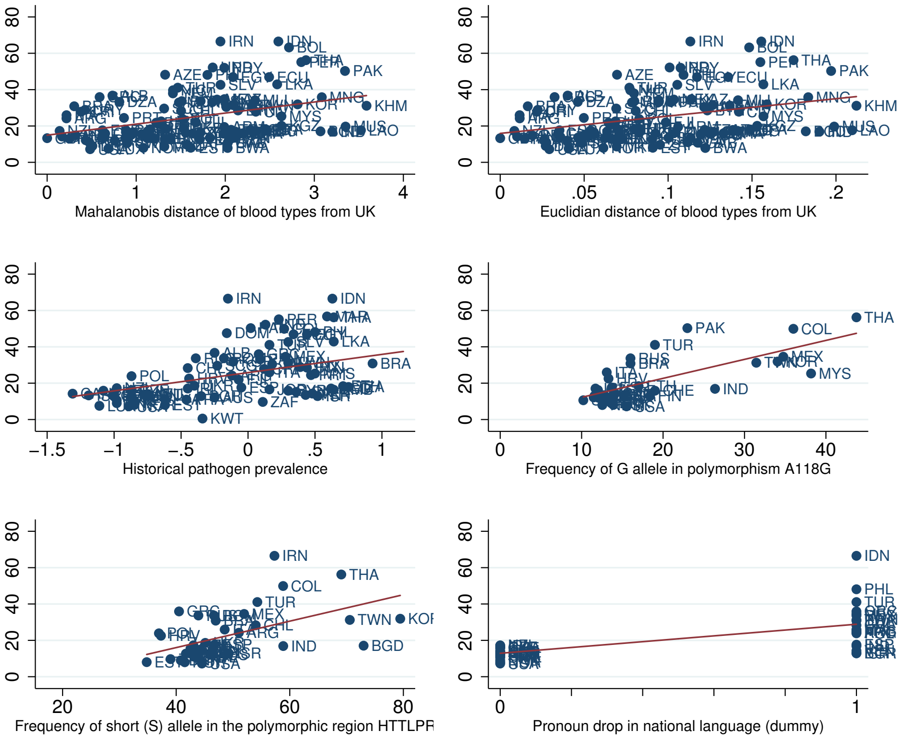


Figure 2.2 Instruments used in the analysis and their relationship to entrepreneurship (measured via the self-employment rate).

the frequency of the G-allele in polymorphism A118G in the μ-opoid receptor gene (Way and Lieberman, 2010), historical prevalence of pathogens (Murray and Schaller, 2010), and a richer set of genetic distance obtained from Spolaore and Wacziarg (2009).[12] Similar to blood types, these biological characteristics are unlikely to influence economic growth *directly*.[13]

As a final robustness check, the analysis exploits the entanglement between culture and language, an aspect which has been emphasized by Tabellini (2008) and Licht (2007). Utilization of language as an instrument for culture may be traced

[12] The historical pathogen dataset of Murray and Schaller (2010) includes nine pathogens: leishmanias, trypanosomes, malaria, schistosomes, filariae, leprosy, dengue, typhus, and tuberculosis.

[13] A potential violation of this requirement may be argued to exist with respect to pathogen prevalence, which may influence growth via channels other than culture, predominately the level of health. However, descriptive analyses show that life expectancy and certain kinds of pathogens are only weakly correlated. For instance, the correlation between the prevalence of the pathogen *Toxoplasma Gondii* and life expectancy at birth in the 2010–2014 period is -0.22. Cross-country data on the prevalence of Toxoplasma gondii is extracted from Pappas et al. (2009).

back to what is now referred to as the “Sapir-Whorf” or the “Liguistic Relativity” hypothesis (Whorf, 1956; Sapir, 1970). As argued by Kashima and Kashima (1998), culture can be linked to linguistic phenomena, particularly to pronoun drop in the case of person-indexing pronouns. For instance, while the English phrase *“I run”* is equivalent to the German expression *“Ich renne”*, neglect of the pronoun is quite common in other languages, such as Spanish and Italian (where this phrase would most commonly be expressed simply as *“corro”*, and the pronouns “*Yo*” and “*Io*” are dropped and the context can be ascertained from the verb). The hypothesis of Kashima and Kashima (1998) is that the requirement of pronoun usage is a result of the psychological differentiation between speakers and their social contexts, where utilization of pronouns is particularly prevalent in individualistic societies. As with the other instruments, it is unlikely that language affects economic performance, thus the required exclusion restriction is likely to be satisfied.

Figure (2.2) displays the relationship between the instrumental variable and entrepreneurship measured via the self-employment rate. The figure underlines the positive correlation that we may expect based on theory. In addition, the relationship between differences in genes and differences in cross-national tendencies to become self-employed is not influenced by the statistical measure used to gauge blood type differences, as the correlation found with respect to the Malahanobis distance strongly resembles the correlation obtained via measurement of blood type distance with the Euclidean distance.

2.4.3 Panel data models

As a second strategy to ensure that a) the analysis identifies a causal link that is b) not influenced by a sample selection bias, we also estimate the effect of self-employment and culture on real per capita GDP growth in a panel data setting. This analysis deviates from the IV strategy in two essential aspects: First, endogenous components are eliminated using lagged variables rather than biological characteristics. Second, as lags are more widely available than genetic information and also vary over time, the number of included country-years is significantly increased.

Equation (2.9) can be transformed into a dynamic panel variant by modeling

$$g_{it} = y_{it} - y_{it-1} = \rho y_{t-1} + \gamma C_{it} + \boldsymbol{\theta}\mathbf{X}_{it-1} + \eta_i + \zeta_t + v_{it}, \tag{2.13}$$

where g_{it} is the growth rate of real per capita GDP, ζ_t is a specific effect of period t, and v_{it} is the idiosyncratic error. The dynamic panel data context of Equation (2.9) further allows for the direct incorporation of fixed effects η_i of country i rather than using continental effects. Equation (2.13) is estimated via the system GMM estimator (Blundell and Bond, 1998), which is widely used in recent growth regressions and

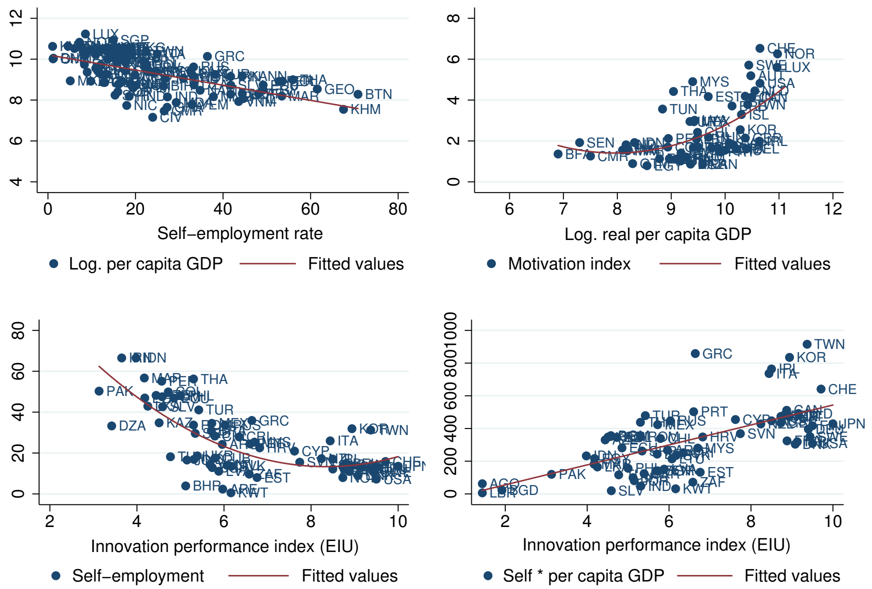


Figure 2.3 Self employment, development levels, and innovation activity. Interaction of self-employment and real per capita GDP (demoted with self × per capita GDP) in 1,000.

documented in detail in the existing empirical literature (see, e.g. Roodman, 2009a,c; Halter et al., 2014; Arcand et al., 2015 and many more).[14]

2.5 Data on entrepreneurship and self-employment and the “Corner-Shop-Effect”

The most challenging pitfall that needs to be circumvented when empirically assessing the effect of entrepreneurship on growth is the observation that variables which are traditionally used to proxy entrepreneurship—primarily, but not exclusively, the self-employment rate—fail to capture the tendency to take risks and to engage in innovation activity as a result of cultural segregation. The reason for this failure is that differences in labor market conditions blanket the “true” entrepreneurship tendency in the data.

[14] Some cultural dimensions are time-invariant *per definition*, as it is passed with little fundamental changes from one period to the next. Contrary to a difference GMM framework, inclusion of such effects is possible when employing system GMM. Asymptotically, the inclusion of time-invariant regressors also does not affect coefficient estimates for other regressors, as all instruments for Equation (2.13) are assumed to be orthogonal to fixed effects and other time-invariant regressors (for a detailed explanation, see Roodman, 2009c).

Figure (2.3) illustrates this problem and suggests a feasible solution. The first graph underlines the negative correlation between self-employment and the development level. As suspected in Section (2.2), this relationship is primarily due to labor market conditions. This is depicted in the second graph, which displays the linkage between the motivation index—reflecting the share of the self-employed that are improvement-driven and opportunity motivated divided by the percentage that is necessity-driven—and per capita GDP. The result is a negative relationship between self-employment and innovation activity, measured by the innovation performance index of EIU (2009) (correlation: -65 %). This, however, contradicts what an entrepreneurship proxy aims to measure and explains why earlier studies based on self-employment data find ambiguous growth effects.

The final graph shows how this problem may be circumvented by employing an interaction between self-employment and initial real per capita GDP. The relationship between the interaction term and the innovation index is strongly positive (roughly 60%). Intuitively, the positive link between the interaction term and innovation activity implies that self-employment is beneficial for innovation in counties with a higher development level, while it is less important for knowledge production in poorer economies.

To incorporate this argument in the empirical analysis, we include the interaction term in the empirical model of Equation (2.9), i.e. the baseline specification augments to

$$y_{it} = \alpha + \gamma C_{it} + \beta C_{it} \times y_{it-\tau} + \boldsymbol{\theta}\mathbf{X}_{it-\tau} + \eta_j + v_{it}. \tag{2.14}$$

We compare the results obtained via approximation of C by the self-employment rate with two adjustments, including the Total Entrepreneurial Activity (TEA) index of GEM (2017) and a cultural variable that more narrowly captures the hypotheses of the theoretical model of Section (2.3), collected from Schwartz (2006). Founded in 1999 by a joint project of Babson College (US) and London Business School (UK), GEM has since developed into a vast international research network including more than 300 research institutions, whose aim is to collect data on entrepreneurial behavior and attitudes of individuals. The TEA index is a central indicator developed by the GEM network, reflecting the percentage of the population aged between 16 and 64, who are either nascent entrepreneurs or owner-managers of a new business.

The Schwartz (2006) measures of culture are collected via survey responses from college-students and schoolteachers, containing 195 samples drawn from 78 countries between 1998 and 2000. While he computes seven cultural value orientations, the *mastery* dimension provides a good proxy for the cultural traits determining the tendency towards entrepreneurship in a society. This dimension gauges active self-assertion in order to master, direct, and change the natural and social environment to attain group or personal goals. As Schwartz (2006) underscores, values such as ambition, success, daring, and competence are especially important in mastery cultures.

2.6 Empirical Results

2.6.1 The effect of entrepreneurship and culture on economic growth

Table (2.1) reports the baseline results, which illustrate the effects of the entrepreneurship variables on economic growth that are identified via a standard OLS setting and the instrumentation strategy documented in Section (2.4). For both estimation strategies, the table reports the results of the reduced model of Equation (2.9), the augmented model that is enlarged by the interaction of entrepreneurship and initial GDP from Equation (2.14), and a comprehensive model that includes control variables for investment and human capital, two key drivers of long-run development identified in the growth literature. The estimates are based on the maximum number of countries for which data is available.

Panel A documents the results when proxying entrepreneurship with the self-employment rate. The parameter estimate of the reduced model shown in Column (1) is negative and statistically significant at the 1% level, indicating a negative relationship between self-employment and economic well-being. This result supports the findings of Salgado-Banda (2007) and others who question the contribution of entrepreneurship in the form of self-employment to economic outcomes. This result is also in line with the negative relationship illustrated in Figure (2.3). However, when estimating this effect by OLS, the direction of causality is unclear. For this reason, Column (4) shows the effect of self-employment instrumented with the Malahanobis distance of blood types relative to the United Kingdom. The outcome of this exercise allows for two key conclusions: first, the effect remains significantly negative, which implies that causality runs from higher degrees of self-employment to lower income levels. This provides support for the hypothesis that high rates of self-employment are the result of a large number of rather unproductive firms that have their origin in poor labor market conditions and "refugee effects". Second, the effect size measured via 2SLS is substantially greater compared with the OLS outcomes, underscoring the need to eliminate endogeneity in the data.

Columns (2) and (5) augment the reduced model by the interaction of self-employment and the development level. This specification dramatically changes the empirically drawn implications: while the effect of the self-employment rate remains significantly negative, the interaction term between self-employment and GDP possesses a positive sign and is statistically significant at the 1% level. This result is obtained in both the OLS and the 2SLS setting, where the parameter estimate is again larger with respect to the instrumental variable approach. Economically, the results of Columns (2) and (5) suggest that self-employment does not contribute to income increases in developing countries, but plays a crucial role in the growth process in advanced economies. These results are further in line with the Unified Growth Theory, which stresses that the key driver of growth changes over the course of the development process (see, e.g., Galor, 2011): while less developed countries mainly grow via physical capital investment and, in later stages, advances in human capital, the primary growth engine in advanced economies is technological progress.

Table 2.1 Entrepreneurship, culture, and economic growth — Baseline estimates.

	OLS Estimates			2SLS Estimates		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A-1: Self-employment rate					
Self-Employment	-0.0232***	-0.290***	-0.233***	-0.110***	-0.411***	-0.330***
	(0.00608)	(0.0218)	(0.0221)	(0.0418)	(0.0542)	(0.0830)
Self × GDP		0.0320***	0.0258***		0.0466***	0.0367***
		(0.00257)	(0.00228)		(0.00813)	(0.00959)
Investment (1970)			-0.141			-0.914
			(0.844)			(1.035)
Years of Schooling (1970)			0.134***			0.0833*
			(0.0300)			(0.0475)
	Panel A-2: First-stage results					
Blood Distance				5.0945***	15.7833**	1.0212**
				(1.8792)	(6.740)	(0.4223)
Countries	113	113	86	113	113	86
R-Squared	0.52	0.83	0.86	0.97	0.76	0.82
F-Stat (second)	22.85***	63.95***	62.39***	9.65***	59.07***	41.64***
F-Stat (first)				7.35***	4.35**	30.44***
AR-Test				21.64***	426.37***	248.67***
UCI interval (self)				[-0.094;-0.0174]	[-0.638;-0.115]	[-0.620;-0.060]
UCI interval (interaction)					[0.016; 0.065]	[0.009; 0.069]
	Panel B-1: Total Entrepreneurial Activity					
TEA	-0.0687***	-0.393***	-0.267***	-0.152***	-0.915***	-0.554**
	(0.00992)	(0.0968)	(0.0662)	(0.0579)	(0.277)	(0.229)
TEA × GDP		0.0421***	0.0280***		0.108***	0.0630**
		(0.0114)	(0.00782)		(0.0353)	(0.0283)
	Panel B-2: First-stage results					
Blood Distance				2.5536**	1.3075***	0.2855*
				(1.0222)	(0.3795)	(0.1451)
Countries	65	65	54	65	65	54
R-Squared	0.72	0.82	0.91	0.46	0.56	0.83
F-Stat (second)	155.4***	43.22***	78.79***	12.99***	16.09***	43.61***
F-Stat (first)				6.24**	3.88*	3.87*
AR-Test				17.68***	7.84***	3.83*
UCI interval (tea)				[-0.225;-0.074]	[-1.161;-0.369]	[-0.976;-0.048]
UCI interval (interaction)					[0.039; 0.138]	[0.001; 0.115]
	Panel C-1: Culture (Schwartz): Mastering					
Culture	0.683	-2.247***	-2.252***	13.15**	-2.641***	-2.253***
	(0.603)	(0.0736)	(0.0505)	(6.220)	(0.504)	(0.167)
Culture × GDP		0.254***	0.251***		0.258***	0.254***
		(0.00234)	(0.00333)		(0.00731)	(0.00400)
	Panel C-2: First-stage results					
Blood Distance				-0.049*	-20.862***	-13.7601***
				(0.0251)	(2.0551)	(2.0463)
Countries	68	68	53	68	68	53
R-Squared	0.52	0.99	0.99	0.96	0.99	0.99
F-Stat (second)	15.64***	2692.6***	3785.0***	2.14*	1307.2***	2715.8***
F-Stat (first)				3.83*	61.45***	99.12***
AR-Test				24.30***	259.04***	187.40***
UCI interval (culture)				[1.881;5.203]	[-3.387;3.492]	[-0.637;-0.057]
UCI interval (interaction)					[0.203;0.273]	[0.227;0.287]

Notes: Dependent variable is the log of real per capita GDP in the 2005–2009 period. Panel A reports the results based on the self-employment rate, Panel B employs the Total Entrepreneurial Activity (TEA) index, and Panel C draws on Schwartz's cultural dimension measuring mastering. Estimation techniques are OLS (Columns 1–3) and 2SLS (Columns 4–6) with robust standard errors reported in parentheses. Instrumental variable is the Mahalanobis distance of blood types A and B relative to the United Kingdom. To economize space, country-dummies are not displayed. First-stage results are reported in Panels A-2, B-2, and C-2. F-Stat (first) gives the F-statistic of the first-stage and the corresponding significance level. AR-Test gives the Anderson-Rubin Wald test to conduct weak-instrument-robust inference. UCI interval reports the union of confidence intervals (UCI) test of Conley et al. (2012) for plausibly exogenous instruments. $*p < .10, **p < .05, ***p < .01$

Consequently, entrepreneurship predominantly spurs growth in countries at later stages in the development process.

Finally, the comprehensive models reported in Columns (3) and (6) indicate that the conditional effect of entrepreneurship possesses a high degree of stability. The marginal effects are, however, somewhat lower. The slight decline in the effect of the interaction term that occurs when accounting for investment and schooling suggests that part of entrepreneurship's growth effect is channeled via education and investment. However, as the decrease in the effect is only marginal, the hypothesis is that innovation activity is the most important transmission channel (which is assessed in Section (2.6.3)).

Panels B and C are based on the same empirical specifications, but use the TEA index and the cultural dimension of mastery from Schwartz (2006) to investigate the stability of the results found in Panel A. Even though data availability results in different sample compositions, the outcomes obtained via the self-employment rate reappear in both Panel A and Panel B.

Naturally, the results crucially hinge on the goodness of the empirical fit and, most importantly, the ability of genetic distance to instrument differences in cultural values and entrepreneurial activity. Panels A-2, B-2, and C-2 provide first-stage results and a battery of statistical tests to investigate the exclusion restriction and instrument strength. R-squared and the (second stage) F-statistic suggest that the estimated models sufficiently fit the data. The first-stage results shown in Panels A-2, B-2, and C-2 demonstrate that blood type distance is a strong predictor of the entrepreneurship variables, being significant in all models and significant at the 1% level in the majority of cases. Instrument strength is further underscored by the F-statistic of the first-stage regressions, which is (strongly) statistically significant in each case. Finally, Table (2.1) reports the outcomes of the weak-instrument-robust test initially developed by Anderson and Rubin (1949). The AR-test provides the advantage that its statistics are robust to weak instruments in the sense that they have the correct size in cases when instruments are weak, and in those when they are not. The null of insignificance of the conditional effect of entrepreneurship on growth is strongly rejected by the AR-test in each of the modeled specifications, pointing to a considerable degree of instrument strength.

A crucial assumption for the validity of the results in Table (2.1) is satisfaction of the exclusion restriction. In the context of our instrumentation strategy, the exclusion restriction essentially means that genetic distance does not affect a nation's wealth via channels other than culture. While there are good theoretical reasons for why this should indeed be the case (see Section 2.4.2), the Union of Confidence Interval (UCI) test of Conley et al. (2012) provides a statistical tool for performing inference while relaxing the exclusion restriction. The test goes beyond the usual investigation of a potential correlation of the instrument and the error term, assessing the consequences for statistical inference if genetic distance were only to be “plausibly exogenous”. To briefly illustrate its implementation, consider a reduced version of the IV setting of Equation (2.11)

$$y_i = \gamma C_i + \zeta \Omega_i + u_i,$$

in which the necessary assumption is $\zeta = 0$. The UCI test deviates from that assumption by using some $\zeta \neq 0$ that is specified by the researcher and returning the union of all interval estimates of γ conditional on a grid of all possible values for ζ, which is reported in Table (2.1) as UCI interval.[15] The results emphasize that even if we substantially relax the exclusion restriction, inference based on the employed instrument would still be informative.

2.6.2 Alternative instrumental variables

The Conley et al. (2012) test displayed in Table (2.1) suggests that even if blood-type distance were only to be "plausibly exogenous", the conditional effect of entrepreneurship and culture on growth would still be significantly positive. However, in order to further evaluate causality, another strategy to investigate the exclusion restriction involves alteration of the instrumentation strategy. Table (2.2) replicates the empirical specifications used to compute the baseline outcomes while using five alternative instrumentation strategies, which are described in Section (2.4.2). Due to the limitation in data availability with respect to the genetic and linguistic information used in Table (2.2), the number of observations declines compared with the baseline table.

The results obtained via alteration of the instrumentation strategy provide strong support for the findings of Table (2.1). Specifically, the conditional effects of entrepreneurship and its cultural traits on economic well-being are positive and strongly significant in each case. There is, however, a minor deviation to be observed with regard to Schwartz' mastery criterion: while the conditional growth effect found with the help of blood type distance is maintained, the results now suggest a (slightly) positive impact of mastery even in less developed economies. Regardless of this initial impact, the growth effect of mastery intensifies during the development process.

The weak-instrument diagnostics reported in the respective panels of Table (2.2) indicate that the alternative instruments are even stronger than blood-type distance. However, as the utilization of these instruments is accompanied with a substantial decline in the number of included countries, Table (2.1) continues to be the preferred specification. It should further be noted that the reduction in data availability necessitates caution in the interpretation of the results. This drawback notwithstanding, the outcomes of Table (2.2) support the baseline findings in identifying a strong conditional effect of entrepreneurship on per capita GDP.

A further robustness check is reported in Table (A2-2) in the appendix, which displays the results based on instruments that are generated following the Lewbel (2012) approach. This technique exploits model heteroscedasticity to construct instruments as functions of the utilized data. The results support the previous findings in models based solely on Lewbel instruments and in those that supplement Lewbel instruments by blood type distance.

[15] Following Persson and Tabellini (2009), the test is implemented based on a regression of growth rates on both the entrepreneurship variables and genetic distance to obtain an estimate of the degree of bias, which serves as an estimate for ζ.

Table 2.2 Entrepreneurship, culture, and economic growth — Alternative instruments.

	Pronoun drop	Genetic dist.	Hist. pathogens	5HTTLPR	A118G
	Panel A-1: Self-employment rate				
Self-Employment	-1.092***	-0.328	-1.398***	-1.566*	-1.907***
	(0.395)	(0.294)	(0.520)	(0.843)	(0.616)
Self × GDP	0.152***	0.0802**	0.226***	0.211**	0.245***
	(0.0421)	(0.0381)	(0.0596)	(0.0877)	(0.0629)
	Panel A-2: First-stage results				
Instrument	103.214**	0.161**	112.294***	1.362***	3.523***
	(47.196)	(0.079)	(26.058)	(0.779)	(1.001)
Countries	36	92	71	37	30
R-Squared	0.84	0.62	0.86	0.70	0.75
F-Stat (first)	48.38***	24.04***	42.89***	67.98***	66.68***
AR-Test	1.6E+05***	311.79***	156.00***	2170.25***	72.73***
	Panel B-1: Total Entrepreneurial Activity				
TEA	-4.709***	-3.089**	-4.213***	-4.961***	-6.477***
	(1.170)	(1.225)	(0.927)	(1.217)	(0.959)
TEA × GDP	0.598***	0.432***	0.551***	0.635***	0.784***
	(0.129)	(0.147)	(0.108)	(0.132)	(0.104)
	Panel B-2: First-stage results				
Instrument	41.191***	0.213***	73.426***	0.635***	1.258***
	(6.908)	(0.054)	(22.046)	(0.168)	(0.412)
Countries	32	56	53	32	25
R-Squared	0.91	0.79	0.80	0.76	0.85
F-Stat (first)	36.78***	52.93***	60.72***	43.75***	22.87***
AR-Test	1.2E+05***	85.18***	73.60***	1785.87***	62.60***
	Panel C-1: Culture (Schwartz): Mastering				
Culture	0.519**	0.225	0.0950	0.0882	0.296*
	(0.219)	(0.180)	(0.165)	(0.222)	(0.160)
Culture × GDP	0.199***	0.230***	0.247***	0.246***	0.226***
	(0.0243)	(0.0215)	(0.0168)	(0.0225)	(0.0158)
	Panel C-2: First-stage results				
Instrument	3.289***	0.016**	23.301***	0.052***	0.126***
	(0.280)	(0.008)	(2.470)	(0.007)	(0.040)
Countries	35	55	50	33	28
R-Squared	0.99	0.96	0.92	0.96	0.99
F-Stat (first)	6218.07***	75.06***	88.36***	26.35***	37.2***
AR-Test	1.6E+05***	136.02***	179.72***	2639.09***	76.20***

Notes: Dependent variable is the log of real per capita GDP in the 2005–2009 period. Panel A reports the results based on the self-employment rate, Panel B employs the Total Entrepreneurial Activity (TEA) index, and Panel C draws on Schwartz's cultural dimension, which measures mastering. Estimation techniques is 2SLS with robust standard errors reported in parentheses. Instrumental variable are (with primary data source in parentheses): pronoun drop (Licht, 2007), genetic distance (Spolaore and Wacziarg, 2009), historic pathogen prevalence (Murray and Schaller, 2010), the frequency of the short (S) allele in the polymorphic region 5HTTLPR of serotonin transporter gene (SLC6A4) (Chiao and Blizinsky, 2010), and the frequency of the G allele in polymorphism A118G in μ-opoid receptor gene (Way and Lieberman, 2010). A detailed description of the instruments is provided in Section (2.4.2). F-Stat (first) presents the F-statistic of the first-stage and the corresponding significance level. AR-Test expresses the Anderson-Rubin Wald test to conduct weak-instrument-robust inference. $*p < .10, **p < .05, ***p < .01$

2.6.3 The effect of entrepreneurship and culture on innovation activity

In line with the hypothesis that entrepreneurship channels to growth mainly via innovation activity (Section (2.3)), Table (2.3) alters the dependent variable, studying the effect of culture and entrepreneurship on three measures of technical progress and innovation: the first is the TFP measure from Hall and Jones (1999) (reported in Panel A), which gauges differences in productivity in the late 1980s. The second and third measures draw on the latest available data on factor productivity and innovation, namely the innovation performance index from EIU (2009) (Panel B), and total factor productivity growth computed by TED (2017) (Panel C). The innovation performance index consolidates information from different sources, including patents granted to applicants (by residence) per million population, but also alternative innovation indicators such as high-technology manufacturing and service output per head, citations from scientific and technological journals, and royalty and license fee receipts per GDP (see EIU, 2009). The TFP growth rates reported by TED (2017) are estimated as the growth accounting residual, capturing changes in output that are not caused directly by changes in labor and capital inputs.

Table (2.3) follows the same structure used in the previous tables, reporting the effect of entrepreneurship and culture while including varying control variables. Overall, the table suggests a strong impact of entrepreneurship and culture on innovation activity. The results further imply that this impact is (much) stronger when employing instrumental variables regressions, pointing to a potential reverse causation between innovation and entrepreneurship. This is in line with the Schumeperian growth models, which stress the growth-effects of improvements made possible by innovations. In addition, the parameter estimates imply that the growth effect is strongest when proxying entrepreneurship via its cultural traits using the mastery criterion of Schwartz (2006).

2.6.4 Historical data

This section turns the focus to a historical perspective, studying the growth effect of culture over the past 500 years. Analysis of macroeconomic history provides the opportunity to overcome the “rare events” problem (Schularick and Taylor, 2012), which is why the recent empirical growth literature increasingly employs data sets that not only reach back decades, but even centuries (see Madsen and Ang, 2016; Madsen et al., 2015; Almunia et al., 2010; Barro, 2009). This literature stresses that economic development has “deep roots” in history (Spolaore and Wacziarg, 2013). While the bulk of empirical studies along this line of reasoning argue for the growth effect of geography (Diamond, 1997; Ashraf and Galor, 2011) and institutions (Acemoglu et al., 2002), Putterman and Weil (2010) show that it is not only the historical legacy of geographic locations, but also the historical legacy of the populations that determines present income levels. This effect has been studied in greater detail by Easterly and Levine (2016), who find that particularly the share of European ancestry matters for development. While these studies conclude that the growth

Table 2.3 Entrepreneurship, culture, and the effect on innovation activity.

	OLS Estimates				2SLS Estimates	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Total Factor Productivity (Hall and Jones, 1999)					
Self-Employment	-0.00817*	-0.0605**	-0.0790***	-0.0710**	-0.339**	-0.191***
	(0.00485)	(0.0237)	(0.0281)	(0.0295)	(0.143)	(0.0733)
Self × GDP		0.00626**	0.00798**		0.0388**	0.0166**
		(0.00287)	(0.00328)		(0.0169)	(0.00742)
N (F-Stat)	85 (5.3***)	82 (7.5***)	76 (5.2***)	85 (4.2***)	82 (6.1***)	76 (5.3***)
TEA	-0.0385***	-0.118**	-0.165**	-0.169**	-0.769**	-1.184**
	(0.0115)	(0.0458)	(0.0618)	(0.0703)	(0.344)	(0.469)
TEA × GDP		0.0103*	0.0155*		0.0920**	0.140**
		(0.00576)	(0.00773)		(0.0433)	(0.0574)
N (F-Stat)	58 (5.6***)	57 (5.1***)	54 (6.5***)	58 (1.7)	57 (2.2*)	54 (1.5)
Culture	0.181	-1.352**	-1.876***	8.693**	-0.976***	5.133
	(0.528)	(0.601)	(0.570)	(3.688)	(0.254)	(4.412)
Culture × GDP		0.123***	0.198***		0.0737**	0.159**
		(0.0199)	(0.0226)		(0.0309)	(0.0763)
N (F-Stat)	55 (3.4**)	54 (13.0***)	50 (22.9***)	55 (1.7)	54 (46***)	50 (1.7)
	Panel B: Innovation Performance Index (EIU, 2017)					
Self-Employment	-0.0531***	-0.337***	-0.124	-0.204***	-0.822***	-0.335**
	(0.0121)	(0.112)	(0.0870)	(0.0770)	(0.304)	(0.140)
Self × GDP		0.0331**	0.0116		0.0887**	0.0375**
		(0.0130)	(0.0101)		(0.0357)	(0.0191)
N (F-Stat)	70 (65.5***)	67 (30.2***)	55 (35.2***)	70 (5.7***)	67 (12.8***)	55 (25.0***)
TEA	-0.179***	-1.204***	-0.272	-0.496*	-3.410**	-0.265
	(0.0409)	(0.327)	(0.261)	(0.262)	(1.412)	(0.705)
TEA × GDP		0.124***	0.0266		0.386**	0.0306
		(0.0373)	(0.0302)		(0.170)	(0.0970)
N (F-Stat)	53 (31.3***)	51 (63.0***)	44 (528.7***)	53 (9.2***)	51 (10.8***)	44 (54.8***)
Culture	-0.881	-6.694***	-6.333***	38.63	-6.234***	-6.853***
	(1.693)	(0.942)	(1.429)	(33.73)	(0.872)	(2.494)
Culture × GDP		0.626***	0.474***		0.788***	0.846***
		(0.0381)	(0.0925)		(0.0782)	(0.300)
N (F-Stat)	51 (8.4***)	51 (66.8***)	42 (52.2***)	51 (0.7)	51 (650.2***)	42 (312.8***)
	Panel C: Total Factor Productivity Growth (TED, 2017)					
Self-Employment	0.0398***	-0.0157	-0.0917*	0.0819***	-0.126	-0.228**
	(0.0142)	(0.0721)	(0.0537)	(0.0187)	(0.108)	(0.106)
Self × GDP		0.000897	0.0109		0.0248**	0.0288**
		(0.00880)	(0.00664)		(0.0126)	(0.0129)
N (F-Stat)	98 (10.3***)	96 (5.5***)	65 (7.4***)	98 (39.3***)	96 (772.4***)	65 (11.4***)
TEA	0.0775***	-0.204	-0.0232	0.144***	-0.404**	-0.890*
	(0.0154)	(0.130)	(0.184)	(0.0368)	(0.166)	(0.474)
TEA × GDP		0.0338**	0.0123		0.0635***	0.130**
		(0.0152)	(0.0241)		(0.0215)	(0.0631)
N (F-Stat)	64 (248.5***)	62 (17.0***)	51 (114.1***)	64 (9.8***)	62 (6.6***)	51 (3.5***)
Culture	0.286***	-0.127	-0.139	0.385***	-0.786	-0.966
	(0.0992)	(1.122)	(1.352)	(0.128)	(1.059)	(1.334)
Culture × GDP		0.0173	0.0249		0.184	0.156
		(0.122)	(0.150)		(0.155)	(0.187)
N (F-Stat)	65 (9.9***)	65 (1.56)	48 (42.1***)	65 (10.0***)	65 (5.3***)	48 (6.2***)

Notes: Dependent variables are total factor productivity compiled by Hall and Jones (1999), the innovation performance index proposed by the Economic Intelligence Union (EIU, 2009), and factor productivity growth in the 2000–2004 period collected from the Total Economy Database (TED, 2017). The specifications of the models are identical to those employed in Table (2.1). $*p < .10$, $**p < .05$, $***p < .01$

effects of ancestral populations work via human capital, only little attention is paid to a cross-country analysis of cultural factors and attitudes towards entrepreneurship.

The most debilitating hurdle to investigate entrepreneurship's long-run effect on growth is the unavailability of long-run time-series. As the series for self-employment and the indicators compiled by GEM only go back decades, the historical analysis rests entirely on the cultural dimension of entrepreneurship. Even though measures of cultural values do not cover a sufficiently long time-span to evaluate their historical contribution to development, a prominent argument raised by Cavalli-Sforza et al. (1994) is that cross-country variations in genetic characteristics were largely determined during the Neolithic migration. In a similar vein, cultural attitudes are passed over multiple generations, resulting in little changes with respect to fundamental values and norms. Hofstede (2001) documents that cultures are "extremely stable over time", stressing the self-regulating quasi-equilibrium between culture and institutions, the latter being both reinforcements and products of the dominant cultural value system. This long-run character allows for the approximation of historical cultural values with the help of current levels of the Schwartz (2006) criterion of mastery.

Table (2.4) illustrates the conditional effects of entrepreneurship culture in five historical periods, including 1500, 1600, 1700, 1820, and 1900. The 1820 period is selected due to data available in the Maddison database, where a substantial leap in country coverage occurs with respect to the early 1820s. The empirical model uses the specification employed in Column (5) of Table (2.1), which includes the mastery criterion and the interaction of mastery with the development level in an IV setting. While the first column of Table (2.4) uses the preferred instrumentation strategy based on blood type differences, the following columns re-estimate Column (1) and apply the alternative instrumentation strategies discussed in Section (2.6.2).

The results obtained from historical data strongly support the conditional effect of entrepreneurship found in the previous sections. This effect turns out be very stable across the different time periods and the different instrumentation strategies. Additionally, the size of the historical effect (0.293 in 1500 and 0.261 in 1900) strongly resembles the magnitude found in the baseline estimates of Table (2.1) (0.254 in the recent period). These results are in line with Comin et al. (2010), who demonstrate that technology adoption in 1500 is a significant predictor for today's income per capita and technology adoption. Interpreting their results through the lens of Table (2.4) suggests that long-lasting attitudes towards entrepreneurship may be the key drivers behind the strong intergenerational correlation of technology adoption. Conversely, this further implies that the transmission channels of entrepreneurship's growth effect in historical periods may strongly resemble the transmission mechanism of today.

Taken together, the outcomes of the historical analysis provide evidence that cultural traits that encourage entrepreneurial behavior have constantly fueled economic well-being over the past 500 years. Thus, historical differences in culture may to some extent have contributed to contemporary income disparities between countries.

Table 2.4 Culture and economic growth — Historical data.

	Blood Types	Pronoun	Genetic dist.	Hist. path.	5HTTLPR	A118G
	Panel A: Incomes in 1500					
Culture	-9.74	-11.02***	0.0105	-10.11	-7.981***	-4.870*
	(10.12)	(3.663)	(6.053)	(6.582)	(2.365)	(2.760)
Culture × GDP_{1500}	0.293***	0.272***	0.257***	0.271***	0.269***	0.265***
	(0.024)	(0.00417)	(0.0105)	(0.00938)	(0.00338)	(0.00428)
N (F-Stat)	27	24	27	27	22	20
	(5499***)	(9115***)	(2853***)	(1.0E05***)	(1.6E05***)	(21.46***)
	Panel B: Incomes in 1600					
Culture	0.229	-11.77**	2.322	-1.994	-3.545	- 1.209
	(6.719)	(4.856)	(4.512)	(5.398)	(3.426)	(3.231)
Culture × GDP_{1600}	0.265***	0.271***	0.254***	0.258***	0.261***	0.258***
	(0.018)	(0.00636)	(0.00734)	(0.00658)	(0.00516)	(0.00486)
N (F-Stat)	25 (7.28***)	23 (7.01***)	25 (3.00*)	25 (7.46***)	21	20
					(11.15***)	(13.27***)
	Panel C: Incomes in 1700					
Culture	-3.105	-12.89**	-0.663	1.480	-0.195	1.964
	(13.62)	(5.969)	(5.027)	(5.285)	(3.423)	(2.981)
Culture × GDP_{1700}	0.283***	0.272***	0.258***	0.254***	0.256***	0.254***
	(0.034)	(0.00830)	(0.00722)	(0.00462)	(0.00420)	(0.00345)
N (F-Stat)	27 (858***)	24	27	27	22	20
		(2851***)	(2860***)	(9535***)	(8864***)	(13.50***)
	Panel D: Incomes in 1820					
Culture	-2.830	-12.89**	-7.104*	44.68	1.870	6.463
	(6.991)	(6.318)	(4.012)	(161.3)	(3.789)	(4.194)
Culture × GDP_{1820}	0.254***	0.271***	0.271***	0.220*	0.254***	0.250***
	(0.008)	(0.00872)	(0.00694)	(0.124)	(0.00398)	(0.00395)
N (F-Stat)	41 (121***)	33	41	41	29	26
		(5325***)	(6391***)	(5121***)	(1.2E05***)	(1.3E05***)
	Panel E: Incomes in 1900					
Culture	-5.737	-11.16	-15.70***	159.1	13.91*	10.04
	(4.652)	(7.079)	(3.943)	(368.8)	(7.280)	(7.120)
Culture × GDP_{1900}	0.261***	0.262***	0.270***	0.205*	0.249***	0.250***
	(0.009)	(0.00467)	(0.00398)	(0.112)	(0.00333)	(0.00319)
N (F-Stat)	44	33	43	42	31	24
	(9398***)	(4930***)	(5704***)	(3359***)	(10.25***)	(9399***)

Notes: Dependent variables are real per capita incomes in 1500, 1600, 1700, 1820, and 1920. Table reports the effect of culture on these variables based on all instruments used in the previous section (with primary data source in parentheses): blood type distance to the UK (Gorodnichenko and Roland, 2017), pronoun drop (Licht, 2007), genetic distance (Spolaore and Wacziarg, 2009), historic pathogen prevalence (Murray and Schaller, 2010), the frequency of the short (S) allele in the polymorphic region 5HTTLPR of the serotonin transporter gene (SLC6A4) (Chiao and Blizinsky, 2010), and the frequency of the G allele in polymorphism A118G in the μ-opoid receptor gene (Way and Lieberman, 2010). A detailed description of the instruments is provided in Section (2.4.2). Estimation technique is 2SLS with robust standard errors reported in parentheses. F-Stat shows the F-statistic of the second stage and the corresponding significance level. AR-Test presents the Anderson-Rubin Wald test to conduct weak-instrument-robust inference. $*p < .10$, $**p < .05$, $***p < .01$

2.6.5 Panel data estimations

The results so far rest on cross-sectional analyses conducted at various time-periods. This section provides a final robustness check on the conditional effect of entrepreneurship based on panel data. The rationale for this strategy is that the estimation technique i) is based on an entirely different philosophy on how to eliminate endogeneity in the data, ii) rests on a considerably larger number of observations, and iii) takes into account serial correlation and unobserved heterogeneity on the country-level. Additionally, the larger number of country-year observations allows for the incorporation of a richer set of growth determinants for which recent empirical studies in growth research traditionally control for. Table (2.5) provides outcomes for four specifications, including the models that account for the effect of entrepreneurship without (Column 1a) and with (Column 2a) covariates, as well as models that augment these specifications by the interaction term of entrepreneurship and GDP (Columns 1b and 2b). As in the previous tables, the results rest on the three proxies of entrepreneurship, which are presented in Panels A–C.[16]

The panel outcomes provide strong support for the hypothesis that entrepreneurship benefits GDP growth in advanced economies. As in the cross-section models, the effect of entrepreneurship is either virtually null or negative if the specification models a linear relationship between entrepreneurship and growth. In the augment models that include the conditional effect of entrepreneurship, the picture changes dramatically, pointing to a strong growth effect of entrepreneurship in countries that are in later stages of the development process. The results are most strongly pronounced with respect to the cultural dimension of entrepreneurship reported in Panel C and somewhat less pronounced with respect to the TEA index. However, due to the restriction in data availability of the TEA index with respect to the time dimension, these estimates are obtained based on significantly fewer observations compared to the remaining panels.

Regarding instrument validity and strength, the lower parts of each panel provide a battery of weak instrument diagnostics. The AR(1) and AR(2) tests emphasize the absence of second-order serial correlation for each model, satisfying a crucial requirement for the validity of dynamic panel GMM estimates. To assess the strength of the internal instruments used in the regression, Table (2.5) reports the weak instrument F-test proposed by Sanderson and Windmeijer (2016) (henceforth SW), which builds on Angrist and Pischke (2009) but allows for separate diagnostics for each endogenous regressor. Test statistics are reported for both the difference and the level equations.[17] The SW test for weak instrumentation shows that in each model specification, the relative IV bias is less than 30 percent in both the level and the difference equations, and is often even (well) below 10 percent. Additionally, the difference-in-Hansen test suggests the superiority of system GMM over difference GMM by confirming the validity of the instrument subsets used for the level-equation in each model.

[16] The analysis includes a set of standard controls used in growth regressions: the investment share, the log of life expectancy, the log of the fertility rate, the average years of school attainment, and government consumption (for a detailed description of the expected growth effects, see Barro, 2003, 2013a). The data is obtained from standard sources, i.e. World Bank (2017), Barro and Lee (2013), and Feenstra et al. (2015).

[17] In general, the SW tests are designed for weak-instrumentation diagnostics in the multiple-endogenous-regressor context of 2SLS settings. These tests can be transferred to dynamic panel GMM frameworks (Bun and Windmeijer, 2010; Newey and Windmeijer, 2009) via construction of the exact GMM instrument matrix for both the difference and the levels equations of the system GMM estimator. This matrix can in turn be used to carry out the standard 2SLS regressions and test.

Table 2.5 Entrepreneurship, culture, and economic growth — Panel estimations.

	Linear model		Interaction terms	
	(1a)	(2a)	(1b)	(2b)
	Panel A: Self-employment rate			
$\text{Log}(\text{GDP}_{pc})(t-1)$	-0.00432 (0.00395)	-0.0193*** (0.00647)	-0.0213*** (0.00756)	-0.0273*** (0.00764)
Self-Employment	0.000338 (0.000229)	0.0000308 (0.000222)	-0.00771** (0.00320)	-0.00342* (0.00200)
Self $\times$ GDP			0.000933** (0.000376)	0.000397* (0.000236)
Investment Share		0.000627* (0.000357)		0.000574 (0.000355)
Log Life Expectancy		-0.00167 (0.0735)		-0.0109 (0.0884)
Log Fertility Rate		-0.0252** (0.0122)		-0.0273** (0.0132)
Years of Schooling		0.00273 (0.00241)		0.00288 (0.00244)
Government Cons.		-0.00113 (0.000877)		-0.00173** (0.000796)
Observations	624	547	624	547
Countries	166	134	166	134
Hansen p-val	0.007	0.159	0.070	0.548
Diff-in-Hansen	0.199	0.812	0.822	0.997
SW F-Stat (diff)	7.96	587.73	26.16	9.29
SW F-Stat (lev)	4.14	6.50	6.74	12.01
SY 30% rel IV bias	4.13	3.91	4.03	4.01
AR(1) p-val	0.003	0.003	0.002	0.002
AR(2) p-val	0.362	0.245	0.284	0.204
Instruments	67	117	87	137
	Panel B: Total Entrepreneurial Activity			
TEA	-0.000993 (0.000774)	-0.000108 (0.000634)	-0.0142 (0.0137)	-0.0153* (0.00912)
TEA $\times$ GDP			0.00147 (0.00153)	0.00179* (0.000990)
Observations	178	168	178	168
Countries	82	75	82	75
Hansen p-val	0.051	0.488	0.021	0.077
Diff-in-Hansen	0.886	0.914	0.635	0.509
SW F-Stat (diff)	6.65	21.71	11.49	7.11
SW F-Stat (lev)	10.27	12.05	7.26	4.54
SY 30% rel IV bias	4.92	4.51	4.86	4.48
AR(1) p-val	0.698	0.549	0.480	0.408
Instruments	23	73	25	52
	Panel C: Culture (Schwartz): Mastering			
Culture	-0.0513 (0.0463)	-0.0434 (0.165)	-0.508*** (0.0500)	-0.483*** (0.101)
Culture $\times$ GDP			0.0540*** (0.00663)	0.0485*** (0.0111)
Observations	602	554	602	554
Countries	108	94	108	94
Hansen p-val	0.120	0.298	0.816	0.336
Diff-in-Hansen	0.557	0.982	0.999	0.994
SW F-Stat (diff)	85.38	9.74	5.22	4.84
SW F-Stat (lev)	16.82	12.05	4.87	11.37
SY 30% rel IV bias	4.14	4.48	4.34	4.41
AR(1) p-val	0.000	0.000	0.000	0.003
AR(2) p-val	0.266	0.602	0.107	0.100
Instruments	63	29	29	31

Notes: Dependent variable is real per capita GDP growth. Estimation technique is system GMM, with Windmeijer-robust standard errors in parentheses. Hansen p-val givers the p-value of Hansen's J-test, Diff-in-Hansen reports the p-value of the Difference-in-Hansen test. SW F-Stat reports the F-statistic of the Sanderson and Windmeijer (2016) test, with critical values presented as SY 30% rel IV bias collected from Stock and Yogo (2005). AR(n) p-val presents the p-value of the AR(n) test. Due to the reduced time dimension with respect to the TEA variable, computation of AR(2) is unfeasible. $*p < .10$, $**p < .05$, $***p < .01$

The tests further underscore that inclusion of the conditional effect of entrepreneurship is statistically necessary: while Hansen's J-test implies that the null of joint validity of the utilized instruments cannot be rejected with respect to Panel C and the augmented models in Panels A and B (in which we are primarily interested), the statistic falls below 0.1 in the models of Panel A and B that neglect the conditional effect of entrepreneurship. Following Roodman (2009b), this may be a sign of an omitted variable problem in the reduced specifications.

Table (A2-3) in the appendix provides additional robustness checks based on alternative panel data techniques, including traditional within-group models and the LSDV bias-corrected estimator of Bruno (2005a,b).[18] The results turn out to be very stable when altering the estimation technique. The conditional effect of entrepreneurship in these cases is even stronger than in Table (2.5). In sum, the results obtained via panel techniques strongly underscore the findings of the previous sections.

2.7 Conclusions

The results documented in this chapter provide robust evidence for a positive effect of entrepreneurship and its cultural traits on economic development. This effect is conditional on the development level: while entrepreneurship facilitates income increases in advanced economies, the aggregate effect on economic well-being in less developed countries is zero or even negative. While innovation is the main conduit of entrepreneurship attitudes, large self-employment rates in developing nations are merely the result of poor labor market conditions rather than the inheritance of attitudes conductive to entrepreneurial behavior. This argument helps to reconcile the strong positive effect of entrepreneurship emphasized in theoretical models with the ambiguity found in empirical studies, all of which model a linear relationship between self-employment—or related indicators—and growth. The positive effect of entrepreneurship can also be found when conducting a historical analysis, suggesting that fundamental differences in cultural socialization have contributed to the great disparity in the world's cross-country income distribution of today. In line with both the findings based on more recent periods and the historical analysis of Comin et al. (2010), the main transmission mechanism of entrepreneurship's long-run impact is the pace of technology adoption.

The present findings provide important policy implications for both developed and developing nations. Many political measures adopted to stimulate growth seek to incentivize new firm foundation, but the effects of such policies may vary across countries. The strong growth effect of entrepreneurial activity found for advanced economies suggests that establishment of an entrepreneurship-friendly environment

[18] Intuitively, the technique first approximates the small sample bias of the within-group estimator (Nickell, 1981) and eliminates this bias using an initially consistent estimator. While the recent literature has introduced a number of estimators that make use of the Kiviet (1995) approach to approximate the bias in the "small T" context (see, e.g., Bun and Kivet, 2003), the Bruno (2005a,b) estimator is the first technique that is also feasible for application in unbalanced panels.

may be an effective strategy to boost growth in richer countries. On the contrary, the results also imply that countries at earlier stages in the development process would benefit more from an investment-based strategy. These conclusions are in line with Acemoglu et al. (2006), who stress that the optimal growth strategy depends upon the development process. In addition, the results also provide support for the Unified Growth Theory in documenting that the primary driver of growth changes over the development process (Galor, 2011).

As stressed by Gorodnichenko and Roland (2017), studying the long-run effects of cultural differences that are deeply rooted in history may contribute to a better understanding of the substantial and slowly changing income disparities in the contemporary world. However, by no means should these results be misinterpreted as implying a "ranking" of cultures. Rather, they should contribute to the achievement of an improved dialogue between cultures and facilitate a departure from "one-size-fits-all" economic development policies.

2.A Appendix of Chapter (2)

Table A2-1 Descriptive statistics of the variables used in the estimations.

	N	mean	std.	min	max
	Panel A: Variables used for cross-sectional regressions				
$y_{i,2005-2009}$	166	8.7796	1.3141	5.4516	11.518
$y_{i,1500}$	31	6.3444	0.2605	5.9915	7.0031
$y_{i,1600}$	28	6.4590	0.35205	5.9914	7.2302
$y_{i,1700}$	31	6.5511	0.39316	5.9915	7.6634
$y_{i,1820}$	58	6.5465	0.37795	5.9840	7.5164
$y_{i,1900}$	50	7.3014	0.63270	6.2785	8.4101
Investment	143	0.2047	0.12754	0.0196	0.7539
Years of schooling	144	4.3803	2.64517	0.06	11.33
Self-employment	131	23.9782	15.46301	0.54	66.50
TEA index	70	11.1404	6.99808	3.10	33.70
Mastery (Schwartz)	69	3.9172	0.15097	3.54	4.25
Blood distance (Mahalanobis)	156	1.7455	0.80813	0	3.59
Blood distance (Euclidean)	156	0.1012	0.04857	0	0.21
Pronoun drop	39	0.5385	0.50504	0	1
Genetic distance	156	152.994	152.914	0	462
Historical pathogens	97	-0.0040	0.64932	-1.31	1.16
5HTTLPR	46	47.8354	12.96262	17.28	80.25
A118G	36	19.5335	10.8504	0.0304	46.478
TFP (Hall and Jones)	116	-0.9134	0.7056	-2.538	0.188
Innovation performance index	76	6.2371	2.1617	1.44	10
TFP growth	121	1.4986	2.5972	-6.8015	11.95
	Panel B: Variables used for panel estimations				
$\dot{y}$	1,804	0.0194	0.04290	-0.34337	0.40893
y	1,806	8.3576	1.30421	5.24302	11.8227
Years of schooling	1,587	5.4892	3.18633	0	13.27
Investment	1,806	23.115	10.87358	1.46107	76.5114
Log life expectancy	2,027	4.12742	0.19950	3.0809	4.4212
Log fertility rate	2,022	1.27487	0.55089	-0.1310	2.2134
Government consumption	1,806	12.1062	9.03880	0.41203	65.4833
Self-employment	629	22.4740	14.9231	0.22	77.7
TEA index	178	11.4145	8.3795	1.975	52.1
Mastery (Schwartz)	748	3.9156	0.1504	3.54	4.25

Notes: Table reports the number of observations (N), the means, standard deviations (std.), minima (min.), and maxima (max.) of the variables used in the empirical specification as described in Section (2.4).

Table A2-2 Entrepreneurship, culture, and economic growth — Evidence from Lewbel (2012) estimates.

	Lewbel Instruments			Lewbel Instruments and Blood Distance		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A-1: Self-employment rate					
Self-Employment	-0.0107	-0.227***	-0.212***	-0.0228**	-0.260***	-0.214***
	(0.0131)	(0.0228)	(0.0215)	(0.0112)	(0.0246)	(0.0212)
Self × GDP		0.0250***	0.0239***		0.0286***	0.0240***
		(0.00266)	(0.00214)		(0.00290)	(0.00214)
Countries	113	113	86	113	113	86
R-Squared	0.50	0.81	0.86	0.52	0.82	0.86
F-Stat (second)	19.91***	52.27***	70.85***	20.90***	54.66***	69.49***
F-Stat (first)	3.16**	4.28***	10.65***	7.60***	5.20***	9.97***
AR-Test	8.21*	30.69***	125.32***	26.57***	91.28***	146.24***
	Panel B-1: Total Entrepreneurial Activity					
TEA	-0.0720***	-0.393***	-0.264***	-0.0784***	-0.393***	-0.260***
	(0.0118)	(0.0914)	(0.0626)	(0.0123)	(0.0913)	(0.0624)
TEA × GDP		0.0420***	0.0276***		0.0421***	0.0278***
		(0.0107)	(0.0074)		(0.0108)	(0.0062)
Countries	65	65	54	65	65	54
R-Squared	0.72	0.82	0.91	0.71	0.82	0.91
F-Stat (second)	132.1***	43.22***	91.56***	97.23***	43.22***	91.56***
F-Stat (first)	19.57***	1126.00***	8326.00***	28.55**	83.25***	370.26***
AR-Test	31.09***	360.04***	147.23***	47.28***	363.68***	250.97***
	Panel C-1: Culture (Schwartz): Mastering					
Culture	5.975	-2.075***	-2.156***	8.915**	-2.072***	-2.165***
	(4.793)	(0.115)	(0.0996)	(3.565)	(0.116)	(0.0895)
Culture × GDP		0.254***	0.253***		0.251***	0.250***
		(0.00420)	(0.00666)		(0.00235)	(0.00328)
Countries	68	68	53	68	68	53
R-Squared	0.50	0.97	0.97	0.52	0.98	0.97
F-Stat (second)	6.29***	2384.6***	3009.8***	3.94**	3668.8***	3134.6***
F-Stat (first)	0.51	3.79***	6.91***	8.23***	11.73***	111.97***
AR-Test	13.97***	25.09***	66.35***	46.70***	94.43***	208.62***

Notes: Dependent variable is the log of real per capita GDP in the 2005–2009 period. Panel A reports the results based on the self-employment rate, Panel B employs the Total Entrepreneurial Activity (TEA) index, and Panel C draws on Schwartz's cultural dimension measuring mastering. The estimation technique replicates the Lewbel (2012) approach, with robust standard errors reported in parentheses. Instrumental variables are constructed as functions of the model's data. The table reports the results obtained by this approach and a separate version where this approach is supplemented by the Mahalanobis distance of blood types A and B relative to the United Kingdom, as used in Table (2.1). F-Stat (first) gives the F-statistic of the first-stage and the corresponding significance level. AR-Test gives the Anderson-Rubin Wald test to conduct weak-instrument-robust inference. $*p < .10$, $**p < .05$, $***p < .01$

Table A2-3 Entrepreneurship, culture, and economic growth — Panel estimations, alternative estimation techniques.

	Self-Employment		TEA index		Culture (Schwartz)	
	(1)	(2)	(1)	(2)	(1)	(2)
	Panel A: Within-Group Estimates					
Log(GDP$_{pc}$)$(t-1)$	-0.0739***	-0.0726***	-0.136***	-0.114***	-0.131***	-0.118***
	(0.0142)	(0.00847)	(0.0468)	(0.0255)	(0.0133)	(0.0131)
Entrepreneur	-0.00391**	-0.00244*	-0.0336***	-0.0167**	–	–
	(0.00184)	(0.00136)	(0.00968)	(0.00707)		
Entrepreneur$\times$GDP	0.000473**	0.000271*	0.00386***	0.00200**	0.0238***	0.0201***
	(0.000218)	(0.000159)	(0.00110)	(0.000795)	(0.00374)	(0.00372)
Observations	624	547	178	168	602	554
Countries	166	134	82	75	108	94
R-Squared	0.19	0.29	0.25	0.34	0.55	0.55
F-Stat	7.706***	11.41***	3.784***	5.856***	28.10***	42.74***
	Panel B: Bias-Corrected LSDV (Bruno, 2005a,b)					
Log(GDP$_{pc}$)$(t-1)$	-0.0173***	-0.0536***	-0.0541*	-0.0394**	-0.0568	-0.0203
	(0.00377)	(0.0129)	(0.0318)	(0.0182)	(0.0481)	(0.0170)
Entrepreneur	-0.0027***	-0.00140	-0.0302*	-0.0102*	–	–
	(0.0009)	(0.00237)	(0.0171)	(0.00602)		
Entrepreneur$\times$GDP	0.00034***	0.000141	0.00342*	0.00124**	0.0297***	0.0230***
	(0.00011)	(0.000267)	(0.00196)	(0.000618)	(0.0008)	(0.00251
Observations	624	547	178	168	601	553
Countries	166	134	82	75	108	94

Notes: Dependent variable is real per capita GDP growth. Panel A reports the results based on the traditional within-group estimator, while Panel B displays the outcomes when a LSDV bias-corrected version of the estimator is used (see Bruno (2005a,b) for a detailed description). The bias-corrected LSDV estimator uses the Blundell and Bond (1998) strategy as a consistent estimator to initialize the bias correction. Robust standard errors are obtained with 100 bootstrapping iterations. The columns labeled (1) and (2) display the results for (1) the reduced specification of Table (2.5) and (2) the comprehensive specification of the panel baseline table that controls for a set of growth determinants. As the mastery criterion from Schwartz (2006) is time-invariant per definition, the within approach neglects the linear effect. Inclusion of the interaction term, however, is computationally feasible, as the interaction term Mastery$\times$GDP is time-variant. $*p < .10$, $**\,p < .05$, $***p < .01$

Chapter 3
The Effect of the Financial Sector on Economic Development

Background The previous chapter highlighted a robust positive effect of entrepreneurship and its cultural traits on growth. This effect manifests itself via a positive stimulus on innovation activity, which accelerates long-run economic growth. However, the wealth of entrepreneurs may often not be sufficient to bear the cost for the projects themselves. For this reason, the financial sector may exert positive growth stimuli as it enables funding of projects. Alongside their effects on innovation and physical capital investment, funds provided by financial institutions may further help to increase education and health, and contribute to a reduction in the fertility rate.

The analysis in this chapter starts with the introduction of the financial sector into the model framework from Chapter (2), which implies a number of potentially positive growth effects of the financial sector, but also points to circumstances in which finance may be an impediment to growth. The latter finding is in accordance with some of the more recent studies conducted to explore the finance-growth nexus that find a negative effect of the financial system on income increases. This result is referred to as the *"vanishing effect of finance"* (see, e.g., Arcand et al., 2015). At first glance, this effect seems to be paradoxical, and the underlying reasons for why such an average negative effect may occur are still poorly understood. This chapter attempts to empirically explain the negative influence of financial institutions on economic well-being found in the recent literature. The results show that this effect is mainly driven by advanced economies and more recent periods, whereas finance was beneficial for income increases in earlier periods and still fosters growth in developing countries. The reason is that the transmission channels through which financial developments is transmitted to growth change over the course of economic development.

3.1 Introduction

There are very few elementary laws which economists regard as universally true. Among this handful of fundamentals, the belief in the positive impact of the financial sector on a nation's wealth occupies a prominent position. This wisdom was derived by numerous theoretical and empirical studies over the past 140 years (Bagehot, 1873; Goldsmith, 1969; King and Levine, 1993a,b; Levine, 1997; Beck et al., 2000).

Experience, however, is the father of wisdom, and after the experience of the Financial Crises of 2007–2008, the view that the financial sector benefits real economic activity came under increasing scrutiny (Schularick and Taylor, 2012). When re-examining the traditional ideas, Rousseau and Wachtel (2011) found that the growth effect of finance has considerably weakened over time. Utilizing data from the time period up to 2004, their results highlight a positive impact of the financial sector during the period from 1960–1989 which vanishes during the post-1990 period and eventually becomes (slightly) negative. Similarly, Arcand et al. (2015) conclude that application of more recent data reveals a weaker relationship between financial development and economic growth. Moreover, the authors provide evidence for the existence of a threshold beyond which the marginal effect of financial development becomes negative. In their—somewhat provocative—title, Rousseau and Wachtel (2011) ask *"What is happening to the impact of financial deepening on economic growth?"*. This chapter seeks to provide an answer to that question.

The key argument advanced in this chapter is that an understanding of the growth effect of financial development requires examination of the transmission mechanisms through which it is channeled to growth. While almost all cross-sectional or panel data studies focus on a direct link running from financial development to growth, very few investigations take into account the channels by which financial deepening is translated into income increases (Levine, 2005). These examinations typically model a linear relationship between financial development and potential transmission channels, reaching a tentative consensus that the financial system fostered fixed investment and productivity in the post-1960s period (King and Levine, 1993a; Beck et al., 2000; Benhabib and Spiegel, 2000) and over the past 140 years (Madsen and Ang, 2016).[19] However, what is common to all of these studies is that they neglect potential conditionalities in the transmission mechanism.

Such conditionalities directly follow from unified growth theory, which stresses that the primary engine of economic growth changes throughout the development process (Galor and Moav, 2004; Galor, 2009). In brief, this theory suggests that investment in physical capital and education are prime sources of growth at earlier stages of development, but their key role is increasingly superseded by that of factor productivity gains as economies become wealthier and credit constraints become less binding. As a result, we may expect changes in the way financial development transmits to growth in the course of the development process and—given the substantial increase in incomes during the past decades—over time.

The analysis of this chapter employs two strategies to identify such changes, utilizing country-year observations drawn from a panel of 129 countries over the past 50 years as well as those from a historical dataset featuring observations from 21 OECD countries between 1870 and 2009. This two-part strategy is motivated by the "rare events" problem (Schularick and Taylor, 2012) and the potential risk

[19] There exists, however, some dispute with regard to the nexus between financial development and savings/investments, both theoretically and empirically (see, e.g., Jappelli and Pagano, 1994; Bandiera and Caprio, 2000).

that the identified effects may stem from country-specific characteristics (Benhabib and Spiegel, 2000). The results from both datasets provide strong support for the conditionality hypothesis, suggesting that financial development was conducive to growth in earlier periods and that this effect has largely disappeared over time and space. The explanation for this "vanishing effect of finance" is that financial development contributes to an increase in physical capital investment and education, and to a reduction in the fertility rate in poorer economies, but these stimulating effects disappear during the development process. In advanced economies, the financial sector primarily fosters productivity gains, leading to a dependency of the financial sector's growth effect on new ideas and potentials for innovation projects.

Since the early 2000s, factor productivity has experienced a severe downward trend in almost all advanced economies, resulting in a global decline in growth rates that has led to the renaissance of the theory of "secular stagnation". Modern theories of secular stagnation emphasize that this phenomenon is primarily caused by faltering innovation activity and a lack of fundamentally new ideas (Gordon, 2012, 2015). As a result, the main transmission channel of the financial sector in rich economies has become increasingly weaker over the past 15 years, reducing the growth effect of financial intermediaries substantially over time.

Taken together, these findings explain why the positive effect of financial deepening on growth has disappeared in the more recent data, while earlier studies found a significant and robust growth effect of the financial sector. The results also imply that financial development may once again be beneficial for growth in the future, provided that technological progress once more permits a sufficient number of innovation projects. However, during phases when technological progress stagnates, forces identified by theories that stress a negative contribution of financial deepening to real-economic activity (Rajan, 2005; Schularick and Taylor, 2012; Philippon and Reshef, 2013; Beck et al., 2014b) may exceed its positive stimuli via intermediation activity, with the results that the overall effect becomes negative.

The remainder of this chapter is organized as follows. Section (3.2) reviews theoretical arguments linking financial development to growth and discusses the main transmission mechanisms of this process. Section (3.3) provides a theoretical model that augments the framework of Chapter (2) and consolidates the key arguments raised in the recent literature on the growth effect of financial institutions. Section (3.4) describes the empirical strategy and the two datasets employed in the analysis. In brief, the estimation strategy uses legal origin (large panel) and agricultural share (historical data) as instruments. Due to a growing concern about weak-instrumentation in the empirical growth literature (Bazzi and Clemens, 2013), the results are reported along with extensive weak instrument diagnostics including tests for weak-instrumentation, weak-instrument robust confidence intervals and rejection probabilities, as well as evaluations of the exclusion restriction. In addition, the outcomes obtained via the application of external instruments are compared with regressions using system GMM and bias-corrected LSDV. Sections (3.5) and (3.6) present the results based on the current and the historical panel of countries. Finally, Section (3.7) concludes.

3.2 Finance and growth in the recent literature

3.2.1 The positive effect of financial development

The first empirical evidence on the finance-growth nexus dates back to Goldsmith (1969) and McKinnon (1973), who found positive correlations between the size of the financial sector and long-run economic growth. However, these descriptive analyses suffered from two essential statistical shortcomings, namely the neglect of controls for other growth-inducing factors and the inability to indicate a causal relationship at work between the variables. The first of these shortcomings was resolved by King and Levine (1993a,b) and Levine (1997), who studied the effect of financial development (henceforth FD) on income increases in a comprehensive empirical model that includes a number of growth determinants. While these investigations confirmed the robust positive effect of FD, the concern of reverse causation persisted. In fact, a prominent line of reasoning initially put forward by Robinson (1952) argues that *“where enterprise leads, finance follows”*, suggesting that financial sectors develop in response to the demand created by the real economy. To tackle this second issue, Levine et al. (2000) employed national legal origin as an external instrument for FD, following the influential papers of La Porta et al. (1997), La Porta et al. (1998), and Beck et al. (2003) that stress a strong link between law and finance. An alternative strategy to eliminate endogeneity was used by Beck et al. (2000), who relied on lagged regressors as internal instruments in panel data estimations. A number of subsequent studies may be seen as variations on the econometric theme established by these papers, concentrating on different aspects of the finance-growth nexus. For instance, Beck and Levine (2004) find a positive effect of both stock markets and banks on economic growth based on panel data approaches, Guiso et al. (2004) show that local FD promotes growth by favoring business creation, market entry, and competition, and Rancière et al. (2008) demonstrate that financial liberalization encourages systemic risk taking, thereby increasing investment and, eventually, economic growth.[20]

3.2.2 The “vanishing effect” of financial development and a new debate

In contrast to the empirical studies conducted during the 1990s and the early 2000s, more recent examinations cast some doubt on the positive effect of FD on economic growth. When re-examining traditional ideas in the wake of the Financial Crisis, Rousseau and Wachtel (2011) find that the influence of finance has considerably weakened over time. Utilizing data from the time period up to 2004, their results highlight a positive impact of finance during the period from 1960–1989 which vanishes during the post-1990 period and eventually becomes (slightly) negative. To provide an explanation for this *“vanishing effect“* of FD, Arcand et al. (2012, 2015) suggest that FD and growth are linked via a non-linear relationship. While initially

[20] For more extensive surveys of the existing literature regarding the growth effect of FD, see Levine (2005), Beck (2008), and Panizza (2013).

providing strong growth effects, FD begins to have a negative impact on living standards once credit to the private sector reaches a critical threshold of 80–100% of GDP. These results were later replicated by a large number of empirical studies (see, e.g., Law and Singh, 2014 and Aizenman et al., 2015).

The recent results have fueled a new debate on the long-run impact of FD on living standards. This debate has been intensified by the findings of Henderson et al. (2013) and Madsen and Ang (2016), who find new evidence for a growth-enhancing effect of FD. However, these seemingly contradictory findings can be traced to the utilized data sample: Henderson et al. (2013) use data from 1960 to 2000, thereby neglecting the past 15 years, during which period the negative effect of FD has intensified. This resembles shearing off the "right-hand side" of the FD-growth parabola. Meanwhile, Madsen and Ang (2016) investigate the effect of finance over a very long time-span based on historical data. Compared with today, however, financial sectors were substantially less developed in the late 19th century. Thus, this study emphasizes the "left-hand side" of the FD-growth parabola.

3.2.3 Possible explanations for the new results

There are several explanations regarding the underlying forces which have contributed to the vanishing effect of finance. First, a prominent argument in the literature stresses that FD is a predictor of both banking and currency crises (see Reinhart and Kaminsky, 1999; Schularick and Taylor, 2012). In an attempt to reconcile these results with the traditional view of a strictly positive effect of FD, Loayza and Ranciere (2006) find that the positive long-run stimuli of FD coexist with some negative short-run effects that are driven by financial crises. Closely related is the argument of Easterly et al. (2000), who show that FD increases output volatility once credit to the private sector has reached a sophisticated size of around 100% of GDP. More provocatively, Rajan (2005) argues that the presence of large and complicated financial sectors increases the potential for a "catastrophic meltdown".

The prospect of efficiency wages may also yield suboptimal allocation of talent, preventing individuals from investing in tertiary education and shifting talented workers from the productive sector into the financial sector (see, e.g., Philippon, 2010, Bolton et al., 2011, Philippon and Reshef, 2013 and Kneer, 2013).

Although not directly addressed in their article, an alternative way to look at the new results has been provided by Demirgüç-Kunt et al. (2013), who illustrate that as economies develop, the marginal impact of bank activities on economic output decreases and financial systems become more market-based. While this finding is consistent with theories arguing that services provided by banks become less important when economies become richer (Allen and Gale, 2000), empirical research has been largely unsuccessful at clarifying the reasons of this change.

Finally, the literature has stressed that the declining importance of intermediation activity is accompanied by ancillary effects that are detrimental to GDP growth: first, decomposing bank lending into enterprise credit and household credit, Beck et al. (2012) demonstrate that lending to households does not trigger any growth effects. In contrast, excessive household lending that gives rise to real estate lending booms exhibits the potential to feed speculative bubbles and macroeconomic recessions (Jordà et al., 2016). Second, Demirgüç-Kunt and Huizinga (2010) argue that during the past decades banks have gradually steered away from their traditional intermediation activities. In the 1990s and the 2000s, non-interest incomes of banks experienced a substantial increase, particularly via the trading of mortgage-backed securities. Rather than triggering growth effects (Beck et al., 2014b), this shift in banking activities increases systemic risk (Stiroh, 2004; Brunnermeier et al., 2012), thereby hampering economic growth (Kroszner et al., 2007).

3.2.4 Transmission channels of financial development to growth

In order to put the previously discussed negative effects of the financial sector into context, it is crucial to study the mechanisms through which FD is transmitted to growth. However, the feature common to the great majority of the empirical studies that aim to link FD to economic development is their focus on reduced-form evidence. This shortcoming has been critcised by Levine (2005), who surveys theories that suggest positive effects of FD on investment via provision of ex-ante information and ex-post monitoring, facilitation of risk-management and diversification, pooling of savings, and supporting the exchange of goods and services.

A second channel of FD identified in the literature is schooling. This literature argues that borrowing constraints lead to under-investment in human capital (Becker, 1960), as financially constrained households are distracted from investing in their children's education due to a lack of the sufficient financial resources necessary for school fees and living costs (Lochner and Monge-Naranjo, 2012). As education works as a major engine of growth (Barro, 2001, 2013a), such a waste of intellectual potential hampers economic development and is greater if the distribution of incomes is unequal (Galor and Zeira, 1993).

Closely related to the education channel, the old age security hypothesis argues that parents decide to have more offspring to ensure that they will receive financial support from their children in old age. Based on data from U.S. counties in the 19th century, Basso et al. (2014) show that more developed financial markets reduce the fertility rate, resolving the issue of families' trade-off between the "quantity and quality of children" (Becker and Lewis, 1973) in favor of having less but better educated children. An additional transmission channel of FD is imbedded in this mechanism, as high fertility rates have been identified as a main obstacle to growth in the empirical literature (see, e.g., Barro, 2003). When analyzing the long-run impact of FD on fertility, this argument further suggests that FD expedites "demographic transition", which plays a central role for the transition from stagnation to growth (Cervellati and Sunde, 2011).

From a theoretical perspective, both investment and human capital possess decreasing marginal returns, which may help to explain the finding of Demirgüç-Kunt et al. (2013) that the growth-enhancing effect of banking activity decreases during the development process. Similarly, De Gregorio and Guidotti (1995) and Huang and Lin (2009) find that FD significantly accelerates growth in low- and middle-income countries, whereas the effect is less pronounced in high-income economies. In line with these results, Aghion et al. (2005) show that FD helps to accelerate convergence.

While these hypotheses necessarily imply that the effect of FD will decrease over the development process, there is a prominent theoretical argument stressing that FD generates permanent growth effects via an increase in knowledge and productivity (Morales, 2008; Buera et al., 2011) which has found support in the more recent empirical literature (Ang, 2011; Ang and Madsen, 2012). This literature stresses that FD increases the ability of firms to innovate and to keep up with advances of the technological frontier (Gorodnichenko and Schnitzer, 2013), and also facilitates the adoption and improvement of new products and (production) processes (Aghion et al., 2005).

While the recent literature called attention to a number of potential transmission channels of FD, very few studies address these channels empirically. Based on cross-sectional data from 57 countries averaged over the period between 1960 and 1989, King and Levine (1993a) show that FD translates to growth through physical capital accumulation and factor productivity. A similar result is reported by Benhabib and Spiegel (2000), who employ data from the 1965–1985 period. In contrast, Beck et al. (2000) and Bandiera and Caprio (2000) find an ambiguous relation between FD and physical capital investment when using data from the 1971–1995 period. More recently, the study of Madsen and Ang (2016) analyzes the transmission channels of FD based on a historical dataset that spans from 1870 to 2009, suggesting that FD stimulates investment, savings, education, and R&D. Common to all of these studies is that they analyze a linear effect running from FD to the transmission channels, ignoring potential conditionalities that may be triggered by the development process.

These conditionalities, however, are crucial to understanding the driving forces behind the “vanishing effect of finance”. This is because the recent growth literature stresses that the main driver of growth changes as economies develop (Galor and Moav, 2004; Galor, 2009). In earlier stages of development, fixed investment and education are prime sources of income increases, which is why FD may be beneficial to growth via its traditional intermediation activity. This effect becomes increasingly irrelevant during the development process for two reasons: first, diminishing returns to reproducible factors yield a decline in the contribution of education and investment to growth. Second, as average incomes rise, the budget constraints of the households become less binding. Consequently, it follows that technological change becomes the primary source of growth in later stages of development. Given this changes in the key driver of growth, it can be hypothesized that the growth potential of FD in advanced economies depends on the ability of countries to produce new ideas: while FD fosters economic development in advanced economies if technological progress is growing rapidly, it provides little growth prospects in times when the potential for promising innovation projects is low. In the latter situation, the previously discussed negative effects of FD may offset its positive contribution, which causes the overall effect to become negative. The conditionality in FD’s impact on the transmission channels over the development process may further explain why studies assuming a linear link between FD and investment find ambiguous results.

3.3 Theoretical framework: The financial sector and economic growth

This section provides an extension to the theoretical model described in Chapter (2) by including the financial sector into the theoretical framework. As the model aims to shed light on the transmission channels of finance on growth, the framework deviates in some details from the version of Chapter (2). By consolidating a number of arguments discussed in Section (3.2) into a simple model of finance and growth, the theory developed in this section lays the foundation for the subsequent empirical work.

3.3.1 The basic model

Considering a continuum of specialized intermediate goods $j \in \mathbb{J}$, the output y_e of firm e can be formulated similarly to Romer (1987, 1990) as[21]

$$y_e = \Psi_e(\kappa_e HK)^{1-\alpha} \int_{\mathbb{R}_+} x_{ej}^{\alpha} dj, \tag{3.1}$$

where Ψ denotes factor productivity, κ_e gives the fraction of human capital HK employed by e and x_{ej}^{α} is the amount x of the intermediate good j used in the production process of e. As each j features diminishing marginal returns, it follows that $\alpha \in (0, 1)$. The term $\int_{\mathbb{R}_+} x_{ej}^{\alpha} dj$ reflects the stock of physical capital.[22] The total amount of human capital employed in the production sector is $K * HK$ where $K = \sum_e \kappa_e$. Equation (3.1) illustrates the production potential if all intermediate goods have been invented. Meanwhile, in each period, there is only a finite number $|\mathbb{J}|$ available in the production process.

Suppose that $|\mathbb{J}| = N$ gives the range or number of capital goods used and let Γ be the total quantity of these inputs.[23] If all firms are equal and $\{\Gamma, N\}$ denotes the list of x_i with constant value $x_i = \Gamma/N$, then Equation (3.1) becomes

$$y_e = \Psi_e(\kappa_e HK)^{1-\alpha} N^{1-\alpha} \Gamma^{\alpha}. \tag{3.2}$$

In this case, output increases with N when holding constant productivity, labor and Γ. Inventions thus boost economic growth as they lead to an increase in the stock of physical capital. Meanwhile, inventions also enhance factor productivity, as a by-product of the invention process is the creation of new knowledge. This knowledge eventually diffuses to competitors, but initially provides an advantage to the inventor. Aghion and Howitt (2009) capture this effect, defining the starting technology of e as

21 See Section (1.2) for an introductory description of this class of models.

22 In general, it is reasonable to use any increasing, strictly concave function $g(x)$ with $g(0) = 0$ to model firm output, depending on capital goods utilization. The special case considered here, however, is analogous to the power function of Dixit and Stiglitz (1977) and assumes the form $g(x) = x^{\alpha}$.

23 Note that this denotation deviates from the original Romer (1987) paper. The definition of N used here refers to Barro and Sala-i-Martín (2004).

$$\Psi_{t-1} = E^{-1} \sum_{e=1}^{E} \Psi_{e,t-1}, \; e = 1, ..., E \,.$$

Therefore, each non-innovating e presumably has the average productivity level of all entrepreneurs in $t-1$, that is $\Psi_{e,t} = \Psi_{t-1}$. Innovating firms, however, have access to $\Psi_{e,t} = \gamma_{j*}\Psi_{t-1}$ where γ_{j^*} represents the importance ("size") of the particular innovation j^*. It follows that $(\partial\Psi/\partial\gamma) > 0$ and $(\partial y/\partial\gamma) > 0$. Let μ denote the probability of an innovation of e and $\gamma = N^{-1}\sum_{j*=1}^{N^*}\gamma_{j^*}$, $j^* = 1, ..., N^*$ be the average size. Then the average productivity across all firms will be (see Section (1.2) for a detailed description)

$$\Psi_t = \mu\gamma\Psi_{t-1} + (1-\mu)\Psi_{t-1},$$

implying that the average productivity grows at a rate of

$$\frac{\dot{\Psi}}{\Psi} = \frac{\Psi_t - \Psi_{t-1}}{\Psi_{t-1}} = \mu(\gamma - 1)\,.$$

Innovating entrepreneurs benefit the economy through two channels. First, increasing numbers of N have a direct effect on physical capital in Equation (3.2). If one entrepreneur creates a new j^*, it can be inserted in the production process of all firms. Second, innovations create new knowledge. After some time, this knowledge becomes available to all firms, enhancing factor productivity and thus output in Equation (3.2).[24]

The innovation j^* makes the innovator a monopolist. Each entrepreneur thus has two incentives to innovate: first, the innovator earns monopolistic profits by selling j^*. As existing capital products can be provided by a range of entrepreneurs, producing j makes e_j a mere price-taker. Second, j^* enhances productivity of e, leading to more efficient production of all capital goods supplied by e_{j^*} and facilitating future innovations.

Households maximize utility over an infinite horizon and have CRRA preferences

$$U = \int_0^{\infty} \left(\frac{c^{1-\lambda} - 1}{1-\lambda}\right) \exp\{-\rho t\}\, \mathrm{d}t\,.$$

Household income is composed of wage w and capital income r on assets a, yielding the common aggregate budget constraint $(\mathrm{d}a/\mathrm{d}t) = wL + ra - c$ and the familiar Euler equation $(\dot{c}/c) = (1-\lambda)\,(r-\rho)$.

3.3.2 The investment decision of the entrepreneur

The decision of an entrepreneur to invest in physical capital or in innovation is determined by the costs and the risk of the project. The capital value earned by creating j^* is

[24] The application of $(d\Psi_e/dt)$, however, depends on the level of human capital available in e. See Nelson and Phelps (1966) and Benhabib and Spiegel (2005).

$$V(j^*) = \int_t^{t+\psi} \pi_{j^*}(v) * e^{-st} \mathrm{d}t,$$

where $\pi(v) \equiv (P_{j^*} - 1)x_{j^*}$ is the cashflow stream at any time $v \in [t, t+\psi]$ and e^{-st} denotes the discount factor.

Each e decides to carry out a project if the expected capital value exceeds the costs η of the particular investment, that is, the expected rent $\mathbb{E}[V(j^*)] - \eta$ is positive. Basically, any investment in j^* under certainty is profitable if $\eta < V(j^*)$. However, the expected capital value will be achieved with a probability $p_{j^*} \in (0,1)$, which includes a risk factor in the entrepreneur's calculation. The parameter p_{j^*} is inversely proportional to the inherent risk associated with j^*. This risk is primarily unknown and must be estimated by the entrepreneur. After assigning a probability of success to j^*, the investment decision becomes

$$V(j^*) > \frac{\eta}{p_{j^*}},$$

emphasizing that higher risk makes an investment in j^* less likely.

3.3.3 The role of financial intermediaries and the equilibrium growth rate

Financial intermediaries can exert a strong influence on the investment decisions of entrepreneurs. Usually, the entrepreneur will not be able to procure financing on his own. In most cases, it is likely, though not certain, that the entrepreneur's initial wealth will not be sufficient for him to cover η himself. Moreover, the model of King and Levine (1993b) indicates that the risk of innovation failure is diversifiable, which makes reliance on any amount of internal finance less efficient. In consequence, the entrepreneur will choose to borrow the funds necessary to finance η from the financial sector.

To decide whether the project shall be supported or rejected, a financial institution—as, for instance, a bank—can determine the feasibility of the investment by paying a cost of $f = f(\eta)$, $(\partial f/\partial \eta) > 0$, where Aghion and Howitt (2009) suggest a mutliplicative function $f = \phi\eta$. In order to break even, the bank must require a repayment of $\phi\eta/p$.[25] Thus, the total cost of the entrepreneur adjusts to $\eta(1 + \phi/p)$.

In equilibrium, it must hold that $\eta(1+\phi/p) = V(j^*)$ under the free-entry condition. Using the production function stated above, it is easy to show that π_{j^*} reaches its maximum at price $P_{j^*} = (1/\alpha) > 1$, yielding[26]

$$\pi = (K * HK)\Psi^{1/(1-\alpha)}\Omega\,. \tag{3.3}$$

[25] Let Π denote the repayment from a feasible innovation project. The expected profit from screening is $p\Pi - \phi\eta$ and thus $\Pi = \phi\eta/p$, as the payment of Π depends on the probability p, whereas the cost of screening will be $\phi\eta$ with certainty.

[26] One way to derive this condition can be found in Barro and Sala-i-Martín (2004). The parameter $\Omega \equiv \frac{1-\alpha}{\alpha}\alpha^{2/(1-\alpha)}$ is inserted for reasons of lucidity.

Applying the free-entry condition, using the maximum price and making use of the condition $\bar{r}(t, v) = [1/(v - t)] \int_t^v r(\xi)\mathrm{d}\xi$ gives

$$r(t) = \frac{\pi}{V(j^*)} + \frac{\dot{V}(j^*)}{V(j^*)},$$

which follows by utilization of Leibniz's rule for differentiation under the integral sign (see Barro and Sala-i-Martín, 2004). As η is a constant and p_{j*} depends on the inherent risk of the project that cannot change over time, it follows that $\dot{V}(j^*) = 0$, simplifying the interest rate to $r = \pi/[\eta(1 + \phi/p)]$.

Inserting π from Equation (3.3), the interest rate becomes

$$r = \frac{K * HK}{\eta(1 + \phi/p)} \Psi^{1/(1-\alpha)} \Omega \,.$$

Substituting the interest rate in the Euler equation gives the growth rate of the economy

$$\frac{\dot{y}}{y} = (1/\lambda) \left[\frac{K * HK}{\eta(1 + \phi/p)} \Psi^{1/(1-\alpha)} \Omega - \rho \right] . \tag{3.4}$$

Equation (3.4) highlights a number of transmission channels through which the financial sector influences economic growth. First, financial intermediaries enable the pooling of funds from small savers to mobilize sufficient resources to cover η. Second, the financial sector facilitates evaluation of investment projects, conducting ex ante screening to estimate p. Such an evaluation typically requires information about future cashflows and interest rates, as well as firm-specific information that is often difficult to acquire. In the majority of cases, individual savers will neither have the time, nor the means or capacity to carry out this assessment. As a result, the absence of a well-functioning financial system hinders the flow of capital to the most promising projects, as in this case the inherent risk of projects is likely to be over- or under-estimated. Whereas the former misjudgment yields a decline in investment projects and innovation activity, the latter may result in too many risky investments that are likely to default. In both cases growth will be reduced. Third, once p is assessed, financial intermediaries choose the path of caution. By supporting investments with high p and rejecting projects that are not feasible, bad investments are filtered out. If a project is supported, financial institutions may also provide means to diversify the inherent risk.

Finally, financial intermediaries possess extensive practical knowledge and experience in various fields. Advice from the financial sector may thus lead to an increase in p. The equation also shows that the higher the screening costs ϕ, the lower both the frequency of innovations and the growth rate will be. A larger financial sector may create economies of scale with regard to the assessment of projects, resulting in a decrease in ϕ that is beneficial to growth.

Equation (3.4) further illuminates the positive growth effect emanating from an increase in factor productivity. As discussed previously, factor productivity grows at rate $(\dot{\Psi}/\Psi) \equiv \mu(\gamma - 1)$, where the size of the innovation γ and the probability of its occurrence μ are crucial in determining productivity gains. An established financial system is more likely to be able to support innovations with large γ. In addition, through the various channels discussed above, the financial sector also contributes to an increase in μ.

While the previous mechanisms emphasize the role of finance in the creation of investment in physical capital as well as in innovation projects, Equation (3.4) highlights a further transmission channel, that of human capital. The model implies that a higher education level yields an increase in economic growth; however, binding household budget constraints may impede investment in education, particularly in the presence of capital market imperfections (see Galor and Zeira, 1993). Through the provision of the funds necessary for education and health investments, the financial sector fosters investment in human capital. This growth effect is particularly strong in economies with highly skewed income distributions (see Benhabib and Spiegel, 2000) and in developing countries with less sophisticated public sectors.

In fact, the growth rate illustrated in Equation (3.4) provides a number of implications in terms of the role of finance in the development process. In poor countries that have not yet approximated the steady state level of per worker capital, the financial system may accelerate conditional convergence by facilitating investment in physical capital (Aghion et al., 2005). In these countries, the absence of advanced financial markets hinders growth, as many investments fail to receive financing without a functioning banking system. This problem is particularly severe if domestic savings are not sufficient to cover the costs η of investment projects, resulting in the need to receive funds from international financial markets. As poor countries experience higher marginal effects of investment in physical capital on growth than advanced economies, the effect of finance may be particularly pronounced in developing economies. Meanwhile, diminishing marginal returns of education provide a further argument for why the effect of finance is likely to be stronger in poorer countries.

In contrast to the positive effects discussed so far, the model also reveals channels through which finance may impede growth. First, the model reflects the argument raised by the "prospect of efficiency wages" hypothesis (see, e.g., Philippon and Reshef, 2013; Kneer, 2013): if high wages paid by financial intermediaries attract an increasing fraction of human capital, K declines and hence causes a reduction in the growth rate. Moreover, efficiency wages may also prevent individuals from investing in tertiary education. A second channel is related to the Demirgüç-Kunt et al. (2013) argument: if the financial sector steers away from traditional intermediation and screening activites and increasingly focuses on non-interest income (e.g. via trading of mortgage-backed securities), an increase in the size of the financial sector will have no effect on the growth rate illustrated in Equation (3.4). Rather, this development may raise inflation and contribute to a higher vulnerability of countries to economic crises (see, e.g., Rajan, 2005; Demirgüç-Kunt and Huizinga, 2010). In such an event, unemployment and hysteresis reduces K, which directly affects the growth rate.

Finally, growth in advanced economies is based largely on technological progress and innovation activity. In times when below-average technological progress allows little potential for beneficial innovation projects—i.e. μ, p and γ are low—, there are only poor growth prospects for advanced economies. In this situation, the negative effects of the financial sector discussed in Section (3.2) may offset its positive contribution, resulting in the overall effect turning negative.

To summarize the theoretical implications, finance may benefit growth by accelerating conditional convergence and by facilitating human capital, investment, and technological progress. In accord with the Unified Growth Theory (see Section (1.2)) and the arguments discussed in Section (3.2.4), we might expect that the importance of these channels changes over the development process. In particular, the positive effects may ebb once incomes and reach more sophisticated levels. In this event, forces described by theories that stress a negative link between FD and growth may supersede the positive effect of FD on economic development.

3.4 Estimation strategy and the data

3.4.1 Specification and estimation technique

To empirically study the hypotheses drawn in Sections (3.2) and (3.3), the analysis follows two steps, specifically (1) estimation of the overall effect of FD on growth and (2) examination of the transmission channels through which this effect translates to income increases. The baseline empirical model assumes the log of GDP per capita (y_{it}) in country i in 5-year period $t, t = 1, \dots, T$ to be a function of the log of initial GDP per capita (y_{it-1}), the level of financial development (F_{it}), human capital (h_{it}), a range of control variables ($\mathbf{X}_{it}$), country- and time-specific fixed effects (η_i, ξ_t), and an idiosyncratic error term (v_{it}), yielding

$$y_{it} = \theta y_{it-1} + \gamma F_{it} + \lambda h_{it} + \boldsymbol{\beta}\mathbf{X}_{it} + \eta_i + \xi_t + v_{it}.^{27} \tag{3.5}$$

As emphasized by Beck et al. (2000), estimation of the impact of FD on long-run economic growth (γ) is impeded by unobserved country-specific effects and endogeneity issues. To tackle the first of theses challenges, Equation (3.5) includes country-specific, time-invariant effects to partial out different institutional environments at the country-level that may affect FD and its transmission to growth. In addition, the model captures time-specific effects to discard short-run economic

[27] As illustrated by Voitchovsky (2005), Bond et al. (2001), and Halter et al. (2014), this expression is derived via

$$\dot{y}_{it} = y_{it} - y_{it-1} = (\theta - 1)y_{it-1} + \gamma F_{it} + \lambda h_{it} + \boldsymbol{\beta}\mathbf{X}_{it} + \eta_i + \xi_t + v_{it}.$$

The selection of the covariates ($\mathbf{X}$) refers to Barro (2003, 2013a), whose framework has been shown to capture the empirical determinants of economic growth quite accurately. Collected from Feenstra et al. (2015), World Bank (2017), Freedom House (2017), and Barro and Lee (2013), these covariates include: the investment share, government consumption, the inflation rate, the degree of openness, political rights, and the log of the fertility rate. The political rights index from Freedom House is re-coded so that higher values are associated with greater political rights. In addition, the log of real per capita GDP in $t-1$ accounts for conditional convergence, and the stock of human capital is captured by the average years of schooling and the log of life expectancy at birth. We exclude measures of physical capital, as their calculation relies on arbitrary assumptions, and follow Barro (2003, 2013a), who assumes that higher levels of y_{it-1} and h_{it} reflect higher levels of capital endowment. Descriptive statistics can be found in Table (A3-1) in the appendix.

shocks, guaranteeing that the estimated parameters reflect the long-run impact of FD on economic development. Second, to rule out the possibility of reverse causation, Equation (3.5) is estimated via three different techniques, namely 2SLS, System GMM, and bias-corrected LSDV. Due to a growing concern about weak instruments in the recent empirical growth literature (Bazzi and Clemens, 2013), we seek to purge the endogenous components in the data via a two-part strategy, following the recent literature by making use of both external instruments (e.g. Levine et al., 2000) and internal instruments (e.g. Beck et al., 2000). As an additional robustness check, the analysis replicates strategies that rely on within-group transformations ("fixed-effects"), as used by, e.g., Rajan and Zingales (1998). However, in order to circumvent a potential dynamic panel bias documented by Nickell (1981), our strategy is based on a bias-corrected version of LSDV developed by Bruno (2005a,b).[28] Application of this estimator further helps to evaluate the econometric concerns recently raised with regard to the system GMM estimator (Murtin and Wacziarg, 2014).

To describe the applied estimation strategies more specifically, the 2SLS strategy uses the legal origin of a country as an external instrument L_{it} for FD, following the extensive literature on the entanglement of FD and law (La Porta et al., 1997, La Porta et al., 1998, and Beck et al., 2003). Specifically, the first-stage regression of Equation (3.5) follows the logic of Beck et al. (2003) via

$$F_{it} = \rho y_{it-1} + \xi \text{SOC}_{it} + \psi \text{ENG}_{it} + \zeta \text{FRE}_{it} + \tau h_{it} + \boldsymbol{\omega} \mathbf{X}_{it} + u_{it}, \tag{3.6}$$

where SOC, ENG, and FRE represent dummy-variables for Socialist-, English-, and French-civil-law countries, and the error term is illustrated compactly as u_{it}. Data on legal origin is collected from La Porta et al. (1998). Our instrumentation strategy needs to satisfy the usual conditions (Angrist and Pischke, 2009): (1) L_{it} must be highly correlated with the level of FD (F_{it}) for each country i and 5-year period t (*strong first stage*), and (2) L_{it} needs to be uncorrelated with any other (unobserved) variable affecting economic growth (*exclusion restriction*). The first requirement is likely to be met, a fact which is extensively discussed in the literature (see, e.g., Levine et al., 2000). A potential concern, however, may be that legal origin stimulates growth through channels other than FD, thus violating the exclusion restriction. To assess whether the inference is still informative if legal origin were only to be *"plausibly exogenous"*, the analysis routinely conducts the Union of Confidence Intervals (UCI) test of Conley et al. (2012) for each of the 2SLS estimates. Briefly put, in the standard IV setting

$$g_{it} = \gamma F_{it} + \varphi L_i + u_i,$$

[28] Intuitively, the technique first approximates the small sample bias of the within-group estimator and eliminates this bias using an initially consistent estimator. While the recent literature has introduced a number of estimators that make use of the Kiviet (1995) approach to approximate the bias in the "small T" context (see, e.g., Bun and Kivet, 2003), the Bruno (2005a,b) estimator is the first technique that is feasible for unbalanced panels.

the necessary assumption is that $\varphi = 0$. The UCI test deviates from that assumption by employing some $\varphi \neq 0$ and returning the union of all interval estimates of β conditional on a grid of all possible values for φ.[29]

Motivated by concerns about a potential violation of the exclusion restriction, the second strategy to identify causality employs internal instruments in system GMM estimations. Application of system GMM in the context of the FD-growth nexus is documented in detail in the empirical literature (see, e.g., Beck et al., 2000; Levine et al., 2000). For a brief overview, consider first the *difference GMM* estimator (Arellano and Bond, 1991), the basic idea of which is to adjust Equation (3.5) to

$$\nabla y_{it} = \theta \nabla y_{it-1} + \lambda \nabla h_{it} + \gamma \nabla F_{it} + \boldsymbol{\beta} \nabla \mathbf{X}_{it} + \nabla \xi_t + \nabla v_{it}{}^{30} \tag{3.7}$$

and to use sufficiently lagged values of y_{it-1}, h_{it}, F_{it}, and $\mathbf{X}_{it}$ as instruments for the first-differences. This strategy, however, features a number of drawbacks: first, difference GMM requires (at least) three consecutive lags for each observation. This requirement, however, yields an asynchronous loss of country-years since the availability of data for FD is typically more limited in less developed countries. Second, transformation of the data via Equation (3.7) results in removal of all level information in the explanatory variables, producing parameter estimates that only exploit the within-variation of the panel. With respect to the influence of FD, application of this technique means that any impact of finance on growth may only be captured if the degree of FD changes significantly over the sample period. Moreover, if for some extreme cases FD possesses a high degree of persistence (as is commonly the case in less developed economies), the lagged levels become weak instruments for the first-differences (Bond et al., 2001).

Arellano and Bover (1995) and Blundell and Bond (1998) show that some of these weaknesses can be addressed if the dependent variables satisfy a stationarity assumption.[31] In this case, the *system GMM* estimator expands the difference GMM framework by utilizing additional orthogonality conditions for the levels in Equation (3.5), employing lagged values of ∇k as instruments for k.[32]

---

[29] Following Persson and Tabellini (2009), we use a regression of growth rates on both FD and our instruments to obtain an estimate of the degree of the bias based on the 95% confidence interval, which serves as an estimate for the lower and upper bounds of φ.

[30] Hereafter, we denote $\nabla k \equiv (k_{it} - k_{it-1})$ for $k \in \{y_{it-1}, F_{it}, h_{it}, \mathbf{X}_{it}, \xi_t, v_{it}\}$.

[31] The assumption is $\mathrm{E}(\eta_i \nabla y_{i2}) = 0 \; \forall i$ and holds if the process is mean stationary: $y_{i1} = \eta_i/(1-\theta) + v_i$ with $\mathrm{E}(v_i) = \mathrm{E}(v_i \eta_i) = 0 \forall \; i$ (Arellano and Bover, 1995). To test this assumption, Difference-in-Hansen tests are routinely reported for all regressions, as suggested by (Roodman, 2009b,c).

[32] The additional moment conditions of system GMM are given by

$$\mathrm{E}[(v_{it} + \eta_i)(\boldsymbol{\Theta}_{it-1} - \boldsymbol{\Theta}_{it-s})] = 0 \text{ for } t \geq 3, \tag{3.8}$$

where the exact choice of $s = 2, \ldots, (t-1)$ refers to the researcher's intended focus. A well-known problem in specifying the instrument matrix is that of "instrument proliferation", which can be countered by either restricting the lag structure or collapsing the (fully-)specified instrument matrix (Roodman, 2009c). This analysis follows a dual procedure, instrumenting FD and the initial development level using a collapsed matrix of all available lags, while all other covariates are instrumented with the second lag. This strategy is motivated by the consensus in the literature that multiple periods are often required for FD to be cultivated and to translate to growth. Additionally, conditional convergence often occurs over several periods. The implemented instrumentation strategy is carefully chosen in accordance with Hansen's J-test of over-identifying restrictions.

It has been shown that System GMM possesses better finite sample properties than Difference GMM (Blundell and Bond, 1998 and Blundell et al., 2000), and that its properties depend crucially on (1) the extent of endogeneity and (2) on the strength of instrumentation (see Bazzi and Clemens, 2013). With respect to (1), the empirical literature clearly emphasizes the role of endogeneity in the FD-growth context (see Beck et al., 2000; Levine, 2005). To assess requirement (2), we replicate the weak instrument tests suggested by Bazzi and Clemens (2013) for each estimation to evaluate the strength of the utilized internal instruments. For the computation of the GMM estimates, the analysis is built on a two-step procedure that weights the moment conditions by a consistent estimate of their covariance matrix, which has proven to be asymptotically more efficient (Bond et al., 2001). To circumvent the problem of downward-biased standard errors in this setting, the Windmeijer (2005) correction is employed to obtain a more accurate inference.

The critique recently leveled against the system GMM estimator is directed at the problems involved with the choice of the number of instruments and lags, as well as at the stationarity condition (see Murtin and Wacziarg, 2014). In particular, if FD is characterized by differential convergence and if these different processes are systematically related to the country-specific effect, then the validity of the stationarity condition may be called into question due to a potential correlation between the country-specific effect and the lagged differences of FD. This problem has led to a renaissance of within-group estimations in the very recent empirical growth literature (see Compton et al., 2011; Murtin and Wacziarg, 2014; Afonso and Jalles, 2014). However, due to the mechanical correlation of the lagged dependent variable and the error-term, traditional fixed-effects estimators typically suffer from a dynamic panel bias in the small T context. For this reason, the third estimation strategy follows the recent literature by employing the bias-correction proposed by Bun and Kivet (2003) and Bruno (2005a,b). Intuitively, this estimator approximates the bias emerging from the fixed panel length and eliminates it based upon an initially consistent estimator.

The analysis of the transmission channels of FD replicates the strategy employed to obtain the reduced-form results. To asses the effect of the transmission mechanisms over the development process and over time, the analyses follow the general logic of Aghion et al. (2005) by inserting interaction terms between FD and per capita GDP, as well as the time variable. These inteacrtion terms are denoted with FD $\times$ GDP and FD $\times$ Time, respectively.

3.4.2 Data and samples: Two strategies to identify the effects of FD

Several measures can be used to obtain proxies for the level of FD. A broad concept of FD typically aims at capturing the financial sector's overall ability to reduce a wide range of costs associated with the intertemporal nature of financial contracts, such as information costs, transaction costs, and enforcement costs (De la Torre et al., 2013; Arcand et al., 2015). This study, however, draws on a much narrower concept using credit to the private sector as a share of GDP (henceforth denoted with CREDIT). There are two reasons for this: First, in contrast to a broader definition, credit to the private sector is more likely to correspond with the growth-augmenting features of

FD summarized in Levine (2005) and with the transmission channels identified in the previous section. Second, while first used by King and Levine (1993a), it has now become the most commonly employed indicator of FD in both the seminal papers on the FD-growth nexus written in the early 2000s (Levine et al., 2000; Aghion et al., 2005) and in more recent studies on the topic (Aghion et al., 2010; Laeven et al., 2015; Arcand et al., 2015).[33] Descriptive statistics are reported in Table (A3-1) in the appendix.

To identify the effects of FD on both GDP growth and the transmission channels, the analysis is based on two different data strategies: The first data set uses credit to the private sector from World Bank (2017), constructing a panel that includes 129 countries between 1960 and 2014 (Dataset 1). Observing a large number of countries during the past five decades enables investigation of the transmission process of FD to growth over the development process and over time. The second data set uses historical data, spanning the period 1870–2009 and including observations for 21 OECD countries taken from Madsen and Ang (2016) (Dataset 2).[34] Analysis of macroeconomic and financial history provides the opportunity to overcome the "rare events" problem (Schularick and Taylor, 2012), which is why the recent empirical literature increasingly employs panel data sets that not only reach back decades, but even centuries (see Madsen and Ang, 2016; Madsen et al., 2015; Almunia et al., 2010; Barro, 2009). Consideration of the past may, however, also be advantageous for a more pragmatic reason. In the early 1870s, the average income level of the included OECD nations resembled that of Botswana, Thailand, Brazil and Costa Rica in the most recent period of Dataset 1. For this reason, utilization of Dataset 2 offers an alternative perspective to study the long-run effect of FD, ruling out the possibility that the effects obtained via Dataset 1 are driven by country-specific characteristics of the contemporary developing world.

3.4.3 Financial development across the globe

Figure (3.1) illustrates the level of FD in the world using observations of the 2005–2009 period, for which data availability is maximized. The figure highlights a substantial degree of variation across nations around the globe. In accordance with the literature on the causes of FD (Rajan and Zingales, 2003; Chinn and Ito, 2006), credit to the private sector is highest in countries with high incomes (55.91%), democratic structures (51.80%), and trade openness (40.45%), whereas FD in the developing world (24.00%), in countries with an authoritarian form of government (22.20%), and in those with a low degree of openness (29.62%) lags substantially

33 Note that re-estimation of the empirical analysis of this chapter with a broader measure of FD such as overall liquid liabilities (M3) to GDP (as used in, e.g., Rousseau and Wachtel, 2002) yields results strongly comparable to those obtained via CREDIT.

34 The countries include Canada, the United States, Japan, Australia, New Zealand, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

behind.[35] During the depicted time-span, the countries with the highest levels of FD were Cyprus (205.99), Ireland (195.72), Iceland (191.97), Denmark (183.79), the United Kingdom (183.31), Spain (177.23), and Switzerland (166.16). The figure also reveals a distinct regional pattern, with FD being high in Europe, Northern America, and Asia, and considerably lower in Sub-Saharan Africa and some parts of Latin America.

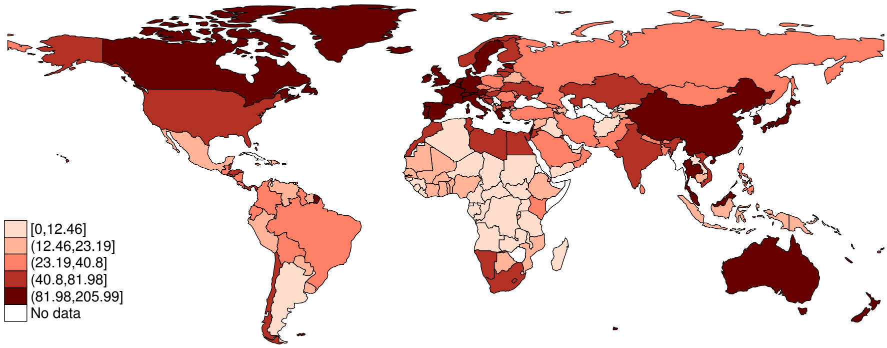


Figure 3.1 The level of financial development in the world, average of credit to the private sector to GDP between 2005 and 2009. Classes refer to the quintiles of the distribution.

On average, the size of the financial sector has increased considerably over the past decades. Whereas the mean of CREDIT was 20.38 in the early 1960s, the subsequent decades saw a substantial increase in the amount of credit channeled to the private sector, particularly during the 1980s and 1990s. In the late 2000s, the mean of CREDIT reached approximately 50 percent of GDP. As documented in Rajan and Zingales (2003), this development varies widely among countries. Figure (3.2) depicts the kernel density of CREDIT in both the 1980–1984 and the 2005–2019 periods, illustrating that the increase in the mean was mainly due to a sharp rise in a minority of countries, whereby there is only a slight increase in the mode of the distribution. This development has resulted in a substantial increase in the standard deviation of CREDIT between the early 1960s (16.88) and the late 2000s (45.77).

The highest absolute increases in CREDIT during the period illustrated in Figure (3.2) took place in the United Kingdom (+165.20), Cyprus (+165.17), Ireland (+164.16), the Netherlands (+161.74), and Denmark (+156.16). On the other side of the spectrum, there was little progress in FD in Uruguay (+5.19), Ecuador (+5.63), and Sudan (+0.23), and even a *decrease* in CREDIT in Gabon (-5.86), Syria (-4.9), Iraq (-3.18), and Côte d'Ivoire (-2.67). Overall, there is a distinct regional and income-related pattern visible in the data, which is presented in Figure (3.3). The left panel of the figure documents that the increase in the mean of CREDIT originates primarily from a substantial rise in the amount of capital directed to the private sector

[35] As benchmark values, the percentages are computed based on: the country classification of the World Bank (income level), the indicator of Freedom House (2017) (where the cut-off is 5), and the level of openness (where the cut-off is the sample mean).

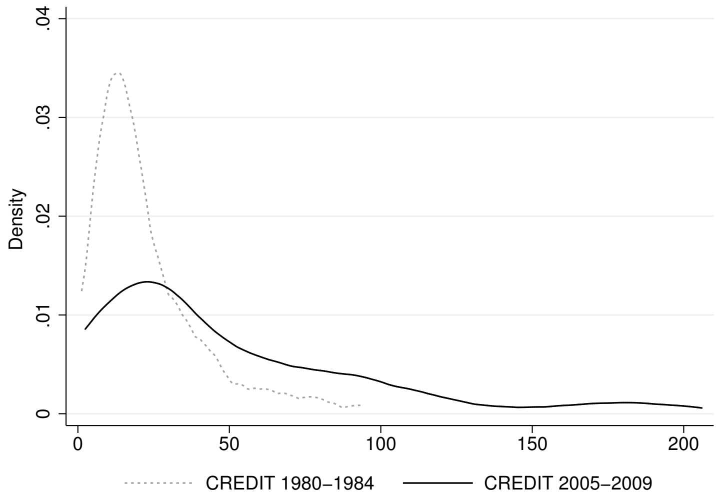


Figure 3.2 The empirical distribution of credit to the private sector to GDP. Kernel density estimate of CREDIT in the 1980–1985 period and the 2005–2009 period. Kernel is Epanechnikov.

in the advanced OECD-countries (grey dashed line). Within this group, the rise in CREDIT was largest in the countries located in North America and lower in European and Central Asian economies. The picture changes dramatically if we look at FD in Sub-Saharan Africa, Latin America, and the Arab World. Credit to the private sector in each of these regions currently amounts to 50 percent of GDP, which is roughly half of the ratio found for Europe and Central Asia, one-third of the OECD-average, and only a quarter of the percentage found for North America. A similar pattern can be seen with respect to income differences (right panel), where the substantial increase in CREDIT observable in high-income countries contrasts sharply with the slight progress made in low-income economies. However, while the level of FD in high-income countries appears to have stagnated since the turn of the millennium, in the early 2000s it experienced a hike in middle-income economies that has since initiated a convergence process.

3.4.4 A first glance at the effect of financial development on GDP and the transmission channels

Figure (3.4) illustrates the relationship between FD and income per capita based on two graphs. The first graph (Panel A) shows the correlation between CREDIT and the log of real per capita GDP, along with a kernel-weighted local polynomial regression line. The relationship between financial and economic development appears to be non-linear, where the positive association between per capita GDP and FD is strongly pronounced until credit to the private sector reaches 60–70% of GDP. From that level onward, the slope of the regression line decreases markedly, approximating zero once

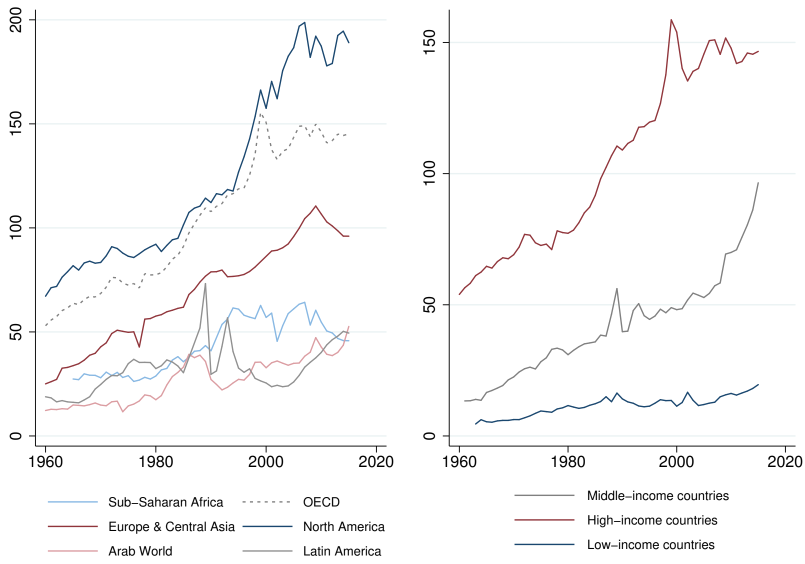


Figure 3.3 Financial development across regions (left panel) and development levels (right panel), 1960–2015.

the critical level of 100% of GDP identified by Arcand et al. (2015) is surpassed. Panel B considers this threshold in the framework of a data-driven regression discontinuity design (RD) following the approach of Calonico et al. (2014b,a, 2017a,b).[36] The depicted model is implemented using spacings estimators, producing an RD plot with evenly spaced bins that mimic the underlying variability of the data.[37] The figure shows that the relationship between credit to the private sector and per capita GDP fundamentally changes once CREDIT surpasses the critical threshold. While there is a strong positive relationship visible in countries with a credit-to-GDP ratio of less than 1 (correlation: 61.7%), this relationship vanishes once the critical value is surpassed (correlation: 19.1%).

The hypothesis presented in Section (3.2.4) stresses that a substantial part of the "vanishing effect of finance" emerges due to diminishing returns on reproducible factors and the change of the primary force of growth, reducing the importance of FD for investment, education, and fertility once these variables reach a sophisticated level. To provide an initial impression of the validity of this hypothesis, Figure (3.5) depicts the relationship between FD and the main transmission variables identified in

36 Note that unlike in a traditional RD setting, we are *not* interested in the local treatment effect around the cut-off, but rather in a potential change in the functional relationship between the variables before and after the critical threshold of CREDIT.

37 Note that Calonico et al. (2014a, 2017b) also provide a second method that uses evenly spaced bins selected to trace out the underlying regression function (IMSE-optimal selectors). RD plot based on this method produce strongly comparable results.

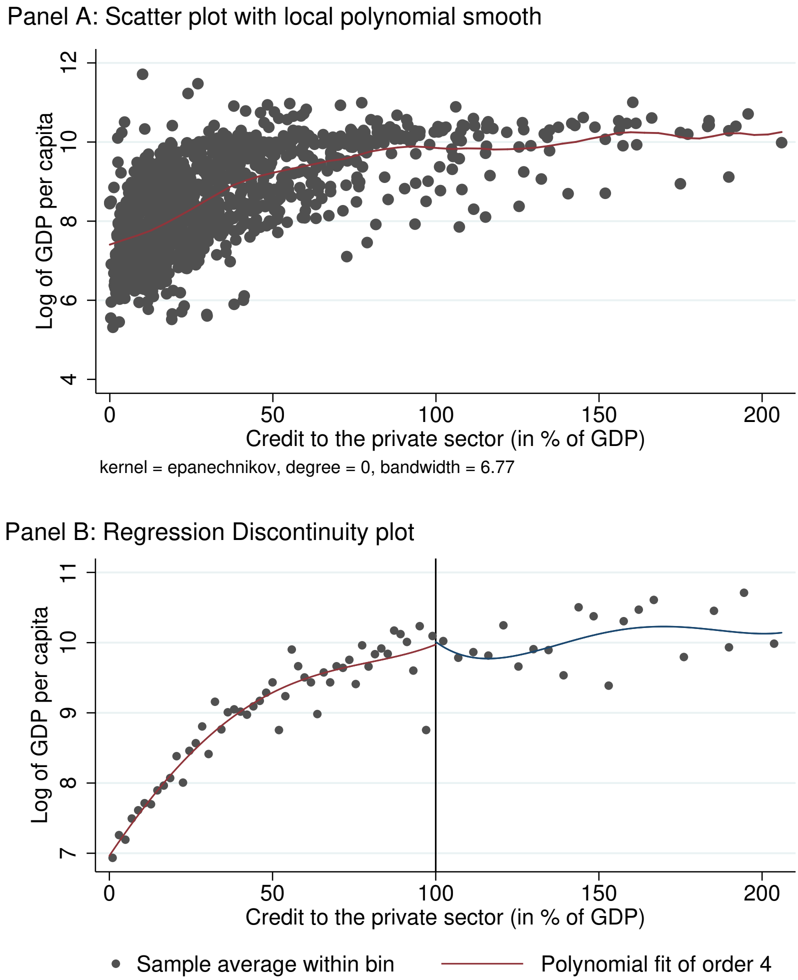


Figure 3.4 The relationship between economic and financial development, 1960–2015 (5-year averages). Panel A: Local polynomial smooth with Epanechnikov kernel (degree=3; bandwith=32.68). Panel B: Regression discontinuity plot with evenly spaced bins mimicking the underlying variability in the data. Implementation via spacings estimators (Calonico et al., 2014a).

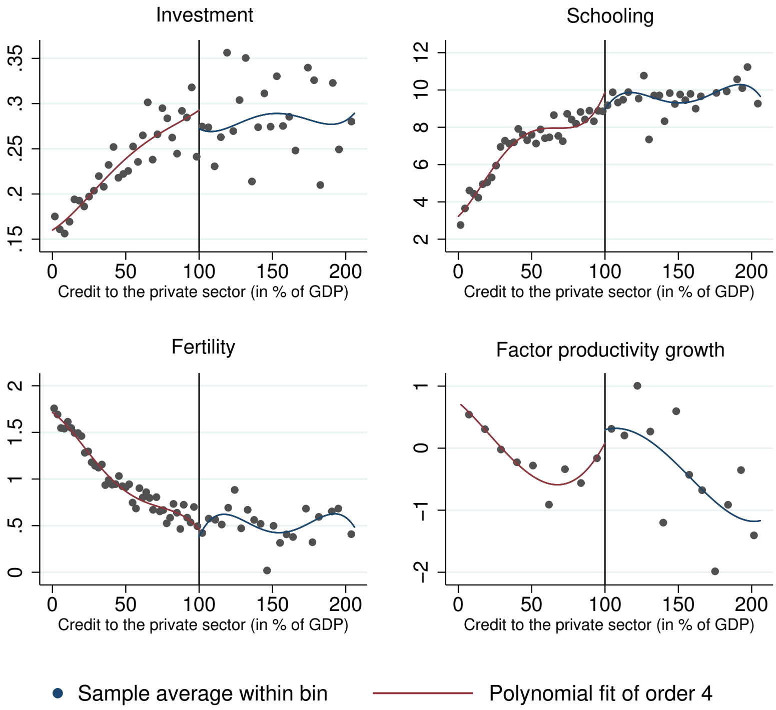


Figure 3.5 The relationship between financial development and the transmission variables: investment share, average years of schooling, log of fertility, and factor productivity growth. Period is 1960–2014 (5-year averages). Regression discontinuity plot with evenly spaced bins mimicking the underlying variability in the data. Implementation via spacings estimators (Calonico et al., 2014a).

Section (3.2.4), following the logic of the RD plot used in Panel B of Figure (3.4). The figure demonstrates that FD is positively related to investment, but it also shows that this relationship weakens during the development process of the financial sectors. A similar pattern is visible with respect to schooling, where the large slope of the RD regression line before the cut-off stands in sharp contrast to the considerably lower slope observable beyond the critical threshold.

There is also a substantial discrepancy between the relationship of fertility and FD before and after credit to the private sector reaches the cut-off point: while contributing to a reduction in the fertility rate in countries with small financial sectors, FD ceases to decrease fertility if the financial system has become mature. This deviation has simple biological roots: the average fertility rate in country-year observations exceeding the cut-off is 1.75, falling below the replacement level fertility. This is typically the case in the final stage of the demographic transition (Cervellati and Sunde, 2011). The implication is that once the demographic transition is finalized, there is no additional contribution of FD on growth via the fertility mechanism.

Contrary to the effect of FD on investment, education, and fertility, there is no distinct pattern visible for factor productivity growth, at least with regard to the sharp cut-off of 100% of GDP. Whereas the figure depicts a slight tendency for a negative relationship between the variables, the estimated non-parametric function assumes a local maximum right *after* the critical threshold.

3.5 Finance and growth during the past 50 years

3.5.1 Baseline results

Figures (3.4) and (3.5) provide a first look at the vanishing effect of finance and the potential mechanisms through which this effect may operate. These bivariate analyses, however, are afflicted with a number of methodological drawbacks, chiefly the neglect of covariates that may affect the growth rate, the potential of reverse causation, and the disregard of unobserved heterogeneity. In this section, we study the effect of FD and its transmission mechanisms in greater depth, using data from 129 countries observed between 1960 and 2014 in eleven non-overlapping 5-year periods (Dataset 1). To ensure comparability between the model specifications, each regression is based on an identical sample of 826 country-year observations. Table (3.1) presents the results obtained via application of the baseline specifications described in Section (3.4.1).

Panel A reports the results of the IV strategy that uses legal origin to instrument FD. Column (1) presents the results from a reduced-form model, leaving all possible transmission channels open. This model yields a point estimate that is negative and statistically significant at the 10% level.

To assess the stability of the effect of FD identified in Column (1), Columns (2)–(5) introduce a number of covariates that have been shown to influence long-run growth. As such, the newly introduced variables are "bad controls" in the sense that they are part of the causal effect we intend to estimate (Angrist and Pischke, 2009). For this reason, the more comprehensive model specifications do not capture (or attempt to capture) the full growth effect of FD. Rather, their comparison with the reduced model in Column (1) highlights potential mechanisms through which FD translates to growth. As discussed in Section (5.2), large parts of the growth effect of FD can be assumed to emerge via its facilitation of schooling and investment. When controlling for these effects in Column (2), the negative effect of FD intensifies from -0.0188 to -0.0486 and becomes statistically significant at the 1% level. This substantial increase in both magnitude and significance implies that much of FD's effect on growth operates through education and investment.

When introducing the inflation rate, openness, government consumption, and political rights, the effect of FD maintains both its size and its significance level, suggesting that these variables do not play a role in the effect of FD on income increases. However, once life expectancy and fertility are incorporated in the models

Table 3.1 Financial development and economic growth — Baseline estimates.

	(1)	(2)	(3)	(4)	(5)
	Panel A: 2SLS estimates				
Log GDP^{pc} $(t-1)$	0.460***	0.0395	-0.112	-0.984***	-1.397***
	(0.0662)	(0.0825)	(0.0954)	(0.312)	(0.286)
CREDIT	-0.0188*	-0.0486***	-0.0422***	-0.0151	-0.00843
	(0.0103)	(0.0145)	(0.0149)	(0.0176)	(0.0152)
Investment		13.45***	12.28***	9.501***	6.524***
		(3.063)	(2.900)	(2.696)	(2.227)
Years of Schooling		0.294***	0.178*	0.285***	-0.0207
		(0.0947)	(0.0929)	(0.0849)	(0.0758)
Inflation			-0.181***	-0.183***	-0.168***
			(0.0163)	(0.0151)	(0.0159)
Openness			0.733**	0.584**	0.517**
			(0.312)	(0.260)	(0.250)
Government Consumption			2.314	0.427	0.745
			(1.869)	(1.816)	(1.862)
Political Rights			0.243**	0.138	-0.0533
			(0.112)	(0.0999)	(0.0913)
Log Life Expectancy				1.935***	4.379***
				(0.613)	(0.530)
Log Fertility Rate					-2.830***
					(0.488)
Observations	826	826	826	826	826
Countries	129	129	129	129	129
Cragg-Donald F-Stat	78.85	53.01	43.66	20.14	28.79
SY 10% rel IV bias	9.08	9.08	9.08	9.08	9.08
SY 30% rel IV bias	5.39	5.39	5.39	5.39	5.39
OP F-Stat	17.08	14.94	13.95	16.31	20.99
OP τ =10%	11.33	11.23	11.51	11.29	12.12
KP rk LM p-val	0.001	0.001	0.001	0.000	0.000
CLR test interval	[-0.04; -0.01]	[-0.07; -0.04]	[-0.06; -0.03]	[-0.06; 0.00]	[-0.03; 0.01]
Wald test interval	[-0.03; -0.01]	[-0.06; -0.03]	[-0.06; -0.03]	[-0.04; 0.00]	[-0.03; 0.01]
UCI interval	[-0.03; 0.03]	[-0.08; -0.02]	[-0.06; -0.01]	[-0.04; -0.01]	[-0.04; 0.01]
F-Stat	51.01***	39.23***	58.46***	57.54***	62.85***
	Panel B: System GMM estimates				
CREDIT	-0.0317***	-0.0221***	-0.0258***	-0.0253***	-0.0204***
	(0.00839)	(0.00730)	(0.00634)	(0.00628)	(0.00652)
Hansen p-val	0.105	0.181	0.424	0.614	0.930
Diff-Hansen	0.505	0.836	0.786	1.000	1.000
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.325	0.507	0.757	0.772	0.668
SW F-Stat (diff)	5.40	4.66	20.34	23.02	35.07
SW F-Stat (lev)	72.55	70.79	38.77	42.28	47.03
SY 10% rel IV bias	11.51	11.30	11.04	10.99	10.95
SY 30% rel IV bias	4.63	4.28	4.02	3.99	3.95
SY χ^2 p-val	0.000	0.000	0.000	0.000	0.001
Instruments	88	106	138	147	156
	Panel C: Bias-corrected LSDV estimates				
CREDIT	-0.0268***	-0.0228***	-0.0255***	-0.0274***	-0.0251***
	(0.00525)	(0.00426)	(0.00561)	(0.00517)	(0.00515)
	Panel D: Non-linear effects of finance				
CREDIT	0.0829***	0.0196	0.0234	0.00157	0.00448
	(0.0247)	(0.0213)	(0.0187)	(0.0167)	(0.0140)
CREDIT squared	-0.00058***	-0.000211*	-0.000229**	-0.000125	-0.000102
	(0.000142)	(0.000113)	(0.0000950)	(0.0000865)	(0.0000693)
SLM p-val	0.000	0.179	0.105	0.463	0.375
Fieller 90% lower	54.64	0.00	0.00	0.00	0.00
Fieller 90% upper	86.28	75.93	77.36	61.56	75.39

Notes: Panel A reports 2SLS regressions using legal origins as instruments, standard errors are cluster-robust. Panel B documents two-step system GMM estimates with period fixed effects and Windmeijer-corrected standard errors. Panel C shows bias-corrected LSDV estimates (Bruno, 2005a,b) with 100 bootstrap-iterations. Panel D presents non-linear effects, employing the system GMM specifications of Panel B. $*p < .10, **p < .05, ***p < .01$

shown in Columns (4) and (5), the parameter estimate of FD shrinks considerably and becomes insignificant.

Naturally, the IV results hinge critically on the ability to instrument FD with legal instruments. Particularly with respect to growth regressions, there is a growing concern in the empirical literature that instruments may (1) be weak or (2) violate the exclusion restriction (Bazzi and Clemens, 2013). Table (3.1) takes this concern seriously, assessing the validity of the results based on a number of statistical tests. First, the table reports the Cragg-Donald F-statistic along with the critical values for a 10% and a 30% IV bias computed by Stock and Yogo (2005). In each case, the Cragg-Donald F-statistic substantially exceeds the 10% critical level. As a second diagnostic for weak instrumentation, the table presents the results of the test proposed by Olea and Pflüger (2013) (denoted with OP), which—in contrast to traditional weak IV tests—is robust to heteroskedasticity, autocorrelation, and clustering. Critical values are reported for the demanding threshold of $\tau = 10\%$, which is exceeded by the OP F-statistic in each case. Third, the p-value of the LM version of the rk test of Kleibergen and Paap (2006) emphasizes that there is no evidence of under-identification in any of the specifications.

As a final weak instrument diagnostic, Table (3.1) reports weak-instrument robust confidence intervals based upon the conditional likelihood ratio test (CLR) developed by Moreira (2003). These intervals are robust to weak instrumentation in the sense that they have the correct size in those cases where instruments are strong as well as those where they are not. For Columns (1)–(3), which highlight a significantly negative effect of FD on growth, the CLR-intervals are located entirely in the negative parameter space, suggesting that even in the presence of weak instrumentation, the effect of FD on growth will still be significantly negative. In addition, the CLR-intervals strongly resemble the non-robust Wald intervals, which implies that legal origin serves as a strong instrument for FD in our setting. This finding is illustrated graphically in Figure (B3-1) in the appendix, which depicts rejection probabilities and confidence intervals for both the CLR and the Wald tests.

While the test reported in Table (3.1) imply that weak instrumentation does not pose a problem, there is still the possibility that legal origin does not satisfy the exclusion restriction. To assess the consequences of a potential violation of this restriction, Table (3.1) presents the outcome of the union of confidence intervals (UCI) test developed by Conley et al. (2012), which asks whether the results change if we relax the exclusion restriction, assuming that the instrument is only "plausibly exogenous" (see Section (3.4.1)). In Columns (2)–(4), the UCI intervals lie in the negative parameter space. This result emphasizes that even if the exclusion restriction is substantially relaxed, inference based on the utilized instruments will still be informative. In the reduced specification and in Column (5), however, the null is included in the UCI interval, implying that the negative effect found in Column (1) is sensitive to satisfaction of the exclusion restriction.

An alternative strategy for both the evaluation of the exclusion restriction and the assessment of the stability of the parameter estimates lies in the alteration of the empirical technique. For this reason, Panels B and C change the estimation strategy, employing system GMM (Blundell and Bond, 1998) and bias-corrected LSDV estimates developed by Bruno (2005a,b). Both Panel B and Panel C highlight a high degree of robustness of the effects identified in Panel A, pointing to a significantly negative impact of FD on growth in each model specification. The magnitude of this effect, however, is slightly lower compared to the outcomes of the IV regressions. As in Panel A, a distinct pattern is that the size of the estimated parameter is substantially reduced when moving from Column (1) to Column (2), the latter of which introduces schooling and investment as potential transmission channels. A similar reduction in the estimated effect—although to a lesser extent—is found if fertility is introduced in Column (5).

To assess instrument strength of the system GMM estimator, Table (3.1) reports two tests proposed by Sanderson and Windmeijer (2016) (henceforth SW). The first is a weak instrument F-test that builds on Angrist and Pischke (2009) but allows for separate diagnostics for each endogenous regressor. The second is the SW χ^2 test for under-identification.[38] The SW test for weak instrumentation shows that in each model specification, the relative IV bias is less than 30 percent in both the level and the difference equations, and is often even (well) below 10 percent. The SW χ^2 test further points to absence of underidentification in both the levels and the difference equations.

As stressed in Section (5.2), the recent empirical literature on the FD-growth nexus emphasizes that FD exerts a non-linear effect on growth (e.g. Law and Singh, 2014; Arcand et al., 2015). Panel D is concerned with the identification of such an effect, introducing the squared level of CREDIT in model specifications (1)–(5). In the reduced model, the inverted-U relationship between FD and growth found in recent studies emerges as a clear empirical pattern, which is reflected in both the parameter estimates and the SLM test of Lind and Mehlum (2010). This relationship, however, disappears once education and investment are introduced in Column (2). In models (2) and (3) the overall effect of CREDIT is essentially negative, and in Columns (4) and (5) it becomes insignificant. The vanishing significance of the "left-hand side" of the FD-growth parabola again suggests that much of the positive effect that FD can have on growth operates via investment and schooling.

3.5.2 Deviations in the effect over space and time

In Section (5.2), we argued that the change in the transmission mechanism of growth over the development process suggested by the unified growth theory may also result in a change of FD's effect on growth. Table (3.2) examines this hypothesis, studying the impact of FD on growth "over space and time". Panel A introduces an interaction term between credit to the private sector and the log of real per capita GDP in the baseline specifications of Columns (1)–(5) in order to assess the growth effect of the

[38] In general, the SW tests are designed for weak-instrumentation diagnostics in the multiple-endogenous-regressor context of 2SLS settings. However, there have been some attempts to transfer these tests to dynamic panel GMM settings (Bun and Windmeijer, 2010; Newey and Windmeijer, 2009) via construction of the exact GMM instrument matrix for both the difference and the levels equations of the system GMM estimator. This matrix can in turn be used to carry out the standard 2SLS regressions and test.

financial sector conditional on the development level.[39] In the reduced specification reported in Column (1), the parameter estimate of CREDIT is positive and significant at the 1% level, suggesting that FD is strongly beneficial for income increases in poor economies. However, this positive effect disappears during the development process, as indicated by the negative sign of the interaction term of CREDIT and per capita GDP. This result is in line with the recent empirical literature (Huang and Lin, 2009; Aghion et al., 2005).

According to the hypotheses drawn in Section (5.2), we might expect that the conditional effect of FD on growth is due to investment in physical capital and education, which are the prime engines for growth in developing economies (Galor, 2009). Column (2) provides empirical support for this argument, where the only difference from the reduced specification in Column (1) is the introduction of schooling and investment. When the specification controls for these variables, the conditional effect of FD becomes insignificant.

Panel B is concerned with a similar analysis, examining changes in the effect of FD on growth over time via introduction of an interaction term between CREDIT and a time trend. The rationale of this strategy is that economies developed substantially during the time span covered in Dataset 1. While average real per capita income in the 109 countries for which data is available over the whole period between 1960 and 2015 was 4,140 USD in the early 1960s, it more than tripled by the end of the observation period (13,203 USD). Consequently, we might expect that the effect of FD over time behaves similarly to its effect over the development level. The reduced specification of Column (1) supports this argument: the parameter estimate of CREDIT is positive and significant at the 5% level, suggesting that on average, credit to the private sector promoted growth during the early periods of the sample. However, as the economies developed over the past 50 years, CREDIT became less and less relevant as a driver of economic growth. This second argument is reflected in the negative sign of the interaction term between CREDIT and the time trend, which is significant at the 1% level. Resembling the findings of Panel A, the positive contribution of FD to growth vanishes in Column (2), which introduces investment and education.

Figures (3.6) and (3.7) provide graphical illustration of the effect of FD over space and time. The graphs demonstrate that FD fosters growth in economies with an average income level lower than 10,900 USD and becomes significantly negative once a development level of 26,900 USD is surpassed. Meanwhile, the growth effect of FD was positive until the early 1990s and became negative around the beginning of the new millennium. At that time, the average income level in the world was 11,429 USD, closely resembling the critical threshold found with respect to the development path. This finding explains why earlier studies using data extending only to the 1990s or early 2000s did not identify a negative impact of FD on growth.

[39] Due to the lack of an applicable instrumental variable for the interaction term, the analysis uses system GMM to eliminate endogeneity drawing on internal instruments. The specifications exactly follow those of Panel B of Table (3.1).

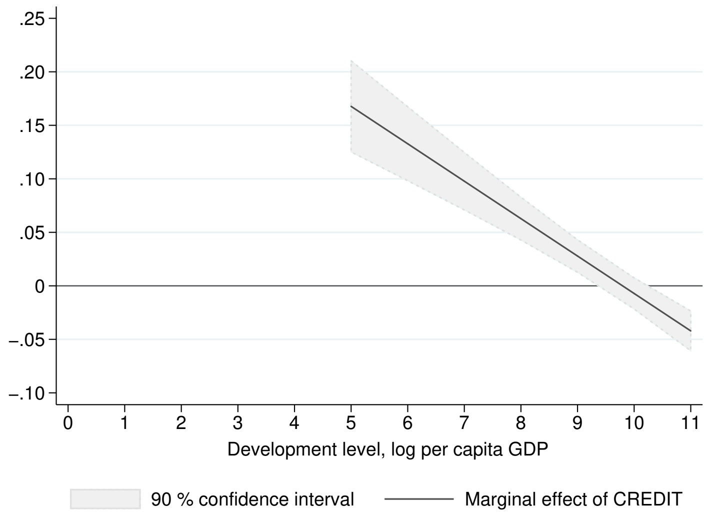


Figure 3.6 The effect of financial development across different levels of economic development. The marginal effects are computed based on the reduced specification reported in Column (1) of Table (3.2) (Panel A).

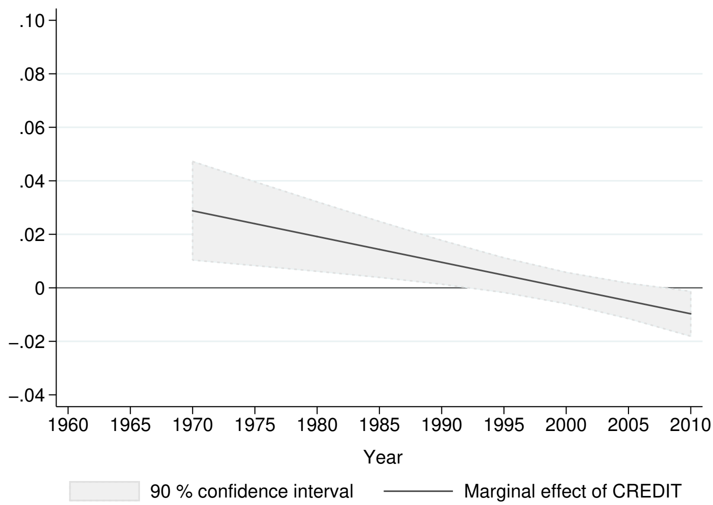


Figure 3.7 The effect of financial development over time. The marginal effects are computed based on the reduced specification reported in Column (1) of Table (3.2) (Panel B).

Table 3.2 Financial development and economic growth — Effects over space and time.

	(1)	(2)	(3)	(4)	(5)
	Panel A: Effect of finance over the development level				
CREDIT	0.343***	0.0843	0.0580	-0.0905	-0.0674
	(0.0550)	(0.0856)	(0.0744)	(0.0677)	(0.0648)
CREDIT × GDP	-0.0350***	-0.00991	-0.00729	0.00716	0.00546
	(0.00545)	(0.00854)	(0.00734)	(0.00685)	(0.00644)
Observations	826	826	826	826	826
Countries	129	129	129	129	129
Hansen p-val	0.539	0.770	0.893	0.952	0.991
Diff-Hansen	0.845	0.977	0.994	0.998	1.000
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.247	0.578	0.725	0.622	0.556
SW F-Stat (diff)	7.32	9.86	12.99	18.60	24.45
SW F-Stat (lev)	126.64	31.36	22.91	18.00	4.14
SY 10% rel IV bias	11.40	11.21	10.99	10.95	10.91
SY 30% rel IV bias	4.39	4.17	3.98	3.95	3.92
SY χ^2 p-val	0.000	0.000	0.000	0.000	0.000
Instruments	141	150	159	168	177
	Panel B: Effect of finance over time				
CREDIT	0.0429**	0.0254	0.0292	0.0208	0.0115
	(0.0167)	(0.0179)	(0.0187)	(0.0171)	(0.0149)
CREDIT × Time	-0.0037***	-0.00350**	-0.00386**	-0.0037***	-0.00238*
	(0.0013)	(0.00158)	(0.00158)	(0.0014)	(0.00132)
Observations	826	826	826	826	826
Countries	129	129	129	129	129
Hansen p-val	0.103	0.378	0.899	0.965	0.995
Diff-Hansen	0.942	0.987	1.000	1.000	1.000
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.386	0.520	0.951	0.828	0.722
SW F-Stat (diff)	26.58	23.95	23.40	24.27	37.19
SW F-Stat (lev)	76.48	34.11	0.80	3.92	6.27
SY 10% rel IV bias	11.40	11.21	10.99	10.95	10.91
SY 30% rel IV bias	4.39	4.17	3.98	3.95	3.92
SY χ^2 p-val	0.000	0.000	0.000	0.000	0.000
Instruments	105	123	154	163	172

Notes: Dependent variable is real per capita GDP growth. Table reports two-step system GMM estimates with period fixed effects and Windmeijer-corrected standard errors in parentheses. The instrument matrix uses all available lags and is collapsed. The specifications replicate those of Table (3.1), here introducing an interaction term between CREDIT and the log of real per capita GDP (Panel A) and a time-trend (Panel B). SW F-Stat and χ^2 p-val report the F-statistic of weak instrumentation and the p-value of the underidentification test according to Sanderson and Windmeijer (2016). Critical values are reported as SY 10% rel IV bias and SY 30% rel IV bias and are drawn from Stock and Yogo (2005). Hansen p-val gives the J-test for overidentifying restrictions. Diff-Hansen gives the Difference-in-Hansen statistic of the instrument subset for the level equation. AR(1) p-val and AR(2) p-val report the p-values of the AR(n) test. $*p < .10$, $**p < .05$, $***p < .01$

3.5.3 Transmission channels

The previous results have shown that the effect of FD in growth regressions crucially depends on human capital and investment. In the next step, we directly examine the channels through which FD translates to economic growth. Section (5.2) has identified four of these channels, emphasizing the roles of education, investment, fertility, and technological progress. While the first three of these factors are included in the baseline specifications of Tables (3.1) and (3.2), the role of technological progress has thus far been neglected. The reason is that reduced availability of variables measuring technical change is accompanied by a substantial decrease in the number of country-years exploitable for empirical research. When employing total factor productivity growth from the Total Economy Database (TED, 2017)—currently the most widely available indicator to proxy technological progress and knowledge accumulation—, the number of observations drops from 826 to 385. This decline is mainly due to a reduction in the time dimension: the present version of the Total Economy Database covers observations for 123 countries, but goes back only to 1990. Due to the strong conditionality of the effect of FD on growth with respect to time documented in the previous section, the prior estimations refrained from inclusion of TFP.

Table (3.3) marks the starting point in the direct analysis of the transmission channels, illustrating the effects of the four transmission variables on growth based on a common sample of 385 country-year observations from 103 countries.[40] The table highlights that each transmission variables possesses the expected sign: while schooling, investment, and technological progress boost economic growth, a high level of fertility is negatively associated with income increases. Each of these effects is significant at the 1% level. Similar to the results of Table (3.1), the effect of FD is significantly negative and sensitive to the inclusion of the transmission variables. In the benchmark specification presented in Column (1) of Panel A, which includes only FD and the initial income level, the point estimate of CREDIT is -0.0269. The negative effect intensifies if the model accounts for the contribution of finance to education (-0.0370), investment (-0.0307), and fertility-reduction (-0.0492). While Panel A does not identify a substantial change in the marginal effect of FD when including technological progress, incorporation of TFP results in the effect of FD becoming insignificant in Panels B and C, which use system GMM and bias-corrected LSDV regressions.

How can this change in the effect of TFP be explained? Table (3.4) seeks to answer this question by directly examining the effects of FD on the transmission variables. The table presents estimates of the reduced effects of finance for each variable, following the specification of model (1) of Table (3.1) (labeled *"reduced"*) as well as a comprehensive variant (labeled *"full"*) that draws on model (5). Due to the strong conditionality of the effect of FD with respect to space and time, Panel A examines the effect of FD on the transmission variables dependent upon the development level, while Panel B assesses its effect over time. To address the strong discrepancy between the number of available observations when comparing the models that include factor productivity with those that neglect TFP, Table (3.4) reports the results based on (i) all available country-years (Panels A1 and B1) and (ii) the common sample employed in Table (3.3) (Panels A2 and B2).

In accord with the implications of the previous estimations, the results in Panels A1 and A2 underscore a significant effect of FD on schooling and investment which is visible in both the reduced specification and the comprehensive model. However, the positive coefficient of CREDIT indicates that this effect is particularly relevant for country-year observations from developing countries. Meanwhile, the interaction

40 While TFP growth rates are available for 129 countries, availability of the covariates yields a further decrease in the number of countries.

Table 3.3 Financial development and economic growth — The effects of the transmission variables.

	(1)	(2)	(3)	(4)	(5)
	Panel A: 2SLS estimates				
Log GDP^{pc} $(t-1)$	0.293***	0.133	-0.0249	0.224***	0.581***
	(0.0760)	(0.0990)	(0.0718)	(0.0609)	(0.176)
CREDIT	-0.0269***	-0.0370***	-0.0307***	-0.0201***	-0.0492***
	(0.00964)	(0.0119)	(0.00909)	(0.00737)	(0.0182)
Years of Schooling		0.263***			
		(0.0956)			
Investment			14.54***		
			(3.713)		
Factor Productivity				0.753***	
				(0.0612)	
Log Fertility Rate					-1.739***
					(0.665)
Observations	385	385	385	385	385
Countries	103	103	103	103	103
Cragg-Donald F-Stat	42.47	37.67	41.64	41.95	32.10
KP rk Wald F-Stat	13.12	11.52	13.29	13.25	10.64
SY 10% rel IV bias	9.08	9.08	9.08	9.08	9.08
SY 30% rel IV bias	5.39	5.39	5.39	5.39	5.39
OP F-Stat	4.18	3.98	3.91	4.22	7.03
OP τ =10%	3.06	2.78	3.17	3.23	2.93
KP rk LM p-val	0.008	0.000	0.000	0.000	0.000
UCI Interval	[-0.08; -0.02]	[-0.08; -0.02]	[-0.08; -0.02]	[-0.08; -0.02]	[-0.08; 0.02]
F-Stat	46.60***	36.10***	38.90***	101.5***	29.68***
	Panel B: System GMM estimates				
CREDIT	-0.0327***	-0.0145*	-0.0210***	-0.00724	-0.0197*
	(0.00938)	(0.00768)	(0.00782)	(0.00927)	(0.0102)
Hansen p-val	0.030	0.080	0.235	0.022	0.105
Diff-Hansen	0.559	0.729	0.964	0.364	0.541
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.629	0.421	0.874	0.683	0.610
SW F-Stat (diff)	88.08	14.91	11.18	7.20	19.36
SW F-Stat (lev)	4.52	30.66	31.46	11.96	30.58
SY 10% rel IV bias	10.91	11.05	11.05	11.30	11.05
SY 30% rel IV bias	3.92	4.03	4.03	4.28	4.03
SY χ^2 p-val	0.000	0.000	0.000	0.000	0.000
Instruments	61	79	70	54	70
	Panel C: Bias-corrected LSDV estimates				
CREDIT	-0.0312***	-0.0350***	-0.0319***	-0.0100	-0.0323***
	(0.00980)	(0.00849)	(0.00868)	(0.00775)	(0.00945)

Notes: Dependent variable is real per capita GDP growth. Panel A reports 2SLS regressions using legal origins as external instruments, standard errors are cluster-robust. Panel B documents two-step system GMM estimates with period fixed effects and Windmeijer-corrected standard errors. The instrument matrix uses all available lags and is collapsed. Panel C shows bias-corrected LSDV estimates using Bruno (2005a,b). The estimator employs 100 bootstrap iterations to obtain robust standard errors. KP rk Wald and Cragg-Donald report the F-statistics of the weak identification test for the case of one single endogenous regressor (2SLS), while SW F-Stat reports the F-statistic of the Sanderson and Windmeijer (2016) test (system GMM), with critical values reported as SY 10% rel IV bias and SY 30% rel IV bias, collected from Stock and Yogo (2005). The KP rk LM p-val documents the p-value of the underidentification test proposed by Kleibergen and Paap (2006). OP F-Stat gives the effective F-statistic of Olea and Pflüger (2013), which is robust to heteroscedasticity, autocorrelation, and clustering, and the critical value for τ =10%. UCI Interval represents the 90% interval of the union of confidence interval test of Conley et al. (2012). Hansen p-val gives the J-test for overidentifying restrictions. Diff-Hansen gives the Difference-in-Hansen statistic of the instrument subset for the level equation. AR(1) p-val and AR(2) p-val report the p-values of the AR(n) test. $*p < .10$, $**p < .05$, $***p < .01$

term between FD and the development level is *negative* and significant at the 1% level, suggesting that FD becomes less relevant for investment in education and physical capital during the process of development. A similar effect can be seen with respect to fertility: in the reduced model, FD contributes to a reduction in the fertility rate in poor economies, thereby indirectly stimulating economic growth. This fertility-reducing effect, however, disappears during the development process. If the economy surpasses a critical average income level of roughly 17,000 USD, FD positively influences fertility.

Finally, Panels A1 and A2 show that credit to the private sector is particularly beneficial for technological progress in advanced economies, which is reflected by the positive parameter estimate of the interaction term between FD and GDP. Table (A3-2) in the appendix provides further evidence on the conditional effect of FD on technological progress based on alternative indicators of innovation activity.[41]

When studying the influence of FD on the transmission variables subject to the time dimension in Panels B1 and B2, the results with respect to investment reveal a conditionality similar to that detected in Panel A. Both the reduced and the comprehensive models suggest that FD was particularly relevant for investment in the earlier periods of the utilized sample, and that this stimulating effect has largely disappeared over time. As previously argued, this finding may again reflect the vanishing contribution of FD to fixed investment over the development process, as many countries developed considerably over the time horizon included in the employed panel. In addition, this deviation explains why recent studies found ambiguous results when modeling a linear effect of FD to investment (Beck et al., 2000; Bandiera and Caprio, 2000).

Finally, Panels B1 and B2 show that financial intermediaries have fostered factor productivity in the past, but this contribution has decreased significantly over the last decades. In sum, Table (3.4) provides a compelling explanation of FD's vanishing effect on growth: in poor economies, where the primary drivers of economic development are education and physical capital investments, FD is strongly beneficial for income increases. Due to diminishing returns of reproducible factors, this effect decreases over the development process. In more advanced economies, technological progress and knowledge replace investment in human and physical capital as the primary engines of income increases, which is why the growth effect of FD is dependent upon its influence on technical change. In principle, this effect is *positive* in advanced economies, as can be seen in Panels A1 and A2. However, Panel B reveals that the growth-enhancing effect of FD via stimulation of technological progress has lessened over time. This suggests that, unlike in earlier periods, the positive effect of FD on technological change does not compensate for its vanishing effect on education and investment. As a result, the overall effect of FD becomes insignificant.

[41] This analysis includes three indicators of innovation: i) The TFP estimates from Hall and Jones (1999), which gauges differences in productivity in the late 1980s. ii) The innovation performance index from EIU (2009), which consolidates information from different sources, including patents granted to applicants (by residence) per million population, but also alternative innovation indicators such as high-technology manufacturing and service output per head, citations from scientific and technological journals, and royalty and license fee receipts per GDP. iii) The number of patent applications per million population (EIU, 2009).

Table 3.4 Transmission channels of financial development to growth.

	Investment		Schooling		Fertility		TFP	
	(reduced)	(full)	(reduced)	(full)	(reduced)	(full)	(reduced)	(full)
	Panel A1: Conditional effect of finance over the development level (all observations)							
CREDIT	0.0041***	0.005***	0.030***	0.023**	-0.03***	0.003	-0.017	-0.132*
	(0.0009)	(0.002)	(0.007)	(0.009)	(0.007)	(0.011)	(0.065)	(0.069)
CREDIT $\times$ GDP	-0.0004***	-0.001***	-0.003***	-0.003***	0.003***	-0.0001	0.0004	0.012*
	(0.0001)	(0.0002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.006)	(0.007)
Observations	1067	826	918	826	1063	826	430	385
Countries	161	129	131	129	161	129	116	103
Hansen p-val	0.447	0.826	0.592	0.971	0.079	0.811	0.061	0.677
Diff-Hansen	0.999	1.000	0.978	1.000	0.986	1.000	0.253	0.942
AR(1) p-val	0.004	0.000	0.018	0.019	0.000	0.000	0.000	0.000
AR(2) p-val	0.496	0.632	0.000	0.000	0.139	0.299	0.442	0.827
SW F-Stat (diff)	565.04	24.34	32.32	24.73	40.36	19.88	9.42	195.02
SW F-Stat (lev)	29.73	20.15	19.15	3.79	31.54	10.24	21.21	6.61
SY 10% rel IV bias	11.32	10.92	11.32	10.92	11.32	10.92	11.32	10.92
SY 30% rel IV bias	4.29	3.93	4.29	3.93	4.29	3.93	4.29	3.93
SY χ^2 p-val	0.000	0.000	0.000	0.048	0.000	0.001	0.000	0.001
Instruments	167	155	125	162	132	156	34	116
	Panel A2: Conditional effect of finance over the development level (common sample)							
CREDIT	0.004***	0.004**	0.025***	0.030***	-0.04***	-0.0149	-0.0195	-0.132*
	(0.001)	(0.002)	(0.0086)	(0.0097)	(0.012)	(0.0132)	(0.0631)	(0.069)
CREDIT $\times$ GDP	-0.0004***	-0.0004**	-0.003***	-0.003***	0.0038***	0.0018	0.00038	0.012*
	(0.0001)	(0.0002)	(0.001)	(0.001)	(0.001)	(0.0013)	(0.006)	(0.007)
Observations	385	385	385	385	385	385	385	385
Countries	103	103	103	103	103	103	103	103
	Panel B1: Conditional effect of finance over time (all observations)							
CREDIT	0.0006*	0.001***	-0.003	-0.004	-0.01***	-0.001	0.009	0.037
	(0.0003)	(0.0004)	(0.003)	(0.004)	(0.003)	(0.002)	(0.018)	(0.023)
CREDIT $\times$ Time	-0.0001**	-0.009***	0.0001	0.0002	0.001***	0.0004	-0.002	-0.005**
	(0.00003)	(0.0001)	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.001)	(0.002)
Observations	1067	826	918	826	1063	826	430	385
Countries	161	129	131	129	161	129	116	103
Hansen p-val	0.044	0.762	0.207	0.670	0.027	0.800	0.023	0.687
Diff-Hansen	0.253	1.000	0.956	0.921	0.763	0.999	0.144	0.951
AR(1) p-val	0.003	0.000	0.017	0.028	0.000	0.000	0.000	0.000
AR(2) p-val	0.583	0.492	0.000	0.000	0.505	0.269	0.517	0.924
SW F-Stat (diff)	46.64	28.54	20.16	27.06	24.47	13.60	7.51	53.94
SW F-Stat (lev)	77.21	6.71	20.43	3.92	79.23	15.79	71.42	4.59
SY 10% rel IV bias	11.32	10.92	11.32	10.92	11.32	10.92	11.32	10.92
SY 30% rel IV bias	4.29	3.93	4.29	3.93	4.29	3.93	4.29	3.93
SY χ^2 p-val	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008
Instruments	131	155	104	142	96	156	55	116
	Panel B2: Conditional effect of finance over time (common sample)							
CREDIT	0.0024***	0.001**	-0.008	-0.0062	-0.01***	0.004	0.0023	0.037
	(0.0006)	(0.0006)	(0.005)	(0.0049)	(0.003)	(0.003)	(0.015)	(0.023)
CREDIT $\times$ Time	-0.0002***	-0.0001**	0.0007	0.0005	0.001***	-0.0002	-0.0012	-0.005**
	(0.0001)	(0.00007)	(0.0005)	(0.0005)	(0.0002)	(0.0003)	(0.0013)	(0.002)
Observations	385	385	385	385	385	385	385	385
Countries	103	103	103	103	103	103	103	103

Notes: Dependent variable is real per capita GDP growth. Table reports two-step system GMM estimates with period fixed effects and Windmeijer-corrected standard errors in parentheses. The instrument matrix uses all available lags and is collapsed. The specifications replicate those of Table (3.1), here introducing and interaction term of CREDIT and the log of real per capita GDP (Panel A) and a time-trend (Panel B). SW F-Stat and χ^2 p-val report the F-statistic of weak instrumentation and the p-value of the under-identification test following Sanderson and Windmeijer (2016). Critical values are reported as SY 10% rel IV bias and SY 30% rel IV bias and are obtained from Stock and Yogo (2005). Hansen p-val gives the J-test of overidentifying restrictions. Diff-Hansen gives the Difference-in-Hansen statistic of the instrument subset for the level equation. AR(1) p-val and AR(2) p-val report the p-values of the AR(n) test. $*p < .10$, $**p < .05$, $***p < .01$

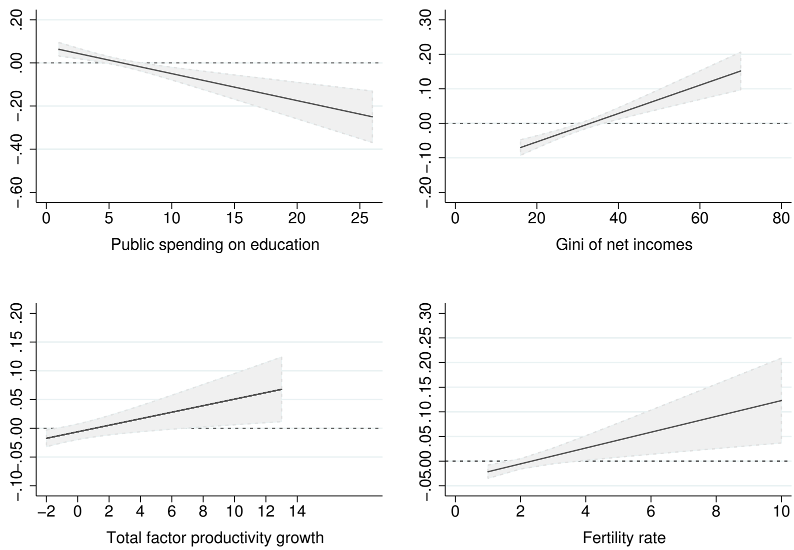


Figure 3.8 The effect of financial development on growth conditional on public spending on education, net inequality, technological change, and fertility. The marginal effects are calculated by replicating the reduced model specification used in Table (3.4).

Figure (3.8) takes a closer look at this transmission mechanism and some accompanying circumstances, illustrating the effect of FD on growth contingent upon public spending on education relative to GDP, the level of net inequality, technological change proxied by factor productivity, and the fertility rate.[42] In accord with the relationship between finance and education found in Table (3.4), FD turns out to be beneficial for growth in countries with underdeveloped public schooling systems, and becomes a negative factor once education is provided freely to all households. In this case, the provision of funds necessary for education is less relevant. Similarly, FD facilitates growth in countries with a high degree of net inequality, where many households face binding budget constraints that prevent their children from receiving a level of education that corresponds with their intellectual potential. In societies with a more equal distribution of incomes, the positive effect of FD vanishes. These results exactly coincide with the capital market imperfections argument pioneered by Galor and Zeira (1993) and Galor and Moav (2004).

The marginal effects of FD illustrated in the upper graphs of Figure (3.8) suggest, essentially, that FD is beneficial for growth in countries with an unequal distribution of opportunity, while it is less relevant—or even an impediment—in nations where

[42] The regressions augment the reduced model of Table (3.4) with the corresponding interaction terms. Data for public spending on education is collected from World Bank (2017), the data source for the net inequality series is Solt (2016).

opportunities are distributed more equally. From this perspective, the results complement the literature stressing the growth effects of inequality of opportunity (see Marrero and Rodriguez, 2013; Roemer and Trannoy, 2016; and Marrero et al., 2016).

The graphs depicted in the lower half of Figure (3.8) focus on two additional arguments, concentrating on innovation and fertility. In line with the results found in Table (3.4), the figure shows that FD is growth-enhancing in times when factor productivity grows at high rates, while it hampers income increases when new ideas are scarce. This outcome supports a crucial theoretical argument discussed in Section (5.2): if the potential for promising innovation projects is high, the financial sector boosts growth via the arguments summarized in Levine (2005), i.e. ex-ante screening of the projects, the provision of funds and knowledge, and the reduction in the inherent risks of the innovation. However, in times when the potential offered by innovations and new ideas is low, financial intermediaries increasingly engage in "non-intermediation" activities (Beck et al., 2014b) which do not contribute to income increases and may enhance the vulnerability of economies to crises (Rajan, 2005; Beck et al., 2014a). In the context of these results, the major decline in productivity growth observable in most advanced economies since the turn of the millennium (Gordon, 2012; Gordon, 2015) may to a large extent explain the negative effect of FD on growth in the post-2005 period.

Finally, the last regression depicted in Figure (3.8) illustrates the effect of FD on growth dependent upon the fertility rate. This analysis shows that FD stimulates growth in countries with high fertility rates, while it is an impediment to income gains in countries with fewer children per woman.

3.6 Finance and growth in the long run, 1870–2009

In the second step, we analyze the effects of FD and its transmission mechanisms in the very long run, covering the period between 1870 and 2009 for 21 OECD countries. These countries include Canada, the United States, Japan, Australia, New Zealand, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. In the early 1870s, many of these countries were at a stage in the development process that roughly coincides with that of today's low- and middle income economies. The rationale for analyzing the long-run perspective is that it provides the opportunity to examine FD's effects through a different lens, ruling out the possibility that the results of Section (3.5)—particularly those contingent upon time and space—are influenced by country-specific factors of the contemporary developing world. Given the parallels in terms of growth which are visible when comparing the OECD countries in the late 19th century with the emerging countries at present, the hypothesis is that many of the effects identified in Section (3.5) may also be found in historical data. The analysis uses long-run time series from Madsen and Ang (2016), Madsen (2014), and Madsen (2008) that draw on a rich set of historical data sources.

3.6.1 Baseline results

Dataset 2 covers 28 non-overlapping 5-year periods. For this reason, the system GMM estimator that is designed for panels with "large N" and "small T" is not feasible in this setting (Miguel and Satyanath, 2010; Bond et al., 2001). Instead, the empirical analysis follows a two-part strategy, building on the traditional within-group estimator (Panel A) as well as an instrumental variable approach that uses the agricultural share to eliminate endogeneity (Panel B). While dynamic panel fixed effects estimates are biased in short panels due to a mechanical correlation between the error term and the lagged dependent variable, Nickell (1981) has shown that this bias vanishes as T increases.[43] Judson and Owen (1999) conclude that in a framework of $T \approx 30$, the traditional within-group estimator works better than other possible alternatives.

The instrumentation strategy employed in the IV setting is theoretically motivated by Rajan and Zingales (2003) and empirically motivated by Madsen and Ang (2016). The argument is that agricultural output and FD stand in a negative relationship due to a conflict between the landed class and the urban merchant class. While the former has an interest in maintaining the status quo, the latter is primarily interested in promoting FD. A high ratio of agricultural output therefore reflects an imbalance between the classes in favor of the landed class.[44]

Table (3.5) reproduces the empirical specification of the baseline regressions reported in Table (3.1) as closely as possible. However, non-availability of historical time series with respect to some of the covariates prohibits exact replication. For this reason, the table distinguishes between two model specifications, including both a reduced variant (labeled with *"reduced"*) that is identical to the specification used in Column (1) of Table (3.1) and a comprehensive model (labeled with *"full"*) that includes the transmission variables analyzed in the previous section. Specifically, investment, schooling, and technological progress are directly included, while non-availability of historical series on fertility makes it necessary to proxy the fertility rate with life expectancy. This strategy is motivated by the theory of demographic transition, which is essential in explaining the transition from stagnation to growth. One of its key features is that—after an initial hike—population growth declines in association with higher life expectancy and lower fertility (Cervellati and Sunde, 2011). In this framework, life expectancy contributes to a reduction in fertility rates due to its positive effects on human capital (Soartes, 2005; Cervellati and Sunde,

[43] The exact magnitude of the bias in the historical analysis can be gauged via (Nickell, 1981):

$$\text{plim}(\hat{\theta} - \theta) \cong \frac{-(1+\theta)}{T-1}.$$

In our case the estimated relationship between the dependent variable in t and $t-1$ is -2.08, which yields a very small dynamic panel bias of roughly 3.8%.

[44] As a second instrumentation strategy, Madsen and Ang (2016) suggest using unionization as a long-run instrument for historical analyses of FD. While the authors provide a detailed discussion concerning the exclusion restrictions of both instruments, our tests for weak instruments and under-identification clearly indicate that the usage of agricultural output is superior to that of unionization (not reported).

Table 3.5 Financial development and growth in the long run, 1870–2009.

	Linear effect		Non-linear and conditional effects		
	(reduced)	(full)	(non-linear)	(over time)	(over space)
	Panel A: Within-Group estimates				
CREDIT	0.560	-0.008	0.012	0.022*	0.036***
	(0.454)	(0.005)	(0.012)	(0.011)	(0.008)
CREDIT SQUARED			-0.004		
			(0.005)		
CREDIT $\times$ Time				-0.001*	
				(0.0005)	
CREDIT $\times$ GDP					-0.003***
					(0.001)
Observations	567	567	567	567	567
Countries	21	21	21	21	21
R-Squared	0.309	0.378	0.310	0.314	0.313
	Panel B: IV estimates				
CREDIT	0.145***	-0.010	0.197***	0.362***	0.717***
	(0.048)	(0.021)	(0.044)	(0.103)	(0.263)
CREDIT SQUARED			-0.100***		
			(0.022)		
CREDIT $\times$ Time				-0.014***	
				(0.004)	
CREDIT $\times$ GDP					-0.057***
					(0.021)
Observations	567	567	567	567	567
Countries	21	21	21	21	21
Cragg-Donald F-Stat	14.63	17.88	53.48	13.00	8.03
KP rk Wald F-Stat	17.12	16.74	69.16	22.86	9.47
SY 15% max IV size	8.96	8.96	8.96	8.96	9.08
SY 25% max IV size	5.53	5.53	5.53	5.53	5.39
OP F-Stat	17.12	16.74	69.16	22.86	9.47
OP τ =20%	12.37	12.37	12.37	12.37	12.12
OP τ =30%	9.65	9.65	9.65	9.65	9.65
KP rk LM p-val	0.000	0.000	0.000	0.000	0.002
CRL test interval	[0.080; 0.279]	[-0.027; 0.004]	[0.124; 0.292]	[0.209; 0.698]	[0.392; 1.734]
Wald test interval	[0.061; 0.230]	[-0.027; 0.003]	[0.117; 0.278]	[0.159; 0.565]	[0.244; 1.191]
UCI interval	[0.022; 0.306]	[-0.049; 0.030]	[0.051; 0.362]	[0.072; 0.745]	[0.108; 1.576]

Notes: Dependent variable is real per capita GDP growth. Panel A reports Within-Group estimates with robust standard errors in parentheses. Panel B documents IV regression results, with financial development instrumented via agricultural output and robust standard errors in parentheses. Column (1) gives the marginal effect in a reduced model, Column (2) augments the specification via inclusion of three transmission channels: schooling, Investment, and health measured by life expectancy. Columns (3) and (4) replicate the reduced specification and introduce interaction terms with a time trend (Column 3) and the log of real per capita GDP (Column 4). KP rk Wald and Cragg-Donald report the F-statistics of the weak identification test in the case of one single endogenous regressor (2SLS), with critical values reported as SY 15% max IV size and SY 35% rel IV size that are collected from Stock and Yogo (2005). KP rk Wald F-Stat and KP rk LM p-val document the weak identification tests proposed by Kleibergen and Paap (2006). OP F-Stat gives the effective F-statistic of Olea and Pflüger (2013), which is robust to heteroscedasticity, autocorrelation, and clustering, and the critical value for τ =20% and τ = 30%. UCI Interval represents the 90% interval of the union of confidence interval test of Conley et al. (2012). CLR and Wald test intervals report the weak-instrument robust interval of Moreira (2003) and the non-robust Wald interval. $*p < .10$, $**p < .05$, $***p < .01$

2007).[45] In Dataset 1, which includes data for both series, the correlation between life expectancy and fertility is -84%. A further advantage of including life expectancy concerns the interpretation of a possible change in FD's effect between the reduced and the comprehensive model: as a reduction in fertility is reflected in higher life expectancy, the effect of each transmission variable operates in the same direction, whereas we might expect to find ambiguous effects when directly including the fertility rate.

The remaining columns of Table (3.5) complement the investigation of the previous section in assessing non-linear effects of FD and conditional effects contingent upon time and space. As in the previous tables, all models are built on a common sample that includes 567 country-year observations.

The first column of Table (3.5) presents the results of the reduced model, which examines linear effects running from FD to growth. In this setting, both the within-group model and the IV specification find a positive parameter estimate. While the effect is insignificant in Panel A, it becomes significant at the 1% level once the model accounts for endogeneity in Panel B. These results lead to two important conclusions: first, the isolated effect of FD on growth has been *positive* in the OECD countries over the past 140 years. Second, there is a potential reverse causation running from economic to financial development. In the more comprehensive model specification in the second column, however, the effect of FD becomes negative and insignificant in both Panel A and Panel B. Resembling the findings of the previous section, this change in the estimated parameter of FD suggests that education, investment, technological progress, and demographic factors are important channels through which FD operates to affect growth.

The second group of regressions presented in Table (3.5) concerns non-linear effects of FD on income increases. The inverted-U relationship between FD and growth found in the baseline table reappears in both panels and is significant at the 1% level for the IV regressions. This underscores the hypothesis drawn in Section (3.5.1) that the overall negative effect of credit to the private sector observable in the linear panel model of Dataset 1 reflects the "right-hand side" of a (much longer) inverted-U parabola. The column labeled *"over time"* asks at which period the former positive effect of FD has become negative, introducing an interaction term between FD and a time trend. Based on both within-group and IV models, the parameter estimate of CREDIT is positive and highly significant, indicating that FD substantially facilitated income increases in the OECD countries during the late 19th century. Meanwhile, the significantly negative sign of the interaction term underlines that this formerly positive effect has gradually vanished over time. Figure (3.9) provides a graphical illustration of this effect, suggesting that the point at which the positive effect of FD disappeared roughly coincides with the threshold identified in Section (3.5.1), implying however that it was reached one 5-year period earlier. This slight difference is due to (1) the longer time span covered in Dataset 1, and (2) the focus on OECD countries, which have reached a considerably higher development

[45] In addition, empirical evidence on a direct relationship between life expectancy and fertility is provided by Zhang and Zhang (2005), who demonstrate that both variables are significantly and negatively related.

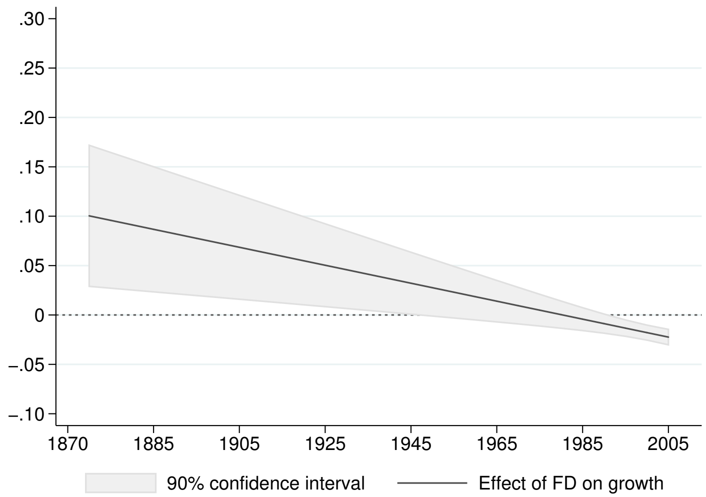


Figure 3.9 The effect of financial development over time, 1870–2010. The marginal effects are computed based on the specification reported in Column (over time) of Table (3.5) (Panel B).

level compared with the average of the countries in Dataset 1 by the turn of the new millennium. Consequently, we might expect that the influence of FD on the transmission channels disappears earlier in Dataset 2.

Finally, the last column of Table (3.5) again highlights that the impact of FD on growth is dependent on the development level, where the marginal effect is strongly positive in less developed countries and becomes increasingly irrelevant during the development process. Both the effect of CREDIT and the interaction term between CREDIT and the development level are significant at the 1% level. Table (3.5) further carries out a number of tests to evaluate the employed instrumentation strategy. These tests point to a considerable degree of instrument strength and demonstrate that inference would still be informative even if agricultural output were only "plausibly exogenous".

Taken together, the results obtained via the examination of historical data provide strong support for the findings in Section (3.5) from on a panel of countries observed during the past 50 years. Consistent with these findings, the conclusion is that FD fostered economic development in the past, and that this effect has vanished over space and time. Meanwhile, the disappearance of the significantly positive contribution of FD to growth between Columns (reduced) and (full) again emphasizes that much of FD's positive effect on growth is channeled via schooling, investment, technological progress, and demographic factors. As these transmission variables developed considerably in the OECD countries during the past 140 years, the growth effect of FD became increasingly weaker over time.

3.6.2 Transmission Channels

The change in the estimated parameters across Columns (reduced) and (full) of Table (3.5) again suggests that much of the deviation in FD's growth effect can be traced to the way it influences the transmission channels over space and time. This section enlarges the analysis of the transmission channels conducted in Section (3.5.3) via inclusion of a historical perspective spanning the period from 1870 to 2010.

Table 3.6 Transmission channels of financial institutions to growth in the long run, 1870–2009.

	Schooling		Investment		Life Expectancy		Knowledge	
	(time)	(develop.)	(time)	(develop.)	(time)	(develop.)	(time)	(develop.)
	Panel A: Within-Group estimates							
CREDIT	0.918**	1.154**	0.035***	0.0108	1.662	5.47**	-0.211*	1.124
	(0.426)	(0.416)	(0.006)	(0.017)	(1.259)	(2.18)	(0.105)	(0.822)
CREDIT $\times$ Time	-0.05**		-0.001***		0.084		0.029*	
	(0.021)		(0.0004)		(0.085)		(0.014)	
CREDIT $\times$ GDP		-0.11***		0.0002		-0.22		-0.079
		(0.0381)		(0.0014)		(0.19)		(0.069)
Observations	567	567	567	567	567	567	567	567
Countries	21	21	21	21	21	21	21	21
R-Squared	0.905	0.892	0.468	0.454	0.952	0.952	0.378	0.384
F-Stat	20.26***	24.452***	2.60***	2.91***	24.89***	5.43***	18.15***	22.01***
	Panel B: IV estimates							
CREDIT	13.52***	20.46***	0.210***	0.401**	56.52***	110.8***	-0.943**	-1.581*
	(2.926)	(5.410)	(0.0650)	(0.163)	(13.78)	(38.05)	(0.398)	(0.938)
CREDIT $\times$ Time	-0.5***		-0.008***		-1.95***		0.043***	
	(0.11)		(0.002)		(0.50)		(0.016)	
CREDIT $\times$ GDP		-1.63***		-0.032**		-8.53***		0.132*
		(0.443)		(0.012)		(3.076)		(0.076)
Observations	567	567	567	567	567	567	567	567
Countries	21	21	21	21	21	21	21	21
Cragg-Donald	12.97	11.54	13.00	8.03	13.00	8.03	13.00	8.03
KP rk Wald F-Stat	23.08	12.74	22.86	9.47	22.86	9.47	22.86	9.47
SY 15% max IV	8.96	8.96	8.96	8.96	8.96	8.96	8.96	8.96
SY 25% max IV	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53
OP F-Stat	23.08	12.74	22.86	9.47	22.86	9.47	22.86	9.47
OP τ =20%	12.37	12.37	12.37	12.37	12.37	12.37	12.37	12.37
OP τ =30%	9.65	9.65	9.65	9.65	9.65	9.65	9.65	9.65
KP rk LM p-val	0.000	0.000	0.000	0.002	0.000	0.002	0.000	0.002
CRL test interval	[7.3;16.9]	[13.3;39.4]	[0.12;0.40]	[0.20;0.99]	[36.7;106]	[67;262]	[-2.1;-0.3]	[-5.6;-0.2]
Wald test interval	[6.2;14.6]	[10.2;30.7]	[0.09;0.33]	[0.16;0.69]	[28.5;84.5]	[44.1;178]	[-1.7;-0.2]	[-3.5; 0.3]
UCI interval	[5.1;16.9]	[8.4;35.6]	[0.04;0.43]	[0.04;0.90]	[28.5;84.6]	[44.0;177]	[-1.9;-0.1]	[-3.9; 0.5]

Notes: Dependent variables are the logs of secondary schooling, investment, and knowledge. Panel A reports Within-Group estimates with robust standard errors in parentheses. Panel B documents IV regression results, with financial development instrumented via agricultural output and robust standard errors in parentheses. Columns (time) and (develop.) replicate the reduced specification of the effect of finance on growth and introduce interaction terms with a time trend and the log of real per capita GDP. KP rk Wald and Cragg-Donald report the F-statistics of the weak identification test in the case of one single endogenous regressor (2SLS), with critical values reported as SY 15% max IV size and SY 35% rel IV size, which are collected from Stock and Yogo (2005). KP rk Wald F-Stat and KP rk LM p-val document the weak identification tests proposed by Kleibergen and Paap (2006). OP F-Stat gives the effective F-statistic of Olea and Pflüger (2013), which is robust to heteroscedasticity, autocorrelation, and clustering, and the critical value for τ =20% and $\tau = 30\%$. UCI Interval represents the 90% interval of the union of confidence interval test of Conley et al. (2012). CLR and Wald test intervals report the weak-instrument robust interval of Moreira (2003) and the non-robust Wald interval. $*p < .10, **p < .05, ***p < .01$

The analysis exactly replicates the approach of Madsen and Ang (2016) to construct time series of the historical transmission variables. Specifically, schooling is measured via the logarithmic fraction of the population in the 13- to 17-year-age cohort enrolled in secondary education, fixed investment is the estimated share of investment relative to the stock of physical capital, and technological progress is measured via the number of patents filed by domestic residents, proxying the level of knowledge available in a given country-year.[46] In addition, due to a lack of comparable historical data on births per women, fertility is again approximated via life expectancy at age 10.

Regarding investment in physical capital and education, the results obtained via application of historical data strongly coincide with those based on Dataset 1. Table (3.6) highlights that in general, credit to the private sector fosters education and investment, but this effect wears off as the development level increases. A similar vanishing effect is visible with respect to time, which reflects the substantial increase in welfare observable in the OECD countries between the late 19th century and today. These effects are strongly significant when using IV in Panel B, and (slightly) less pronounced with regard to the within-group estimator in Panel A.

The results also imply a vanishing effect on life expectancy. During the 19th century, FD contributed to a substantial increase in general health and longevity. While the average life expectancy in the OECD countries was 48.19 years in the 1870–1874 period, it experienced a considerable leap during the following century, climbing to roughly 65 years by the early 1970s. The vanishing relevance of FD for health visible across both space and time may be due to a considerable increase in public health provision that commonly accompanies economic development: once the income level of the households and the social security systems reaches a sophisticated level, a substantial portion of the available medical treatments becomes affordable for a large part of the population. In this case, the effect of FD decreases, while preferences and lifestyle become decisive health factors (Contoyannis and Jones, 2004). In addition, due to the close long-run relationship between fertility and life expectancy, we assess these effects as a tentative implication that the fertility-reducing influence of FD becomes less relevant over the development process, which corresponds to the results obtained based on Dataset 1 in Section (3.5.3).

Finally, the last set of transmission analyses reports the impact of FD on technological progress. The penultimate column examines the change in the effect of FD on the knowledge variable across time, illustrating that FD did not foster idea production during the late 1870s, but became increasingly important over time. This effect deviates from the results identified in Section (3.5.3), suggesting that the declining effect of FD on ideas production found in the more recent data covering the past decade(s) is more than canceled-out by the substantial influence of FD on knowledge accumulation in the very long run.[47] In addition, the final column implies that the change over time may also be due to the vast income increases in the OECD countries over the past 140 years. The IV estimates in this case suggest that FD was particularly

[46] All variables are based on a large number of different historical sources which are described in detail in the online appendix of Madsen and Ang (2016). To obtain a more accessible coefficient size, the knowledge variable is divided by 10^6.

[47] When reducing the time horizon of Dataset 2 to match that of Dataset 1, the vanishing effect of FD on knowledge re-appears (not reported), underscoring that the deviation in the effect is indeed due to the time span and not due to data inconsistencies.

beneficial for knowledge production in richer countries, supporting the findings based on the past 50 years reported in Section (3.5.3).

Overall, the findings obtained via historical data strongly coincide with the conclusions drawn based on the broad panel of countries featuring information from the past 50 years in Section (3.5).

3.7 Concluding remarks

Based on both a broad panel of country observations from the past 50 years and a historical dataset including information for 21 OECD nations between 1870 and 2009, this chapter highlights that the formerly positive influence of FD on economic growth has disappeared since the early 2000s. This "vanishing effect of finance" is mainly driven by advanced economies, whereas FD is still beneficial for income increases in developing countries. The reason is that growth and FD are associated via a nonlinear relationship which is due to a fundamental change in the transmission mechanism of finance across different levels of economic development.

Developing countries benefit from FD via its positive stimulus to education, physical capital investment, and health, as well as via a reduction in the fertility rate. As average incomes rise, these transmission channels become increasingly irrelevant, particularly if public spending on education is high and inequality of net incomes is low. In contrast to earlier stages in the development process, the growth effect of FD in advanced economies crucially depends on technological change, as knowledge production becomes the primary transmission mechanism of FD once countries have achieved a sophisticated level of development.

However, while this result suggests that FD channels differently to growth depending on the stage in the development path, it does not explain why the overall effect of FD on growth has vanished during the past 15 years. The key for this explanation lies in the remarkably reduction in factor productivity growth that has occurred since the turn of the millennium, leading to a renaissance of the theory of "secular stagnation" (Gordon, 2012, 2015). In fact, when looking at the recent trend in TFP growth among the OECD countries included in the historical sample of Dataset 2, a striking empirical pattern is that productivity gains declined substantially and cross-nationally between the early 1990s and 2016 (see Figure (B3-2) in the appendix): out of 21 countries, 19 experienced a negative trend in factor productivity during the past 25 years.[48] The cross-national decline in TFP growth may explain why FD has ceased to translate to growth in the advanced economies since the early 2000s.

A direct implication of this argument, however, is that FD could regain its beneficial role in promoting growth if future rates of innovation were to approximate those of the past. While supporting the faltering innovation activity for the United States based on the number of patent grants per million of the US population, Korotayev et al. (2011) find that knowledge production has been subject to long-run phases

[48] Even in the two countries that possess a positive trend at the end of the illustrated time period (Spain and Japan), TFP trend *growth rates* were still negative over the overwhelming number of periods.

of up- and downswings in economic history. Given these long-run dynamics, the negative effect of FD on growth may not be a process that will persist forever.

While the goal of this chapter is on the analysis of the vanishing effect of FD, the reduction in TFP growth may also provide an explanation for why the effect of FD became *negative* in the more recent periods of the sample. If opportunities to financially support innovation and investment projects are limited, financial institutions have a greater interest to develop new business segments, particularly if monetary policy is expansive (Borio et al., 2017). Rather than triggering growth effects, these “non-intermediation activites” (Demirgüç-Kunt and Huizinga, 2010) mainly contribute to a higher vulnerability of countries to economic crises in mature financial sectors (see Beck et al. (2014b); Beck et al. (2014a); Rajan, 2005).

3.A Appendix of Chapter (3)

Table A3-1 Descriptive statistics of the variables used in the estimations.

	N	mean	std.	min	max
	Panel A: Variables used in Dataset 1				
$\dot{y}$	1,624	0.0219	0.0407	-0.302	0.3210
y	1,615	8.3886	1.3072	5.3173	11.802
CREDIT	1,249	34.297	33.388	0.0100	205.99
Years of Schooling	1,584	5.9005	3.0634	0.0400	13.090
Investment Share	1,626	0.2065	0.1114	-0.013	1.6846
Inflation Rate	1,656	36.105	262.41	-6.628	6962.8
Openness	1,822	0.7600	0.4862	0.0199	4.3781
Government Consumption	1,626	0.2050	0.1180	-0.024	0.9337
Political Rights	1,624	4.0835	2.1948	1	7
Log Life Expectancy	2,016	4.1274	0.2001	3.0809	4.4224
Log Fertility Rate	2,018	1.2815	0.5508	-0.137	2.2134
TFP Growth	605	0.6444	2.5609	-15.036	16.381
	Panel B: Variables used in Dataset 2				
$\dot{y}$	588	0.0272	0.0306	-0.145	0.1759
y	588	10.8802	2.1518	3.2717	16.249
CREDIT	588	0.5018	0.4086	0.0036	2.1414
Investment Share	588	0.1079	0.0221	0.0281	0.1800
Log Secondary Schooling	582	2.8270	1.3769	-2.758	5.0821
Life Expectancy	588	58.319	7.1670	40.158	72.659
Knowledge	588	0.0678	0.2171	0.0000	2.3238
	Panel C: Instrumental variables				
Socialist Legal Origin	2,079	0.1640	0.3704	0	1
English Legal Origin	2,079	0.3333	0.4715	0	1
French Legal Origin	2,079	0.4180	0.4933	0	1
German Legal Origin	2,079	0.0317	0.1754	0	1
Scandinavian Legal Origin	2,079	0.0265	0.1605	0	1
Agricultural Share	588	0.1818	0.1600	0.0064	0.8565

Notes: Table reports the number of observations (N), the means, standard deviations (std.), minima (min.), and maxima (max.) of the variables used in our empirical specification as described in Section (3.4.1).

Table A3-2 The conditional effect of financial development on technological progress — Alternative indicators.

	OLS Estimates (1)	(2)	IV Estimates (1)	(2)
	Panel A: Total factor Productivity (Hall and Jones, 1999)			
CREDIT	-0.0255	-0.0254	-0.274***	-0.137***
	(0.0174)	(0.0166)	(0.0333)	(0.0337)
CREDIT × GDP	0.00338**	0.00321**	0.0270***	0.0142***
	(0.00163)	(0.00155)	(0.00333)	(0.00322)
Investment		-0.284		-0.350
		(1.140)		(1.510)
Years of Schooling		0.0243		-0.0285
		(0.0269)		(0.0366)
Political Rights		0.00198		-0.0851*
		(0.0448)		(0.0446)
Countries	108	99	108	99
R-Squared	0.36	0.35	0.00	0.53
F-Stat	48.24***	17.95***	34.55***	27.55***
	Panel B: Innovation Performance Index (EIU, 2009)			
CREDIT	-0.156***	-0.113***	-0.322***	-0.167***
	(0.0254)	(0.0252)	(0.116)	(0.0640)
CREDIT × GDP	0.0185***	0.0129***	0.0342***	0.0178***
	(0.00238)	(0.00238)	(0.0110)	(0.00593)
Countries	71	68	71	68
R-Squared	0.70	0.84	0.56	0.98
F-Stat	87.12***	69.11***	48.55***	859.7***
	Panel C: Patent Applications per Million (EIU, 2009)			
CREDIT	-8.655***	-8.536***	-58.18**	-38.19**
	(2.253)	(2.781)	(27.09)	(16.75)
CREDIT × GDP	1.025***	1.005***	5.729**	3.820**
	(0.253)	(0.292)	(2.589)	(1.616)
Countries	74	71	74	71
R-Squared	0.32	0.31	0.01	0.01
F-Stat	8.01***	4.72**	4.53**	3.10**

Notes: Dependent variables are total factor productivity estimated by Hall and Jones (1999), and two indicators compiled by EIU (2009), including an innovation performance indicator, and patent applications per million population. As this data is only available for a cross-section of country observations, the empirical models builds on OLS and IV, the latter replicating the instrumentation strategy of the baseline model. Robust standard errors are reported in parentheses. $*p < .10$, $**p < .05$, $***p < .01$

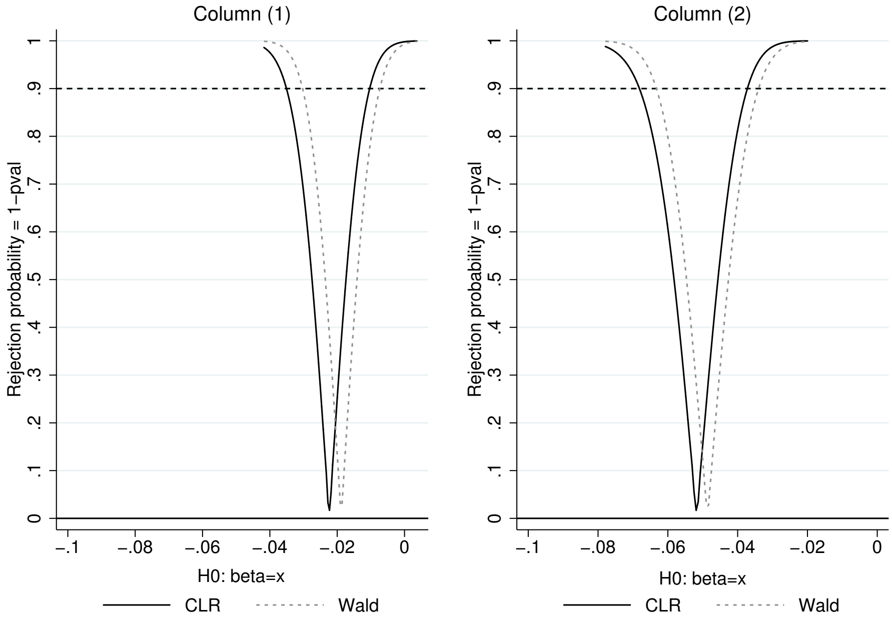


Figure B3-1 Weak instrument diagnostics of legal origin. Figure depicts weak-instrument robust confidence intervals and rejection probabilities implied by the conditional likelihood ratio test (CLR) of Moreira (2003) and the Wald test for the models reported in Columns (1) and (2) of the baseline specification in Table (2.1).

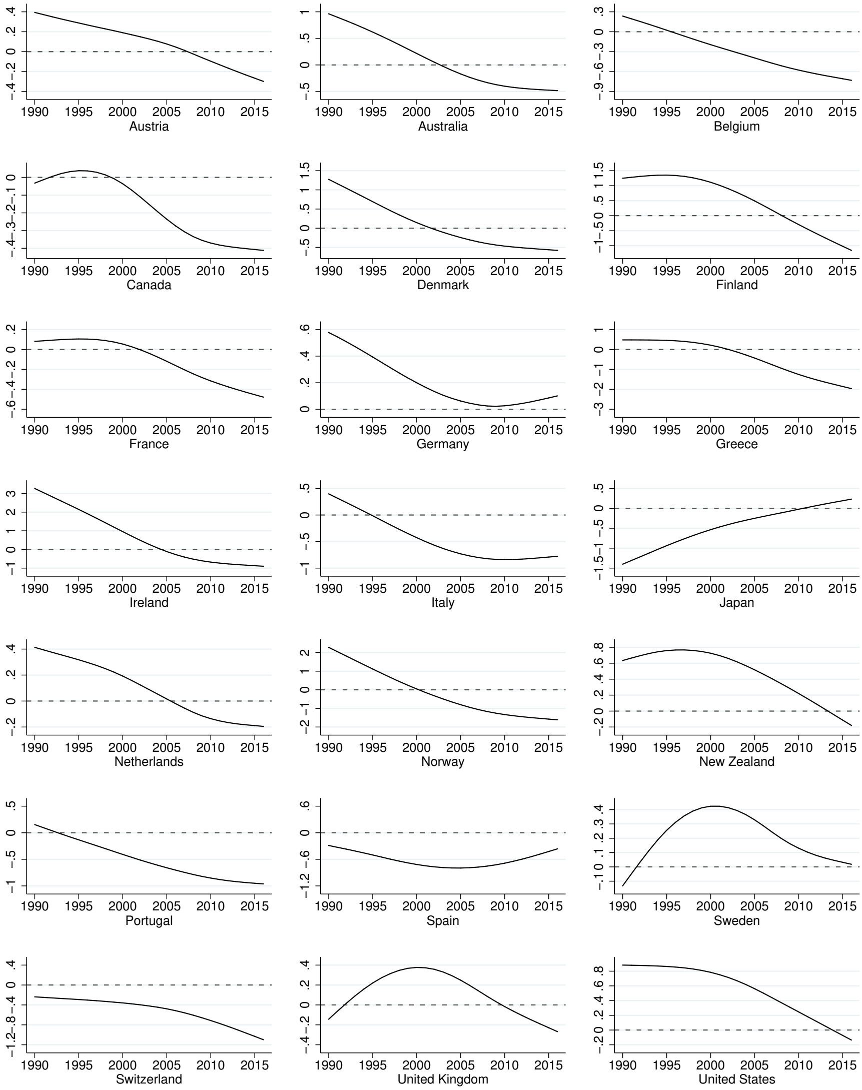


Figure B3-2 Trend in total factor productivity (TFP) growth for the countries included in Dataset 2, 1990–2016. The trend is estimated using the HP filter. Data source is TED (2017).

Chapter 4

Secular Stagnation in the Advanced Economies and the Declining Growth Rate of Germany[49]

Background The previous chapter showed that an increase in the size of the financial sector is particularly detrimental to growth in times when factor productivity grows at low rates. The reason is that financial institutions tend to steer away from their traditional intermediation activity if the potential for beneficial investments and innovation projects is low. In fact, productivity gains have considerably declined since the turn of the millennium in the overwhelming majority of advanced economies, yielding disproportionately low growth rates compared to the historical development of incomes during the last 70 years. This observation has led to renewed interest in the theory of "secular stagnation", a concept originally introduced in the late 1930s by Hansen (1939).

This chapter analyzes the causes of the reduction in growth observed in the advanced economies with a particular focus on the German economy. Germany has realized tremendous growth rates in the aftermath of World War II. Since the early 1970s, however, growth rates have declined and had settled down at a more or less constant rate of 2 percent per year, only to experience a renewed negative trend around the early 2000s. In many respects, the German situation can be thought of as a blueprint of the situation in the developed economies, as the results of this chapter imply that the below-average development of incomes is a result of global trends that affect most affluent countries equally.

4.1 Introduction: The global growth crisis

After World War II, the German economy has realized income increases that do not seem to be reproducible today. Figure (4.1) illustrates the development of the German growth rate and its trend from 1950 to 2010. While per capita income in the 1950s and the 1960s grew by an average of, respectively, 8.3 and 4.5 percent per year, these increases have declined significantly since the end of the 1960s. Between

[49] A previous version of this chapter has been published as Berthold and Gründler (2015).

the early 1970s and the late 1990s, growth rates settled down to a more or less stable level. However, since the beginning of the new millennium, the downswing of the trend has sharpened once more. During the 2000s, incomes grew by only 1.2 percent per annum.

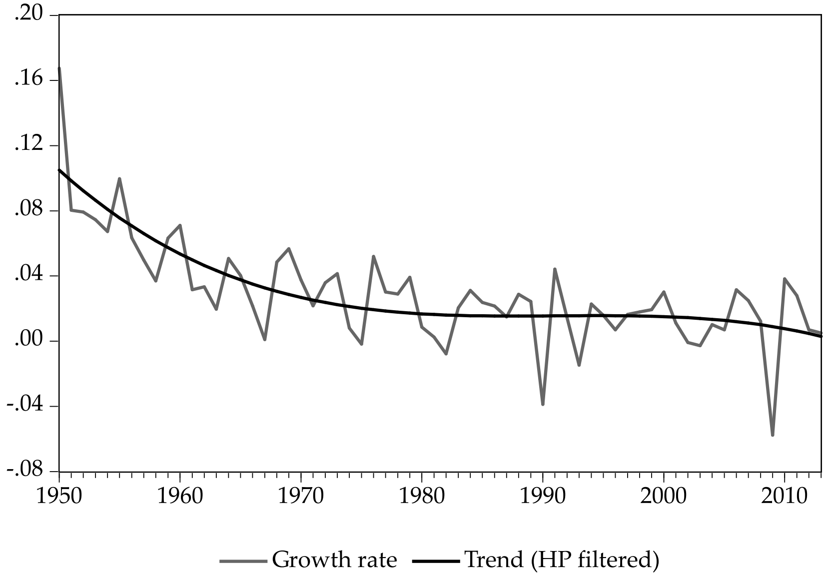


Figure 4.1 Per capita GDP growth rates and trend in Germany, 1950–2013. Data source is Heston et al. (2012), trend is calculated using the Hodrick-Prescott filter.

This phenomenon is not only specific to Germany, but can be identified in almost all developed countries. Figure (4.2) shows the evolution of the average growth rate of the 25 most affluent countries in terms of per capita GDP and its trend from 1970 to 2010.[50] Astonishingly, the development of per capita growth mimics the growth path of Germany quite accurately. After a period of high increases of per capita income during the 1970s, growth rates decreased, finally achieving a more or less stable level at the end of the 1970s. In the early 1990s, average growth in the top 25 increased tremendously. However, this increase was only of short duration. Around the year 2000, growth in the advanced economies exhibited a renewed negative trend. Figure (B4-1) in the appendix illustrates the trend of growth rates of each of the 25 richest countries in the world in terms of per capita income individually. With the exception of the oil-exporting countries Qatar, Kuwait and the United Arab Emirates, all of the top 25 countries experienced a more or less comparable decline in growth around the year 2000. Excluding oil-exporting economies, the mean value

[50] The group includes countries that were among the 25 richest nations in 2012 as measured by the classification of World Bank (2012b). Due to data availability, Qatar, Brunei Darussalam and the United Arab Emirates are excluded.

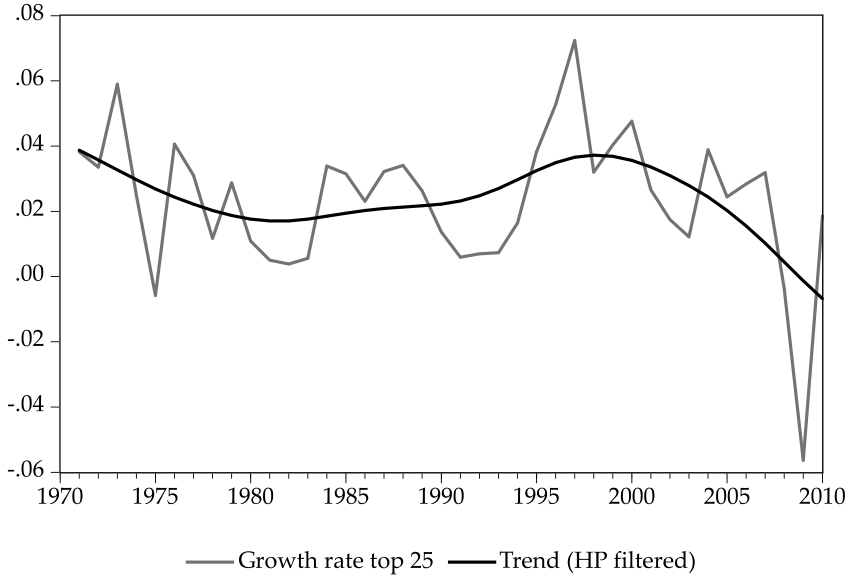


Figure 4.2 Average per capita GDP growth rates and trend in the developed economies, 1950–2013. Data source is Heston et al. (2012), trend is calculated using the Hodrick-Prescott filter.

of growth in the advanced economies was 3.1 percent in the 1970s, 2.3 percent in the 1980s, 2.1 percent in the 1990s, and 1.3 percent in the 2000s.

Recently, the phenomenon of declining growth rates across the advanced economies has renewed the interest in "secular stagnation", a concept initially proposed by Hansen (1939) and revisited by Summers (2014). Generally, secular stagnation describes a downward tendency of the real interest rate, which reflects an excess of desired savings over desired investment that eventually results in a decline in long-run growth. The theoretical explanations for why such a downturn might occur can be consolidated into four categories: (1) increasing saving rates due to conditional convergence, i.e. high growth rates in developing economies, (2) declining relative prices of investment goods, (3) declining rates of population growth, and (4) falling investment rates due to a dearth of attractive investment opportunities (see Eichengreen, 2015).

This chapter aims to provide an explanation for why the decrease in growth rates over the past 50 years and the particularly noteworthy decline since the start of the new millennium has occurred. The analysis follows Gordon (2015) by drawing on a supply-side explanation of secular stagnation. Whereas the focus of the analysis is on the developments observed in Germany, the findings suggest that the essential part of the explanation also holds for the group of advanced economies, transforming the German problem into a blueprint of the current situation in most advanced economies.

The chapter proceeds as follows: first, the main drivers of long-run economic growth are outlined from a theoretical viewpoint, where large parts are based on the growth theories discussed in Chapter (1.2). This consideration highlights the importance of technological progress and illustrates the role of human capital accumulation in the growth process. Subsequently, Section (4.3), is concerned with an empirical examination of the theoretical hypotheses, investigating the effects of technological change and human capital on economic growth in a panel of 187 countries analyzed between 1970 and 2010. The results underline the significant contribution of technological change to long-run growth. In addition, it turns out that human capital accumulation is a crucial factor in explaining historical growth rates. Likewise, conditional convergence emerges as a clear empirical pattern. In the case of Germany—and similarly the majority of the advanced economies—the post-war convergence effects following World War II did not fade until the late 1960s. This provides the important implication that the relevant time span to explore the origins of the recent growth crisis is 1970 to present.

In what follows, the analysis in Section (4.3) turns to a detailed examination of the evolution of human capital and technological progress in the advanced economies during the past decades. This synopsis shows that the growth potential provided by human capital are different across the industrial countries. In particular, the analysis implies that Germany lags behind the average of the developed countries with regard to education. As human capital is a direct input factor in the production function and is necessary for closing the technological gap—i.e. the transfer of scientific research into marketable goods and production processes—countries with below-average levels of human capital grow at lower rates. The second part is concerned with the analysis of technological progress in the world. Contingent upon the education level necessary for its implementation, the rate of technical change is the most important factor that affects long-run growth in the developed economy.

The exploration of technological progress during the last 200 years suggests strong fluctuations in the rate of technical inventions over time, with a considerable reduction observable since the turn of the millennium. The recent slowdown of worldwide innovation activity implies a lack of radically new ideas, which coincides with the reduction in long-run growth. Section (4.4) concludes.

4.2 Determinants of long-run growth: Convergence, human capital, and technology

The first step to examine the causes of the decline in growth is to analyze the determinants of long-run growth from a theoretical viewpoint. The growth theory of the past 60 years has proposed a number of important determinants that influence long-run improvements in living standards. Whereas country-specific characteristics are found to have a huge impact on economic development, there are some essential mechanisms that affect the wealth of individuals among different countries in a quite similar manner. In particular, the theory on economic growth emphasizes that the initial development level, educational achievements, as well as the increase of knowledge and its application are the most decisive factors in the development process (see Chapter (1.2) for a brief overview of the existing literature). For these reasons, a supply-side explanation of the current slowdown in growth must be made up of three parts. These are (1) conditional convergence, (2) human capital, and (3) technological progress.

Perhaps the most important hypothesis of the neoclassical growth model of Solow (1956), Swan (1956), Koopmans (1965) and Cass (1965) is that of convergence. The description of this model provided in Chapter (1.2) demonstrates that poor economies will eventually catch up to rich countries and that per capita incomes converge. Similarly, the model predicts that the growth rate of any economy declines as it approximates its steady state. However, empirical results indicate that the starting position and the growth rate are negatively correlated only when holding constant some variables that distinguish the countries. Therefore, reconciling the convergence hypothesis with data requires examining the concept of *conditional* convergence. In fact, the seminal contributions of Barro and Sala-i-Martín (1992), Mankiw et al. (1992), Barro and Sala-i-Martín (2004), and Barro (2003, 2013a) demonstrate that poorer countries do indeed grow faster than economies that have approached their steady state level of capital. Once this level is reached, human capital and technological progress become the main engines of growth (see, e.g., Romer (1986, 1987, 1990), Lucas (1988), Grossman and Helpman (1991), and Aghion and Howitt (1992, 1998)).

Consider first the case of human capital. The growth model of Hanushek and Woessmann (2012) emphasizes the importance to distinguish between the different factors that shape the skill level of workers. More specifically, the model considers human capital h to be a function of family input F, individual abilities A, schooling quality q, schooling quantity Y and other relevant factors Z that include health and labor market experience. This yields

$$h = \lambda F + \phi(qY) + \eta A + \alpha Z,$$

where $\lambda, \phi, \eta, \alpha \in \mathbb{R}^+$ represent the marginal impacts of the particular determinant. As family input and individual abilities can hardly—if at all—be improved, schooling quality and quantity as well as health and labor market experience are the most important factors that distinguish the advanced countries. The decision of individuals to invest in one unit of these factors of human capital at any time t can be described as (see, e.g., Johnes, 1993)

$$\int_0^Q C(t)\exp\{-rt\}\mathrm{d}t \leq \int_Q^T R(t)\exp\{-rt\}\mathrm{d}t. \tag{4.1}$$

$C(t)$ denotes the costs of achieving one marginal unit of human capital, $R(t)$ represents the rent of the educational program, Q refers to the time at which the training program is completed, and T marks the time of retirement. Thus, low costs of education, low interest rates, high returns to education, and a young population lead to positive growth stimuli in advanced economies via facilitation of educational achievements.

Closely related to the skill level of the workforce is the amount of knowledge that is available for individuals and firms. Therein lies the second main growth engine emphasized by the endogenous growth theory (see Chapter (1.2)). In fact, there is a strong relationship between the evolution of technological progress and human capital. The models of Nelson and Phelps (1966) and Benhabib and Spiegel (2005) illustrate the interaction between the technological frontier and factor productivity, where the rate at which the technological gap is closed depends on the level of human capital and the fraction μ of h working in the research sector. Let $T(t)$ be the theoretical technological level that measures the stock of knowledge or body of techniques available to individuals. Suppose that ω denotes the time lag between the invention and its adoption. On an aggregate level, the technology used in practice equals, on average, the technological frontier ω years ago. It follows that

$$\Psi(t) = T\left[t - \omega(\mu h)\right], \; \frac{\partial \omega(\mu h)}{\partial h} < 0.$$

Chapter (2) illustrated how the innovation process of entrepreneurs leads to an increase in $\Psi(t)$ if new capital goods j^* are introduced. Whenever the distance between the technological frontier and the presently implemented knowledge is high, the prospect to earn monopoly rents provides strong incentives for entrepreneurs to close the technological gap. To enable the creation of j^*, however, the technological frontier $T(t)$ must provide a sufficiently large potential to generate new blueprints. This raises the important question of how the technological frontier develops over time. An influential strand of the endogenous growth theory argues that $\partial T(t)/\partial(t)$ is subject to strong fluctuations over time (see, e.g., Jovanovic and Rousseau, 2005). The reason is that some ground-breaking technological improvements boost the technological frontier to a large extent, while others yield only little increases in $T(t)$. The impact of some technological inventions is so strong that they allow for a multiplicity of adoptions, and have a protracted influence on all industries. These technological advancements are called “general purpose technologies” (GPTs).[51] In the model of Nelson and Phelps (1966), the creation of these technologies leads to an instantaneous leap forward of the technological frontier. This widens the technological gap and creates a large potential for factor productivity gains, assuming that the stock of human capital is sufficiently large enough to master the new technologies.

Due to disembodied technological know-how flows, technological progress in one country also augments the technological frontier of a second country, given that these countries are in interaction with each other. International trade and openness therefore enhance the potential for factor productivity gains, which eventually translates to an increase in incomes. Regardless of the origin of technological advancements, growth potentials emerge through two channels: first, technological inventions increase the potential for productivity gains. These enhancements in productivity have a direct effect on per capita growth, as the output can be produced more efficiently. Second, technological inventions allow for marketable adoptions and thus trigger

51 Helpman and Trajtenberg (1994) characterize GPT’s as consisting of three major attributes: pervasiveness, an innovation-spawning effect, and scope for improvement.

indirect growth stimuli, which is illustrated in endogenous growth models with an expanding variety of products (see, e.g., the Romer (1990) model). These additional capital goods in turn enable improvements and variations κ_{j^*}, which further increase the output.

In fact, the process of improving existing j^* is a crucial growth determinant in the "Schumpeterian" growth models (see, e.g., Aghion and Howitt, 1992, 2009). Whenever a sufficiently large bundle of innovations j^* allows for a wide range of improvements κ_{j^*}, firms are likely to improve existing capital goods rather than invest in entirely new products. The reason is that both the cost and the risk accompanied by the creation of κ_{j^*} may often be substantially lower than in case of j^*. However, as improvements become more and more costly with each step up the quality ladder, the profits earned by creation of the $(\kappa + 1)$th improvement are lower than in case of the (κ)th variant. The decline in profits eventually provides incentives for entrepreneurs to invest in entirely new blueprints that have been made available by the technological frontier.

The diffusion process of a new technology j^* can be formulated more specifically by using an epidemic model of the functional form (see, e.g., Petsas, 2003)

$$\frac{\dot{\pi}(t)}{\pi(t)} = \phi\,[1 - \pi(t)]\,, \tag{4.2}$$

where $\dot{\pi}(t) \equiv \mathrm{d}\pi(t)/\mathrm{d}(t)$ gives the change in the fraction of industries using the new technology and ϕ denotes the rate of diffusion. More generally,

$$\pi'(t) = \phi\pi(t)\,[\Pi - \pi(t)]$$

describes the approximation to any upper limit Π.[52] Disregarding the time index for reasons of lucidity, the diffusion of technology in Equation (4.2) is very slow if π is small in relation to the upper limit. However, the speed of diffusion increases (approximately) exponentially up to a certain point. For high degrees of saturation, the marginal rate of diffusion converts to zero. The solution of Equation (4.2) provides the fraction of industries using a new technology as a function of time. This function possesses a sigmoid shape that can be described by

$$\pi = \frac{1}{(1 + \exp\{-(\gamma + \phi t)\})}, \tag{4.3}$$

where γ denotes the integration constant.[53] In $t \to \infty$, each industry uses the new technology. The first derivative of Equation (4.3) gives the growth rate of π. Consider that the new technology benefits each industry to the same extent and that within each industry, there is a comparable fraction of individuals to potentially carry out improvements or variations κ_{j^*}. Then the development of the amount of improvements follows the growth rate of diffusion π'. Figure (4.3) illustrates this process. Depending on the assumption, the functional form can be based on various

52 Since $0 < \pi(t) < 1$, the upper limit equals one in Equation (4.2).

53 See Sydsæter and Hammond (2008) as well as Petsas (2003) for the derivation of the solution for Equation (4.2).

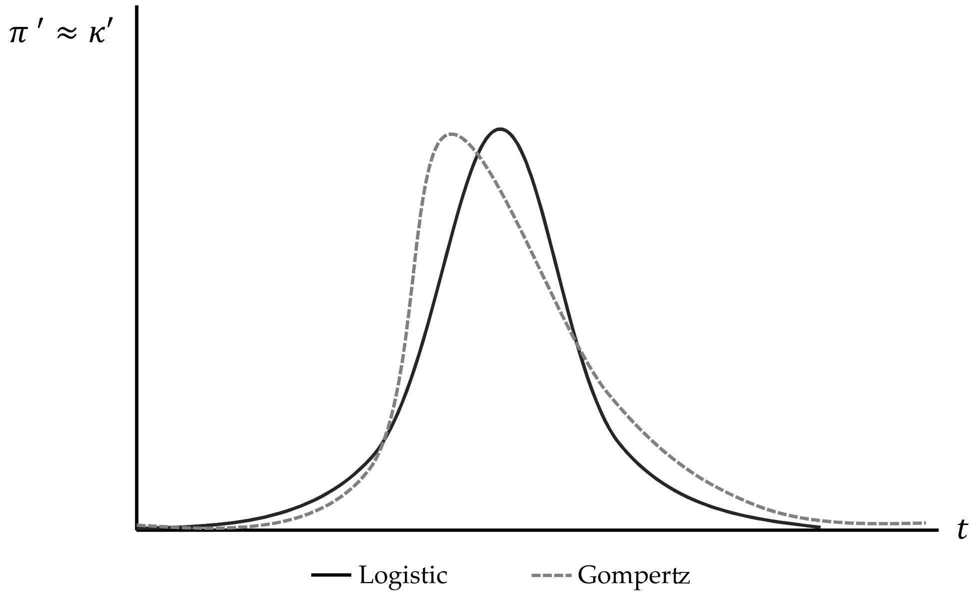


Figure 4.3 The rate of diffusion of new technologies, logistic and Gompertz function as functional relationships to model processes of technological diffusion.

members of the class of sigmoid functions. However, the theoretical considerations may present an argument for a right-skewed Gompertz specification. This is because the potential for various improvements may never run out completely, but rather become increasingly irrelevant as new technologies are introduced to replace the vintage ones.

The technological frontier $T(t)$ enables the introduction of j^*, i.e. it stimulates an increase in the number of capital goods (see Equations (2.1) and (2.2) of Chapter (2)). The improvements and variations following this introduction develop according to Figure (4.3). As described above, the extent of human capital is crucial in determining this process for various reasons, as human capital increases both the pace of technology adoption and innovation.

In sum, the long-run growth rate of a steady state economy is crucially determined by the theoretical technological level. As such, it may substantially vary over longer periods. If the technological gap is small, growth potentials are low. These situations occur whenever there is a lack of fundamentally new ideas. Romer (1993) describes such a situation as an "idea gap". The same result is obtained in the case of regular or even large technological gaps that are accompanied by poor rates of human capital accumulation. Intuitively, the position on the function illustrated in Figure (4.3) determines the long-run growth potential of the entire group of advanced economies, while differences in human capital account for different growth potentials of the individual countries.

4.3 Empirical evidence: Supply-side analysis of secular stagnation

4.3.1 The empirical effects of human capital and technology on long-run growth

The previous section underscores that the combination of human capital and technological change is decisive for long-run economic development. This section is concerned with an empüirical exploration of these effects. The specification of the empirical model builds on Barro (2000, 2003, 2013a) and Acemoglu et al. (2008). In these models, real per capita GDP growth is assumed to be a function

$$\dot{y} \equiv \frac{\mathrm{d}y}{\mathrm{d}t} = F(y_{t-\tau}, h_{t-\tau}, \Xi_{t-\tau}), \tag{4.4}$$

where $y_{t-\tau}$ is the (initial) logarithmic value of per capita GDP, $h_{t-\tau}$ denotes the (initial) stock of human capital, and $\Xi_{t-\tau}$ contains a number of environment and control variables suggested by the standard growth model and endogenous theories. Each regressor is lagged by $0 < \tau < 1$ periods. The analysis does not directly employ capital endowment since data on physical capital is quite unreliable, as its calculation depends strongly on arbitrary assumptions about depreciation and approximated values of both initial capital endowment and investment flows. The specification follows Barro (2003, 2013a) in assuming that higher levels of y and h are correlated with a larger stock of physical capital, so that the combination of both variables may serve as an appropriate proxy.

Controlling for some crucial growth determinants that distinguish the countries, the coefficient of $y_{t-\tau}$ gives the rate of conditional convergence. Conditionality is of great importance, as empirical growth research—as carried out, for instance, by Mankiw et al. (1992) and Barro (2003, 2013a)—suggests that the absolute convergence hypothesis of neoclassical growth theory cannot be confirmed empirically. In fact, the relation between the initial level of GDP per capita and the growth rate must be examined while holding constant some crucial variables that capture country-specific potential for economic growth. Likewise, Atkeson and Kehoe (2000) show for the two-country-case that the steady state is a function of initial conditions and thus varies between the economies. For these reasons, the incorporation of country-specific state and environment variables is compulsory.

Following Barro (2003, 2013a), Equation (4.4) is estimated using panel data from 187 economies between 1970 and 2010 in a simultaneous equation model (SEM). Each equation of the SEM covers a five-year period, so that each equation reflects a cross-sectional estimation at a certain point in time. Thus, the SEM comprises seven equations which are jointly estimated with GMM. This approach is determined by the long-term perspective of growth regressions, the need to smooth short-term fluctuations, and the availability of data. Estimating the influence of the variables on growth using annual panel data would lead to severe biases and a contradiction of the implications of growth theory. Yet, to guarantee comparability to some of the previous studies, the sensitivity analysis compares the results obtained via GMM estimations based on 5-year averages to the outcome of panel data regressions that draw on annual data.

More specifically, the empirical model is specified as

$$\begin{aligned}\dot{y}_{it} \quad = \quad & \alpha_0 + \beta y_{it-1} + \gamma_1 \text{YSCHOOL}_{it} + \gamma_2 \text{LIFEEX}_{it} + \gamma_3 \text{FERT}_{it} \\ & + \gamma_4 \text{DEM}_{it} + \gamma_5 \text{OPEN}_{it} + \gamma_6 \text{HOF}_{it} + \gamma_7 \text{GOVC}_{it} + \gamma_8 \text{INVS}_{it} \\ & + \gamma_9 \text{LATINAM}_{it} + \gamma_{10} \text{SUBSAH}_{it} + \rho_1 \text{PAT}_{it} + \rho_2 \text{CIT}_{it} + u_{it},\end{aligned} \tag{4.5}$$

where the growth rate of real per capita GDP $\dot{y}_{it}$ of country $i = 1, \ldots, N$ is estimated based on 5-year intervals $t = 1, \ldots, T$. The fertility rate (FERT) accounts for the negative effect of population growth on the steady-state ratio of capital to effective worker in the standard growth model. In addition, higher rates of fertility reflect a higher fraction of resources devoted to child-rearing. The investment share (INVS) incorporates the preference for saving and GOVC denotes government consumption. DEM is a dummy variable that assumes a value of 1 if the country is democratically organized and HOF is a rule of law index covering the extent of economic and political freedom. Countries with a high level of rule of law are connected with low values of HOF, so the coefficient of HOF must be interpreted reversely. In order to attend to the specific environments of Sub-Saharan and Latin American countries, Equation (4.5) includes dummy variables LATINAM for Latin American countries and SUBSAH for Sub-Saharan nations. Meanwhile, the degree of openness (OPEN) accounts for international spillovers and the gains from trade.

Holding constant these variables that are decisive for the country-specific growth potentials, the analysis in this section seeks to examine the role of human capital and technological progress for long-run development. Human capital in Equation (4.5) is approximated using average years of schooling (YSCHOOL) and life expectancy at birth (LIFEEX). Technological progress is measured by patent applications in relation to GDP (PAT) and citations (CIT). The latter comprises the number of citations achieved by patents granted in a country within the respective five-year interval. There is some advantage in using citations rather than patent applications, as citations reflect the quality of innovations more accurately. In addition, the tendency to register innovations at the patent office varies between countries, especially in heterogeneous samples that cover a large number of different nations. However, there is reason to suspect that "high-quality" patents are always registered, as the prospect of monopoly rents provides strong incentives for patent protection. The data sources of the variables are shown in Table (A4-1) in the appendix of this chapter.

The estimation technique of the basic regression is GMM using Newey-West HAC standard errors, as autocorrelation and heterogeneity are likely to occur due to the persistence in macroeconomic time-series as well as the heterogeneous sample of 187 countries. Many of the regressors in Equation (4.5) must be expected to be endogenous, which is why the analysis includes lagged variables of each regressor in the list of instruments. Including lagged instruments also reduces the possibility of a reverse causation, which is particularly likely in the case of education. Moreover, diminishing returns to education may disguise the effect of schooling. For these reasons, the lagged average number of years devoted to primary schooling (PSCHOOL) is included as a surplus instrument. In fact, there is a considerable variation with respect to PSCHOOL in the group of low-income countries. While pupils in countries with annual incomes less than 12,736 USD on average devote roughly 3 years to primary schooling, the standard deviation is high (1.8 years). Particularly in Afghanistan (1.6 years), Mozambique (1.0), Niger (1.1), and Yemen (1.7), primary schooling lags alarmingly behind.

Table (4.1) reports three versions of the baseline regression model, where each of the versions is further split into a first examination covering the whole sample and a second estimation including only middle and high-income countries as classified by the World Bank. Estimations that are only based on high-income countries would cause severe losses in degrees of freedom, leading to the possibility of biases in the estimations. When estimating the growth effect on middle and high-income countries, the regressions neglect the country dummies for Latin American and Sub-Saharan countries, as both regions are minimally represented in the high-income sub-samples.

Column (1) of Table (4.1) shows the results of the baseline specification in the whole sample of countries and focuses on the role of human capital. The analysis shows that fertility is negatively associated with growth, while a higher degree of openness and higher investment shares foster economic development. The influence of rule of law and democracy is somewhat more ambiguous. Whereas HOF does not affect growth at all in the baseline estimation, the democracy dummy possesses a negative sign. Acemoglu et al. (2008) discuss the influence of democracy on incomes in greater detail, suggesting that there may be a correlation but no causation between both variables. Furthermore, they show that historical factors appear to have shaped divergent development paths of economic well-being and political institutions. Likewise, Barro (1996) demonstrates that the favorable effects of democracy on growth include maintenance of the rule of law, free markets, low government consumption, and a high stock of human capital. Once these effects are held constant, democracy and growth are negatively correlated. Another explanation often brought forward in the literature is that the progress towards democratization enhances growth at low levels of political freedom, but depresses the rate of income increases once a moderate level has been achieved.[54] The most compelling explanation, however, is that a dichotomous measurement of democracy is not sufficiently detailed to capture the various institutional aspects of the countries. This pitfall is analyzed in detail in Chapter (6).

Controlling for the effects of the covariates, the outcome highlights a strong influence of human capital on economic growth. Both school attainment and life expectancy significantly contribute to increases in per capita incomes. This influence is quite robust, as there is little variation between the whole sample and the high-income estimation.

The second version of the baseline estimation is concerned with the influence of technological progress as measured by patent applications. The results show that

[54] See Barndt et al. (2005) for a historical perspective.

Table 4.1 Baseline regression results on the effect of convergence, schooling, and technology. Dependent variable is real per capita GDP growth.

	(1) Human capital		(2) Patents		(3) Citations	
	all	developed	all	developed	all	developed
y_{t-1}	-1.289*** (-19.25)	-1.711*** (-26.49)	-1.617*** (-28.43)	-2.324*** (-32.86)	-2.094*** (-32.48)	-2.914*** (-38.95)
YSCHOOL	0.062** (2.49)	0.060*** (2.87)	0.059*** (2.77)	0.112*** (6.00)	0.074*** (4.49)	0.113*** (7.22)
LIFEEX	0.037*** (3.94)	0.047*** (4.39)	0.063*** (6.00)	0.042*** (3.24)	0.011 (1.40)	0.041*** (4.96)
FERT	-2.388*** (-12.39)	-2.502*** (-16.10)	-2.429*** (-18.31)	-2.128*** (-15.40)	-2-947*** (-25.17)	-2.426*** (-26.20)
DEM	-0.224* (-1.70)	-0.342*** (-2.91)	-0.365*** (-3.05)	-0.144 (1.35)	0.525*** (-6.62)	-0.148* (-1.95)
OPEN	0.008*** (11.02)	0.008*** (13.99)	0.012*** (13.93)	0.014*** (16.00)	0.012*** (15.47)	0.015*** (20.01)
HOF	0.026 (0.87)	0.003 (0.10)	-0.006 (-0.23)	-0.135*** (-5.67)	-0.901*** (-4.02)	-0.094*** (6.16)
GOVC	0.001 (0.17)	0.009 (1.70)	0.003 (0.41)	-0.009 (1.20)	-0.023*** (-3.36)	-0.032*** (-5.77)
INVS	0.034*** (7.17)	0.032*** (5.32)	0.23*** (5.54)	0.030*** (6.42)	0.042*** (8.65)	0.024*** (5.80)
LATINAM	-0.556*** (-5.25)		-0.121*** (-1.40)		0.164*** (2.04)	
SUBSAH	-1.379*** (-7.08)		1.180*** (-6.68)		-3.020*** (-19.98)	
PAT			2.294*** (11.44)	2.585*** (10.84)		
CIT					0.193*** (13.15)	0.281*** (24.40)
Observations	794	512	602	438	448	388
R-squared	.09,.22,.08, .12,.28,.15, .16	.11,.35,.07, .12,.31,.13, .18	.15.21,.24, .16,.36, .20, .14	.28,.42,.36, .28,.46,.12, .18	.23,.42,.11, .24,.42,.30, .30	.15,.40,.25, .28,.44,.26, .41
Standard Errors	3.02, 2.77, 3.82, 4.80, 2.54, 3.17, 3.43	2.75, 2.30, 2.68, 3.27, 2.53, 3.17, 3.07	2.39, 2.66, 3.36, 4.86, 2.35, 2.85, 3.66	2.30, 2.32, 2.86, 2.98, 2.16, 3.00, 3.19	2.39, 2.09, 2.55, 3.03, 2.08, 2.43, 2.87	2.52, 2.21, 2.71, 2.90, 2.56, 2.68

Notes: Estimation method is OSGMM with HAC standard errors (Newey-West), t statistics are in parentheses. The dependent variable is real per capita GDP growth in the periods 1970–1975, 1975–1980, 1980–1985, 1985–1990, 1990–1995, 1995–2000, 2000–2005, 2005–2010. Due to data availability, the included periods in columns (2) and (3) are 1975–1980, 1980–1985, 1985–1990, 1990–1995, 1995–2000, 2000–2005, 2005–2010. The column labeled "'developed" refers to the sub-sample of middle and high-income countries. Instruments are lagged exogenous variables. Surplus instruments are primary school attainment (PSCHOOL) and the democratization index of Vanhanen (2012). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

countries with a high rate of patent applications in relation to GDP are able to achieve greater income increases than economies with little innovation activity. The influence of patents increases slightly when considering only middle and high income countries, but in general proves to be quite robust. Column (3) conducts a similar analysis, but measures technological advancement by the number of national citations rather than by patent applications. As in the previous case, growth is significantly related to innovation activity. The effect of citations is stronger in high-income countries, but similar to the regressions in Column (2), technological progress is positive and significant in both sub-samples.

The regressions in Table (4.1) also imply that technical knowledge does not necessarily need to be created domestically, but can also be imported via international trade activity. The positive sign of openness (OPEN) provides indication that spillovers from abroad also lead to factor productivity enhancements. The more open an economy is, the more it can benefit from internationally available technological knowledge and the more capital goods can be used in the production of the output.

In general, the system is relatively robust across the various estimations in Table (4.1). However, the incorporation of PAT and CIT leads to more plausible results with respect to government consumption and rule of law. Columns (2)–(3) demonstrate that public expenditures that do not directly affect productivity, but create distortions in the private sector, lead to a reduction in growth. In the models that neglect innovation activity, the influence of government consumption is virtually insignificant. One explanation for the change in the sign of government consumption may be that different forms of government consumption may have different effects on growth.[55] A quite similar result is achieved with respect to the rule of law index. In Columns (2)–(3), HOF exhibits a negative influence, indicating that rule of law increases growth rates due to the reverse coding of the HOF variable. In general, HOF and innovation activities are negatively correlated (between -10 and -20 percent depending on the measure and the period). There is much reason to expect rule of law to boost the number of patent applications, as the existence of property rights increases the incentives to invest in research activities.

Conditional convergence emerges as a clear empirical pattern in Table (4.1). In each of the regressions, the initial value of real per capita GDP significantly reduces growth rates. This means that after holding constant some crucial county-specific variables, economies with a low initial income level tend to grow faster than richer economies. Thus, poor economies will eventually catch up with prosperous countries and per capita incomes will converge asymptotically, even though divergent steady state levels are likely to prevent the achievement of identical income levels. The rate of convergence in Table (4.1) lies between -1.289 and -2.914 and is stronger when considering only middle and high-income countries. Similar rates of conditional convergence have been reported in Barro (2003, 2013a), where the coefficient assumes values between -2.11 and -2.90. The strong effect of convergence also implies that advanced economies whose capital stocks abruptly decline as a result of war or natural disasters are very likely to experience above-average growth rates while re-approximating their steady states.

[55] Note that government consumption as measured by Heston et al. (2012) includes a large variety of different governmental activities, such as education expenditures and expenditures for the social system.

4.3.2 Sensitivity analysis

One concern of the baseline results may be that the findings are strongly influenced by the specification of the system based on 5-year averages and the utilized GMM estimator. For this reason, Table (4.2) investigates whether the outcome changes when applying different estimation strategies. The sensitivity regressions are divided into two branches: the first class of models, illustrated in Column (1), maintains the basic specification of Equation (4.5) but applies different estimators, W2SLS and 3SLS. The second class, reported in Columns (2) and (3), explores the influence of the growth determinants in a dynamic panel based on annual data. The table reports two different forms of dynamic panel regression, GMM with Newey-West HAC standard errors and Within-Group regressions.

The regressions using alternative estimators in the basic simultaneous equation model yield results highly comparable to those of the baseline outcomes. Conditional convergence again plays an important role in the evolution of growth rates, indicating that poor countries tend to catch up with advanced economies when holding constant some variables that distinguish the countries. Supporting the baseline results, Column (1) illustrates that education and innovation activity positively contribute to growth, although the effects are slightly less pronounced in case of 3SLS. In addition, life expectancy and the investment share decrease in importance in both the W2SLS and the 3SLS estimates. As in the baseline output, the degree of openness exerts a significant influence on growth rates, indicating that international spillovers affect welfare increases to a large degree.

The estimations thus far use 5-year averages in order to smooth cyclical fluctuations. It can reasonably be argued that averaging the data is crucial when investigating the long-run determinants of economic growth. In addition, modeling long-run relationships with lagged variables helps to address the econometric problems that occur when dealing with growth rates, endogeneity, and autocorrelation. Nevertheless, estimating the baseline model with annual data provides an advantageous robustness check of the basic findings. To disentangle cyclical effects from long-run growth, the specification uses 5-year moving averages of the growth rate as the dependent variable. In order to ensure comparability to the baseline regressions, the time lag used with respect to $y_{t-\tau}$ and the instruments is five years.

Both dynamic panel estimations yield highly comparable results and strongly support the findings of the baseline results. In contrast to the regressions in Column (1), the positive effect of human capital accumulation becomes clearly visible, as both school attainment and life expectancy significantly contribute to income increases. Moreover, Column (2) supports the earlier findings that innovations and international spillovers significantly influence growth rates. The annual estimates further confirm the conditional convergence hypothesis. However, when including fixed effects, the marginal impact of the initial income value is notably stronger. The comparison between the Within-Group model and the panel GMM estimation also reveals that individual country effects are responsible for large parts of income increases, as the inclusion of fixed effects allows an explanation of 63 percent of the variation in growth rates, whereas R-squared in the GMM model without fixed effects is only 25

Table 4.2 Sensitivity analysis of the baseline results. Dependent variable is real per capita GDP growth.

	(1) 5-year panel		(2) Annual panel: patents		(3) Annual panel: citations	
	W2SLS	3SLS	Panel GMM	Within-Group	Panel GMM	Within-Group
y_{t-1}	-1.521*** (-7.85)	-1.684*** (-8.42)	-1.873*** (-15.75)	-7.443*** (-27.97)	-2.225*** (-11.65)	-7.811*** (-6.47)
YSCHOOL	0.137* (1.90)	0.579 (0.79)	0.073* (1.87)	0.647*** (8.30)	0.131** (2.46)	0.386*** (3.39)
LIFEEX	0.003 (0.09)	0.043 (1.26)	0.085*** (4.09)	0.165*** (5.98)	0.150*** (4.52)	0.253*** (2.89
FERT	-3.131*** (-6.59)	-2.990*** (-5.86)	-3.843*** (-14.70)	-1.316*** (-3.51)	-2.542*** (-7.73)	-2.305*** (3.95)
DEM	-0.219 (-0.55)	-0.186 (-0.46)	5.866*** (3.36)	-0.006 (-0.03)	0.246 (1.08)	-0.844*** (-4.38)
OPEN	0.011*** (3.74)	0.012*** (3.61)	0.008*** (5.53)	0.026*** (10.14)	0.009*** (5.12)	0.034*** (9.58)
HOF	0.073 (0.80)	0.037 (0.39)	0.051 (1.13)	-0.019 (-0.38)	-0.163*** (-2.77)	-0.078 (-0.95)
GOVC	0.005 (0.19)	0.015 (0.55)	-0.63*** (-3.61)	-0.267*** (10.60)	0.004 (0.18)	-0.186*** (-3.39)
INVS	0.011 (0.68)	0.018 (1.11)	-0.040*** (-2.96)	0.171*** (18.81)	0.033* (1.89)	0.194*** (13.00)
LATINAM	0.005 (0.01)	-0.176 (-0.53)	-0.504*** (-2.77)		-1.126*** (-4.62)	
SUBSAH	-0.494 (-0.93)	-0.949* (1.78)	-0.190 (-0.62)		-1.126** (-2.15)	
PAT	2.501** (2.515)	2.100** (1.96)	0.150** (5.28)	0.206*** (5.70)		
CIT					0.052* (1.76)	0.001 (1.11)
Observations	602	602	2,214	2,214	1,140	1,140
R squared	.19, .34, .30, .08, .40, .15, .02	.37, .56, .13, .34, .56, .36, .41	.25	.63	.30	.63
Standard Errors	2.33, 2.44, 3.44, 5.01, 2.27, 2.93, 3.89	2.05, 1.99, 2.81, 4.31, 1.96, 2.55, 3.02	2.56	1.93	2.58	1.96

Notes: Table reports W2SLS, 3SLS, Panel GMM, and Within-Group estimations. The statistical appendix provides a detailed description of the applied techniques. The t statistics are documented in parentheses. The dependent variable is real per capita GDP growth in the periods 1970–1975, 1975–1980, 1980–1985, 1985–1990, 1990–1995, 1995–2000, 2000–2005, 2005–2010. Due to data availability, the included periods in columns (2) and (3) are 1975–1980, 1980–1985, 1985–1990, 1990–1995, 1995–2000, 2000–2005, 2005–2010. The dependent variables in Columns (2) and (3) are 5th order moving averages of annual GDP growth. Instruments are lagged exogenous variables. Surplus instruments are primary school attainment (PSCHOOL) and the democratization index of Vanhanen (2012). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

percent. Column (3) utilizes citations rather than patent data as a proxy of innovation activity. While the basic estimates strongly resemble the outcome of Column (2), CIT is only significant in the panel GMM estimation.

Overall, the sensitivity analysis supports the findings of the baseline regressions. As predicted in the theoretical section, human capital, innovation activity, and international spillovers turn out to be robust determinants of the growth rate. Conditional convergence, meanwhile, emerges as a clear pattern, illustrating that countries with capital stocks below the steady state level are able to grow at a much faster rate. The results suggest that once the steady state is reached, human capital and innovation activity become the main drivers of growth.

4.3.3 The link between human capital and innovations

The regression results demonstrate that human capital and innovations critically influence the steady state rate of growth, and it should be noted that these factors do not evolve independently from one another. In fact, there is a close connection between the two determinants. Figure (4.4) shows the average annual TFP growth between 1960 and 1995 and the technological gap in 1995 in a sample of 84 countries.[56] When consulting data on patent application, the most innovative country in the post-2010 period is the United States (527,059 applications), followed by Japan (339,610), South Korea (185,632), and Germany (60,799).[57] Due to the position of the United States as the technological leader in the world, the theoretical technological frontier is set equal to the total factor productivity of the US. The technological gap is the logarithmic distance of the TFP of country i in relation to the TFP of the US, that is

$$T^g \equiv \log\left(\frac{TFP_i}{TFP_{USA}}\right).$$

In many respects, this examination is similar to the catch-up hypothesis originally proposed by Gerschenkron (1962). Figure (4.4) illustrates that nations that have been able to close their technological gap did a much better job of establishing factor productivity growth. Countries that have not been able to catch up with the technology of the US—such as Tanzania or Zambia—possess very poor TFP growth rates. The correlation between the variables in Figure (4.4) is high (67 percent).

Additionally, Figure (4.5) demonstrates the relationship between the average human capital endowment in 1960 and annual factor productivity growth between 1960 and 1995. The time-lag is introduced to rule out the possibility of a reverse causation. As in the empirical analysis, human capital is proxied using the average years of schooling. The correlation between the two variables is clearly positive, suggesting that economies with a larger initial stock of human capital succeed in generating higher rates of factor productivity growth. However, the variation around the regression line is high. While Asian nations such as Taiwan, South Korea, Thailand,

56 The selection of the countries and the time period is determined by the availability of data.

57 Numbers refer to patent applications of both residents and non-residents and reflect the average values of the 5-year period 2010-2014. Data source is World Bank (2016).

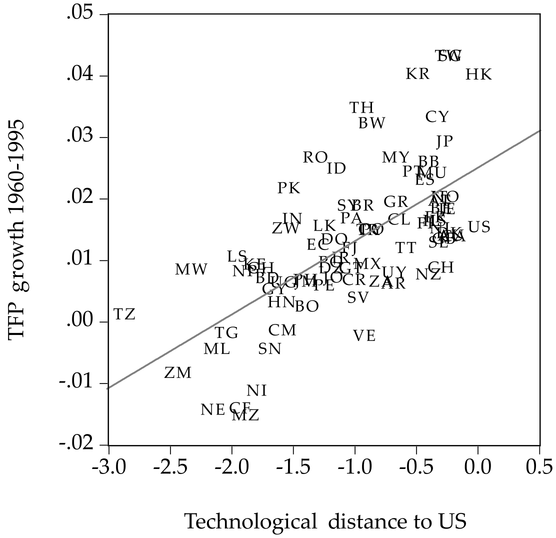


Figure 4.4 The relationship between the technological gap and factor productivity growth. Data source is Benhabib and Spiegel (2005) and own calculations.

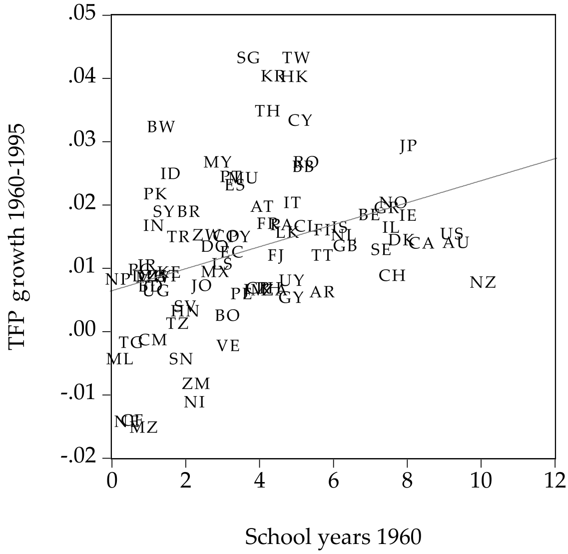


Figure 4.5 The relationship between human capital and total factor productivity growth. Data source is Benhabib and Spiegel (2005), Barro and Lee (2013), and own calculations.

and Hong Kong have been able to realize above-average TFP growth rates, some South and Central American nations—e.g. Venezuela and Nicaragua—as well as a variety of African states including Zambia, Mozambique, and Niger have experienced poor TFP increases. Overall, the correlation between TFP growth and initial human capital is 36 percent.

The connection between human capital and technological progress is crucial when analyzing the evolution of these factors in particular countries, as Figures (4.4) and (4.5) indicate that the possibility to substitute both determinants are limited. In general, progress on the technological frontier will benefit long-run growth, but it will have a lower impact in countries with relatively low human capital endowment.

4.3.4 Post-war convergence in the advanced economies

Conditional convergence emerges as a distinct empirical regularity in each of the regression results in Tables (4.1) and (4.2). When aiming to explain the evolution of growth rates in countries whose capital stocks have been depleted during World War II—such as many European and Asian countries—convergence can be seen to play an important role. This particularly holds for the German capital stock. In the aftermath of the war, German production was only 38 percent of the prewar level.[58] In accordance with the prediction of the standard growth model and the empirical evidence on conditional convergence, Germany subsequently experienced tremendously high growth rates of GDP. Figure (4.6) demonstrates how GDP growth has developed as the capital stock gradually recovered. The abscissa a illustrates the logarithmic distance of output in period $t \in [1947, ..., 1970]$ in relation to the prewar level in 1938, that is

$$a = \log\left(\frac{GDP_t}{GDP_{1938}}\right).$$

The ordinate gives the associated growth rates in t. The correlation is strongly negative (-83 percent). This indicates that growth rates were exceptionally high when the capital stock was heavily depleted. Yet as output approximated its prewar level, the growth rates declined. By the end of the 1960s, the effects from convergence had dissipated. This result explains the first major decline in German growth rates around the year 1970. By the 1970s, German output had fully recovered from the effects of World War II. The average German growth rate between 1918, the end of World War I, and 1939, the beginning of World War II, was 2.7 percent per year.[59] Assuming that without the war the German economy would have continued to grow at this rate, it is easy to calculate a hypothetical "non-war" growth path. Comparing the realized output with this hypothetical path shows that production has caught up with the hypothetical level in 1971.

However, not only Germany, but also a number of other advanced economies experienced similar convergence effects after World War II. Figure (4.7) shows the

[58] Data source: Maddison (2013).
[59] Data source: Maddison (2013).

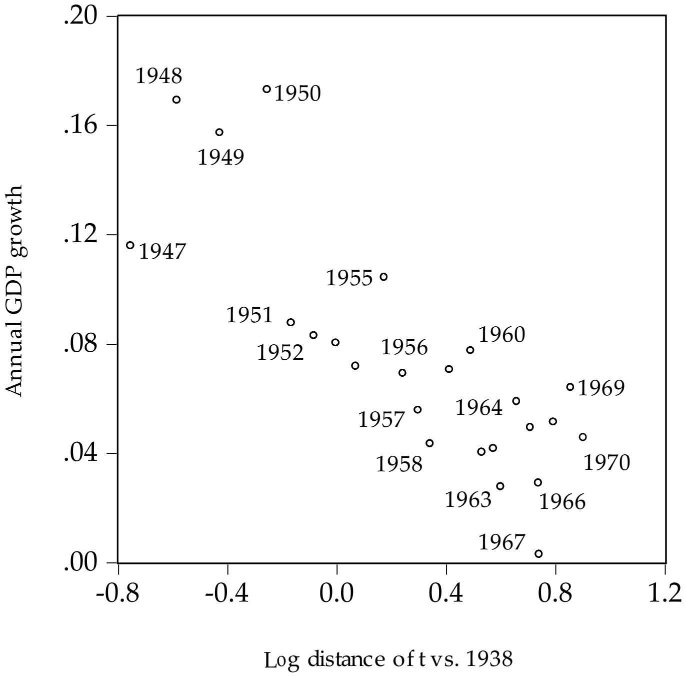


Figure 4.6 Postwar convergence in Germany. Destruction of the capital stock and growth rates, 1947–1970. Data source is Smolny (2000), Maddison (2013), and own calculations.

development of the logarithmic value of real per capita GDP in Germany, France, Japan, and Italy between 1900 and 2010. The figure demonstrates a severe reduction in output in the aftermath of the war. However, very much resembling the case of Germany, the economies of France, Japan, and Italy exhibited extraordinarily high growth rates, which emerged due to convergence effects. The figure also shows the slowdown in growth that was initiated around the year 2000. The countries illustrated in Figure (4.7) may stand exemplary for a large number of European countries that have been affected by World War II. In fact, a quite similar development can be found in most industrial nations in Europe.

In sum, this section suggests that the analysis of the current growth crisis may only take into consideration the period from 1970 onward, as growth rates in Germany and other advanced economies in Europe and Asia were strongly influenced by convergence effects that benefited income increases during the 1950s and the 1960s.

4.3.5 Growth in the advanced economies

In order to better understand the growth potentials implied by human capital and innovation activity that are discussed in the next sections, it is worthwhile to first take a closer look at the relative positions of the advanced economies in terms of historical GDP growth. When considering real per capita GDP, the 25 richest countries in

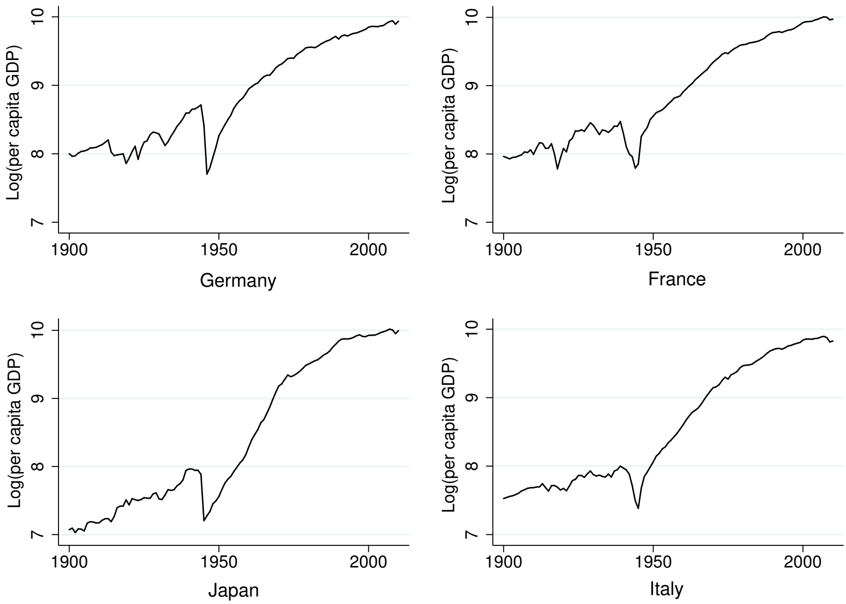


Figure 4.7 Postwar convergence in Germany, France, Japan, and Italy. Depicted is the development of $\log(\text{GDP}_{pc})$ between 1900 and 2010. Data source is Bolt and van Zanden (2014).

2012 were (in descending order) Luxembourg, Qatar, Singapore, Norway, Kuwait, Brunei, Switzerland, the United States, the United Arab Emirates, the Netherlands, Austria, Ireland, Sweden, Denmark, Canada, Australia, Germany, Belgium, Finland, Iceland, Equatorial Guinea, the United Kingdom, France, Japan, and Italy.[60] This sample includes some countries whose wealth is based primarily on oil resources, such as Brunei, Equatorial Guinea, Qatar and Kuwait. In order to investigate inherent growth mechanisms, there is a strong argument for exclusion of these countries in the subsequent consideration.

Figure (4.8) shows the average annual growth rates of the top 25 economies within the periods 1970–2010 and 1990–2010. Neglecting the oil-exporting countries, the average growth rates of the top 25 were 1.92 percent (1970–2010) and 1.49 percent (1990–2010), respectively, for the two periods. Growth in the developed countries declined in nearly every economy. Only six of the top 25 nations succeed in growing at an average rate of 2 percent or higher between 1990 and 2010, two of them primarily as a result of their having rich oil reserves. Clearly, the phenomenon of declining growth rates is not a problem specific to Germany, but can be identified in almost all developed countries. Within the group of the top 25, German growth rates correspond almost exactly to the mean (1.83 percent in 1970–2010 and 1.39

[60] This classification is based on data of World Bank (2012c) and considers the average of per capita GDP between 2005 and 2011.

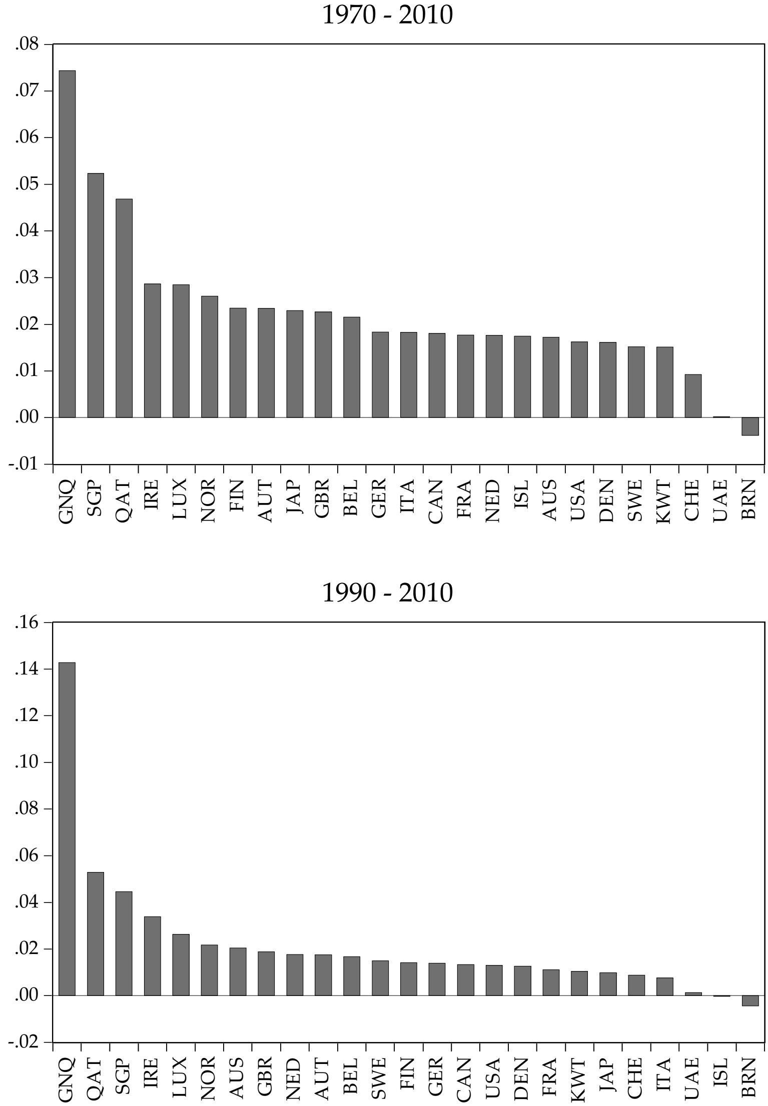


Figure 4.8 Average annual real per capita GDP growth rates of the top 25 economies. Periods 1970–2010 and 1990–2010. Data source is Heston et al. (2012).

percent in 1990–2010). The t-test indicates that there is no statistically significant difference between the German growth rate and the mean value of the top 25.[61]

4.3.6 Human capital in the advanced economies

The analyses in the previous sections imply that human capital and technological progress are the major drivers of income increases in developed economies. In the following, both determinants are explored in greater detail, beginning in this section with the examination of the current level of human capital. One important dimension of human capital is education. In their study of growth accounting between 1890 and 1970, Goldin and Katz (2007) suggest that the improvement in educational attainment in the United States contributed 0.35 percent points per year to real GDP per capita growth. Yet, it is not only school attainment, but also schooling quality which influence economic growth. Particularly in the sample of developed countries where school attainment is highly comparable, schooling quality distinguishes the countries. Figure (B4-2) in the appendix of this chapter illustrates the link between schooling quality as measured by PISA (Programme for International Student Assessment) scores, TIMSS (Trends in International Mathematics and Science Study) scores and the index of cognitive skills of Hanushek and Woessmann (2012). There is strong evidence that the quality of education determines the level of per capita incomes, supporting the results of the regressions in Tables (4.1) and (4.2).

Figure (4.9) shows the differences in schooling quality among the advanced economies. The dotted line marks the median value of the sample. When analyzing test-scores, there is always a trade-off between the evaluation of future potential and current abilities. More recent test scores reflect the abilities of students who are still in school, and their average skill level can easily deviate from that of their parents. This implies weighting the measures more towards students and less towards workers. If a measure seeks to assess the skill level of the workforce, then there is no choice but to use older data. Bearing this important difference in mind, Figure (4.9) illustrates recent scores obtained from the PISA waves of 2009 and 2012, as well as from the fifth TIMSS cycle collected in 2011. These measures reflect the potential human capital endowment of future workers. The figure further shows the cognitive skills index from Hanushek and Woessmann (2012), which provides insights into the skill level of the workforce. The index uses data from international test scores achieved in math and science between 1964 and 2003, and normalizes the data to match the PISA norm of mean 500 and standard deviation 100, divided by 100.

The data shows that the average cognitive skill level of the workforce is particularly high in Japan, Switzerland, Singapore, Finland, Australia and Belgium. In contrast, the skill levels in Germany, Denmark, Norway, and the United States are below the median value of the advanced economies. These countries are shaped by a high disparity of education outcomes. For instance, when accounting only for the share of top-performing students, the cognitive skills index shows that Germany (10.5 percent) ranks considerably above the median (8.80 percent). This suggests that

[61] The test gives $p = .6156$.

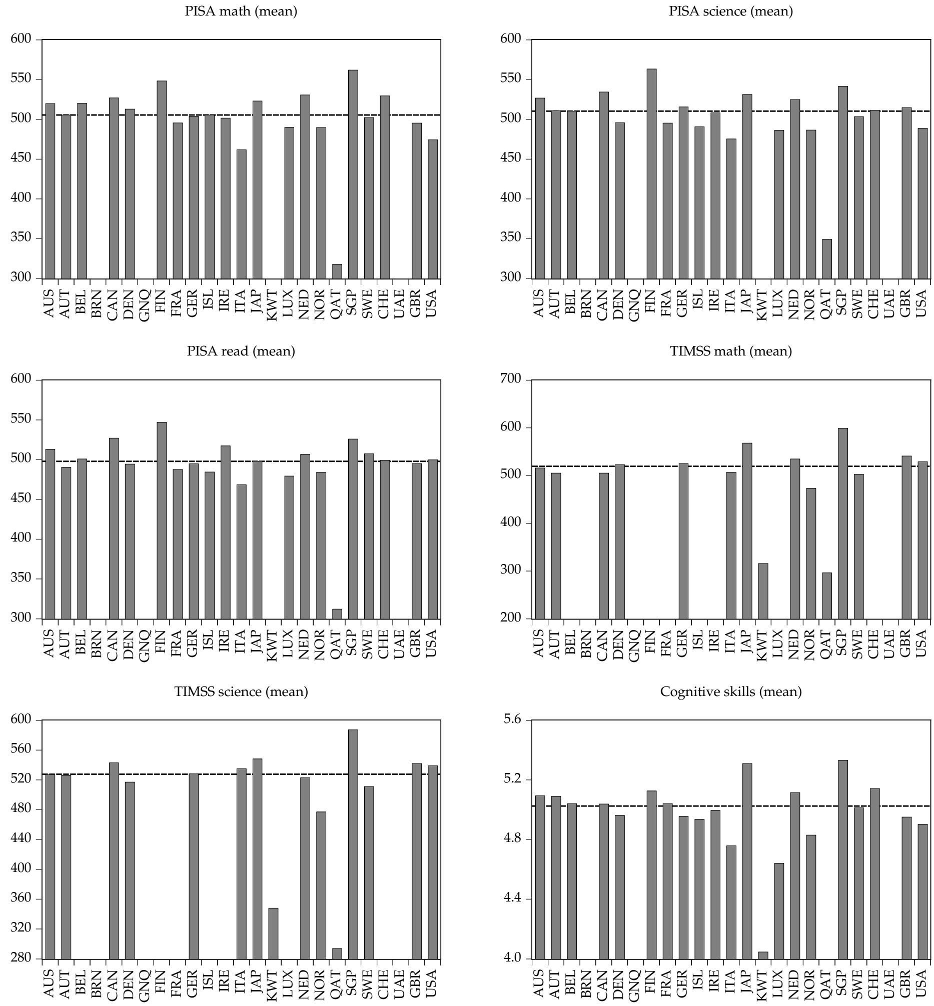


Figure 4.9 Schooling quality in the advanced economies. Data source is Hanushek and Woessmann (2012), World Bank (2012a), and own calculations. Reported are scores obtained from the last two PISA waves, conducted in 2009 and 2012. TIMSS data are based on the fifth cycle, collected in 2011.

Germany still has an elite of highly educated individuals, but the mean value of education lags significantly behind this level, reflecting high educational inequality.

When considering the scores obtained in PISA and TIMSS tests that evaluate future potential, the picture changes only slightly, suggesting that the educational systems in Singapore, Finland, Canada, and Japan are particularly successful. Whereas some of the scores in science and math give cause for optimism, Germany, the United States, and the United Kingdom lag behind the median education level in nearly each category.

Does this indicate that these countries invest too little in education? Figure (B4-3) in the appendix illustrates that educational expenditures in percent of GDP in Germany are considerably below the median. It is noticeable, however, that countries such as Singapore and Japan, whose test results turn out to be very positive, invest a relatively small share of GDP in the education sector. The United States, on the other hand, invest more substantially in the education system than the median. The benefits to education are nevertheless below average. These results suggest that a mere increase in educational expenditures does not necessarily lead to an improvement in schooling quality. In general, the correlation between education expenditures and schooling quality is moderate (PISA math: 36 percent, PISA science: 27 percent).

Whereas the future potential for human capital is distributed unequally across the advanced economies, the case of Germany reflects a situation in which the deficits in human capital accumulation are very likely to increase in the future. This becomes clear when considering some of the crucial variables that influence the individual decision to invest in human capital which are implied by Equation (4.1) of Section (4.2). First, the German population on average is relatively old. Figure (B4-3) in the appendix shows that Germany has the second highest fraction of people aged 65 or older (20.6 percent), surpassed only by Japan (23.4 percent). Second, the fertility rate in Germany (1.39) is significantly below the median value (1.87) of the top 25. As indicated in the theoretical section, the decision of an individual to invest in human capital depends on the length of the period during which the individual receives education rents. As the population ages, these incentives decline. This is currently the case in Germany, where the aging of the population reduces the average incentive to invest in human capital. Likewise, the decline in the fertility rate (see Figure (B4-3) in the chapter appendix) reduces the potential for future increases in the human capital stock. Human capital is thus likely to accumulate at an even lower rate in the future, which decelerates economic growth in the long-run.

The empirical results of Table (4.1) furthermore suggest that health is an important dimension of human capital. The most common indicator of health is life expectancy. Despite its aging population, the average life expectancy in Germany is relatively low and below the median of the top 25 (see Figure (B4-3) in the appendix). Citizens of countries such as Australia, Canada, Iceland, Japan and Switzerland are on average much healthier. The oil-exporting countries prove to be major outliers in this statistic, as incomes are highly unequally distributed among the population in

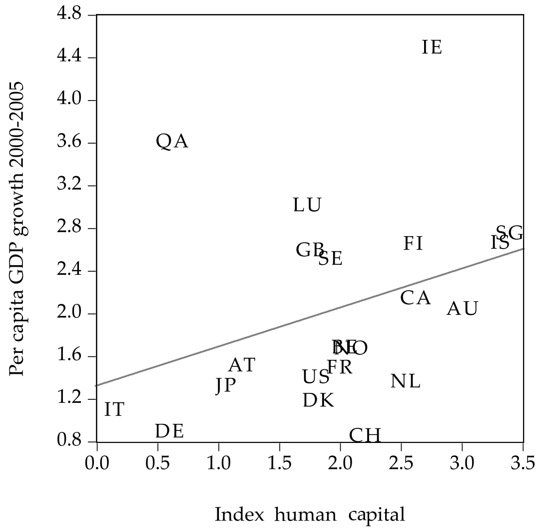


Figure 4.10 The relationship between the human capital index and real per capita GDP growth, 2000–2005. Sample includes the top 25 economies.

these countries. For example, the Gini coefficient of net incomes in the United Arab Emirates is about 46.72 percent.[62]

The previous observations illustrate that human capital consists of an array of different elements, among which schooling quality is only one. Combining schooling quality with health, fertility, and demography allows for computation of a more detailed index of human capital. Let μ_x and σ_x^2 be the empirical mean and variance of a given variable x. Utilization of these empirical moments enables calculation of an index of human capital $\tilde{h}$ that covers the sum of the normalized distances to the mean value of the key dimensions of human capital as follows

$$\tilde{h}_i = \left[\frac{\text{FERT} - \mu_{FERT}}{\sigma_{FERT}} + \frac{\text{LIFEEX} - \mu_{LIFEEX}}{\sigma_{LIFEEX}} + \frac{\text{MAT} - \mu_{MAT}}{\sigma_{MAT}} - \frac{\text{OLD} - \mu_{OLD}}{\sigma_{OLD}}\right] - \min(\tilde{h}_i).$$

The index is based on four dimensions. These are the fertility rate (FERT), life expectancy (LIFEEX), the mean value of mathematical skills as measured by PISA (MAT), and the fraction of the population that is of age 65 or older (OLD). Normalization of the particular distances ensures that each determinant contributes to $\tilde{h}$ with the same weight, regardless of the underlying scaling. The index is furthermore adjusted by the minimum value of the countries in the sample to fit the domain $(0, \infty)$. Figure (4.10) plots this measure against the average rate of real per capita GDP growth in the period 2000–2005. The time span ends in 2005 in order to prevent the distortive effects of the financial crisis from influencing the examination.

[62] See UTIP (2012).

Due to data availability, Figure (4.10) does not include all of the top 25 economies. The figure highlights strong potential for future human capital accumulation in Finland, Canada, Singapore, the Netherlands, and Iceland. In contrast, the growth potential suggested by human capital in Germany, Italy, and Austria is much weaker. The regression line marks the average growth rate implied by the human capital potential in the sample. While the slope of this line is positive, there are some considerable outliers to be observed: Ireland and Qatar were able to boost production by a rate that substantially exceeds the theoretically possible increase, given their stock of human capital. As for Qatar, the outstanding growth rates are mainly due to the export of oil. In contrast, countries such as Germany, Japan, Denmark, the United States, and Italy grew at slower rates than implied by their potential. However, even if Germany could have managed to realize the full growth potential emerging from the human capital stock, per capita GDP growth would still lag behind the rest of this group. The reason is the potential itself: the low fertility rate, the below-average test scores, and the aging population lead to a comparatively low value of $\tilde{h}$. None of these factors can be expected to be significantly improved in the medium term. Human capital induced growth potentials in Germany are thus likely to remain weak in the short to medium term.[63]

Aside from the average level of schooling, educational inequality is a decisive factor for growth potentials. An equally educated population is crucial for closing of the technological gap and for creation of technological advancements. In fact, the analyses in Chapter (2) imply that entrepreneurship can only unfold its full potential if the individuals possess an education level sufficient to implement new technological inventions. Thus, a high level of educational inequality is harmful for long-run growth.

The schooling dataset of Barro and Lee (2013) enables computation of a Gini index of educational inequality. Let $\bar{H}$ be the average number of years the population aged 15 or older has devoted to education. Suppose further that ζ_k denotes the cumulative average years of schooling of each educational level k, where $k \in \{0, 1, 2, 3\}$ represents (0) no schooling, (1) primary, (2) secondary, and (3) tertiary education. Finally, n_k denotes the share of the population that has achieved the respective educational level. The Gini coefficient of educational inequality ($Gini^h$) can be computed via (see Castelló-Climent, 2010)

$$Gini^h = \frac{1}{2\bar{H}} \sum_{i=0}^{3} \sum_{j=0}^{3} |\zeta_i - \zeta_j| n_i n_j.$$

The detrimental effect of schooling inequality is illustrated in Figure (4.11). The figure shows the relationship between educational inequality and the average years of schooling in the sample of high-income countries based on 5-year averages in the period between 1960 and 2013. This illustration underscores a significantly negative correlation between schooling inequality and the average number of years the individuals devote to education. Countries that are shaped by a high degree of

63 Note, however, that this situation may change in light of the high rate of net migration observable since 2015.

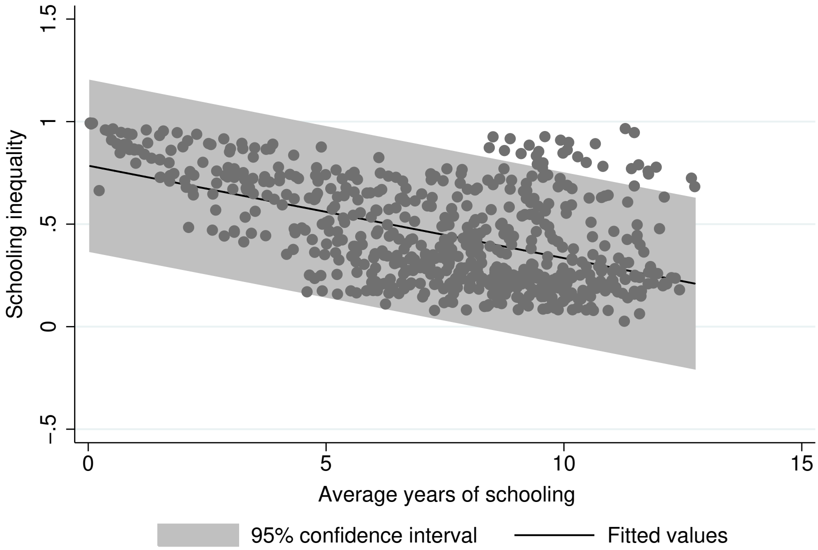


Figure 4.11 The relationship between educational inequality and average years of schooling. Sample includes high-income countries according to the classification of the World Bank (income higher than 12,736 USD).

schooling inequality exhibit a lower average education level, which is plausible as most individuals quit the education sector at the very latest once they have finished tertiary education. Due to diminishing marginal returns to education, a high level of schooling inequality leads to a waste of growth potentials via two channels. First, schooling inequality directly reduces the level of human capital that is employed for production of the output. Second, a low number of tertiary educated individuals hampers the implementation of technological advancements.

Figure (4.12) depicts the development of the Gini index of schooling inequality for Germany, Japan, Sweden, and the United States in 5-year intervals covering the period between 1960 and 2013. The figure reflects the transnational trend of declining schooling inequality that is observable in the overwhelming majority of nations on the globe. However, the cross-country differences in schooling inequality are as distinct in the post-2010 period as they have been in the early 1960s. While the mean value of the educational Gini index was .58 in the early 1960s, it declined to .40 in the period between 2010 and 2013. During the same time, the standard deviation remained at a constant level of 24 Gini points. When referring to the advanced economies, the United States are among the countries with the highest level of schooling inequality, only surpassed by the United Kingdom (.79 Gini points) and Switzerland (.72). At the other end of the spectrum, Japan (.08) has the most equalizing schooling system in the world. Similarly low levels of educational inequality can be found in Hong Kong

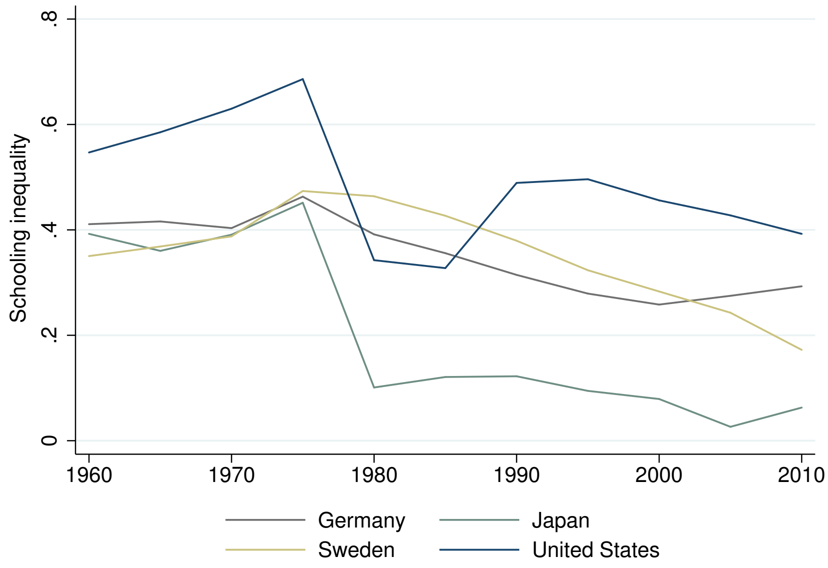


Figure 4.12 The development of the Gini index of schooling inequality in Germany, Japan, Sweden, and the United States.

(.17) and Canada (.19). The development observed in Sweden is typical for a number of Scandinavian countries: while schooling disparities were strongly pronounced in the early 1980s, a number of attempts to modernize the education sector yielded a substantial decline in educational inequality. Similar tendencies have taken place in Finland, Denmark, and Norway.

Germany is close to the average of the advanced economies in terms of schooling inequality. Resembling the development of most Scandinavian economies, the country benefited from a decline in educational disparities between 1960 and 2013. However, Germany is also among the few countries with an increase in schooling inequality at the current edge of the sample.

4.3.7 Innovation activity in the advanced economies

The second major driver of long-run development identified in Tables (4.1) and (4.2) is technological progress. The results also suggest that international spillovers matter and that factor productivity and capital goods may be imported from abroad, given that domestic human capital is capable of mastering internationally available knowledge. This suggests that the technological frontier is not country-specific. In fact, if economies and societies are closely linked, $T(t)$ is composed of technological and scientific contributions from a large number of countries. While this potential is theoretically equal for each country with a sufficient degree of openness, the adoption of technology differs between the economies. As outlined in Section (4.2), inventions may occur randomly over time. In times when research does not lead to significant progress, $T(t)$ may develop in small, continuous steps. However, in times when new general purpose technologies arise, the technological frontier may increase by leaps and bounds.

New GPT provide the potential for a number of new products (j^*), as well as their improvements and variations (κ_{j^*}). As the costs of investing in any κ_j are much lower than of investing in j^*, the tendency to adopt entirely new technologies will become lowered after new GPT have been introduced. However, since the κth step is easier to achieve than the $(\kappa + 1)$th step, improving and modifying existing goods will gradually become more difficult. At some point in this process, the prospect of a monopolistic position will encourage entrepreneurs to invest in new, hitherto unexploited inventions (see Chapter (2) for a detailed illustration and Chapter (3) for the role of the financial sector in this process). This consideration has compelled a variety of authors, such as Schumpeter (1911, 1939) and—in more recent times—Jovanovic and Rousseau (2005), Bresnahan and Trajtenberg (1995), and Helpman and Trajtenberg (1994) to argue for the development of innovation activity in waves.

This section examines the historical development of global innovation activity and the current growth potential provided by the technological frontier. The starting point of this analysis is data on patent applications from the United States, Germany, and the world between 1790 and 2011. These series are detrended using a polynomial of the form

$$\log(\rho)_t = \alpha + \sum_{n=1}^{\mathfrak{N}} \beta_n \tau_t^n + \varepsilon_t,\ \mathfrak{N} = 1, ..., 4, \tag{4.6}$$

where τ^n denominates the nth degree trend variable. The selection of n refers to the minimization of the estimated p-values. The residuals of this estimation illustrate the up- and downturns in innovation activity. In order to smooth short-term fluctuations, the analysis uses moving averages of ξth degree. The innovation index can thus be computed via

$$\Lambda(t) \equiv \bar{\varepsilon}(\xi) = \frac{\varepsilon_t + \varepsilon_{t-1} + ... + \varepsilon_{t-(\xi-1)}}{\xi}. \tag{4.7}$$

Section (4.2) suggests that the diffusion of GPT can be modeled using a logistic or Gompertz function and that the development of improvements and variations of that technology follows the first derivative of these functions. Each of these circles may be denoted with κ_r', while the long-run innovation index $\Lambda(t)$ may be expected to consist of multiple concatenations and overlaps of individual κ_r'.

Figure (4.13) illustrates the innovation index $\Lambda(t)$ for the United States with $\xi = 5$ and depicts the time points at which major inventions appeared. The frequency of patent applications in the past was evidently subject to strong fluctuations. These fluctuations do indeed have strong similarities with the diffusion cycles κ_r'. Comparing the variations with the occurrence of GPT, there is a strong indication that the frequency of patent applications increased at above-average levels whenever

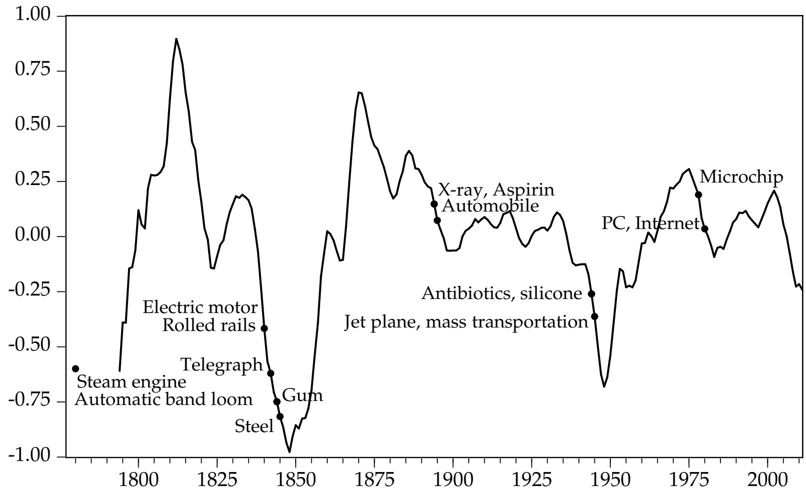


Figure 4.13 The innovation index (Λ_{it}) for the United States, 1790–2011. Dating of the GPT is based on Haustein and Neuwirth (1982), Van Duijn (2013), Silverberg and Verspagen (2003), and Gordon (2012). The timing of the emergence of the railroad refers to the invention of the rolled rails, whereas the steam locomotive was developed in 1824.

radically new inventions appeared. Yet, with respect to most of the GPT featured in Figure (4.13), the increase in patent applications occurs only after a time delay of approximately 5 to 10 years. This lag may occur due to the time requirement for the development of new infrastructure and the corresponding skills necessary to master the new technology. The delay argues in favor of a sigmoid shape for the diffusion function of GPT. Some time after the adoption of the new technology, the potential of transfer applications and improvements is exhausted, resulting in a slowdown in $\Lambda(t)$.

$\Lambda(t)$ also allows assessment of the current growth potential provided by technological advancements. Starting in the early 2000s, the index implies a sharp downturn in worldwide innovation activity. In fact, Figure (4.13) illustrates that the innovation index has not fallen to a comparably low level since the late 1940s. However, the figure clearly brings to light the fact that there have been numerous fluctuations in the evolution of $\Lambda(t)$ during the past 200 years. Historically, each downward movement could be offset after some period of time.

Since different methods of detrending can easily yield divergent results, Figure (B4-4) in the appendix compares the innovation index $\Lambda(t)$ with a comparable index using the HP filter. The shape of the curve is in fact very similar, although some fluctuations tend to be less pronounced. Figure (B4-4) furthermore compares $\Lambda(t)$ with the German innovation index, acquired by applying the same method described above using patent data from Germany. The German index follows $\Lambda(t)$ with some delay, which reflects the effect of technological diffusion. As many inventions depicted in Figure (4.13) have their origin in the United States, international spillovers cause the innovation index in Germany to rise after a certain amount of time. The reduction in the delay between the innovation cycles of Germany and the United States that occurred during the last two decades illustrates the effects of globalization, which triggered a substantial increase in the diffusion rate.

Overall, technological progress currently seems to provide poor potential for long-run growth in the advanced economies. Figure (4.13) demonstrates that innovation activity has declined since the early 2000s. The HP filtered innovation cycle and the German innovation index (both illustrated in Figure (B4-4) in the appendix of this chapter) support this assessment. In addition, Figure (4.14) shows the innovation index of the world, derived by the method described in Equations (4.6) and (4.7) using aggregated worldwide patent data. This index reveals a similar turning point in patent application growth around the year 2000. Although one might argue that patent data must be considered biased and inappropriate for gauging "real" innovation activity, it is still the only variable available for a large sample of countries and a sufficiently long time span. Nevertheless, it is useful to compare the results to other innovation indicators to prevent the drawing of unfounded conclusions. Figure (4.14) (second part) shows the HP filtered trends of multifactor productivity growth rates in the United States and the advanced economies. The limited time span is the result of limited availability of data. Both indicators illustrate a negative trend in the long-run, which is particularly pronounced after the year 2000. The short recovery period around 1980 resembles the development of $\Lambda(t)$.

The figures presented in this section support the supply-side explanation of secular stagnation and yield implications similar to those of the work of Gordon (2012, 2015), who argues that faltering innovations are currently leading to a slowdown of the US growth rate. As Gordon (2015) points out, the main impact of the computer and internet (ICT) revolution has reached its climax in the dot-com era, with the initially strong influence on labor productivity withering away since the early 2000s.

Most firms in the advanced economies have already benefited from the internet and web revolution, and the methods of production have been subject to little change during the last decade. Having utterly changed the way offices function, the digital revolution reached its crest in the early 2000s. By that time, most places of work were equipped with personal computers allowing for utilization of word-processing software without repetitive retyping, as well as instantaneous transmission and acquisition of information as a result of linkage to the internet (Gordon, 2015). Quite similarly, Jovanovic and Rousseau (2005) show that electrification spread much faster than ICT technologies and did so more broadly over different sectors, which is why its impact on growth was considerably more important. In addition, electrified machinery comprised a larger fraction of physical capital when electrification was at a stage of development similar to that of ICT today.

In fact, when looking at the data, ICT already seems to be on the wane. Figure (4.15) shows investment in ICT in Australia, France, Germany, Japan, the United States, and the United Kingdom. In the early 1990s, ICT goods and services were of minor importance in each of the depicted countries. With the increasing ubiquity of

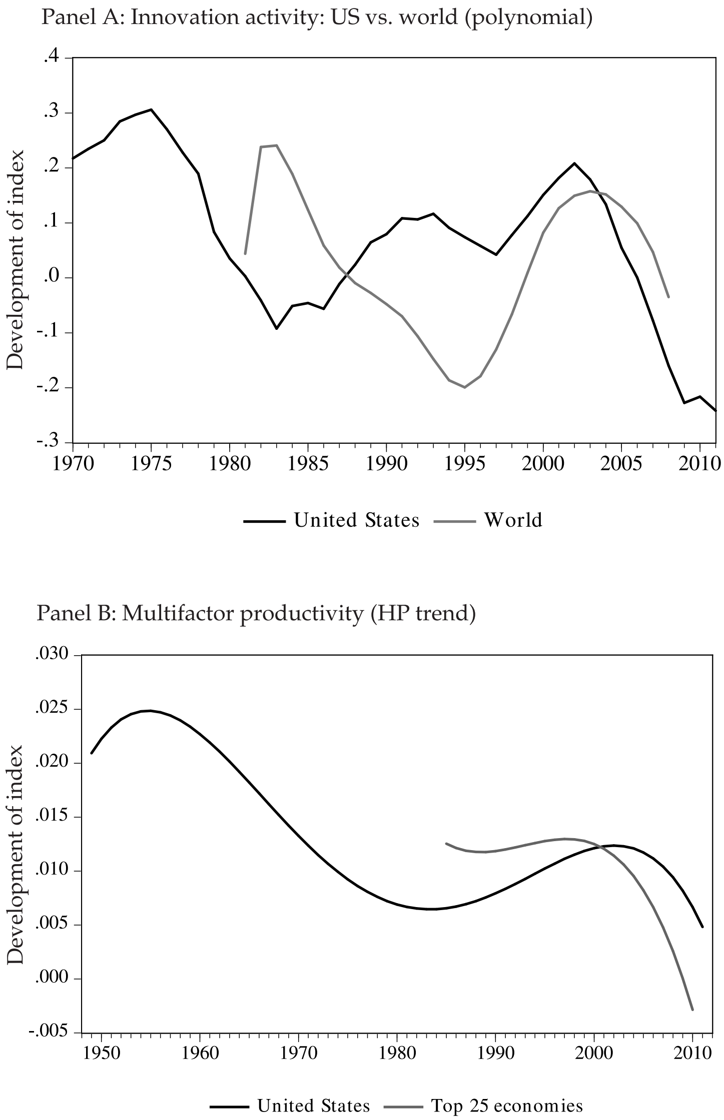


Figure 4.14 Innovation activity in the United States, the developed economies (top 25), and the world. Data source is OECD (2013a,b), Bureau of Labor Statistics (2013), World Intellectual Property Organization (2013), and own calculations.

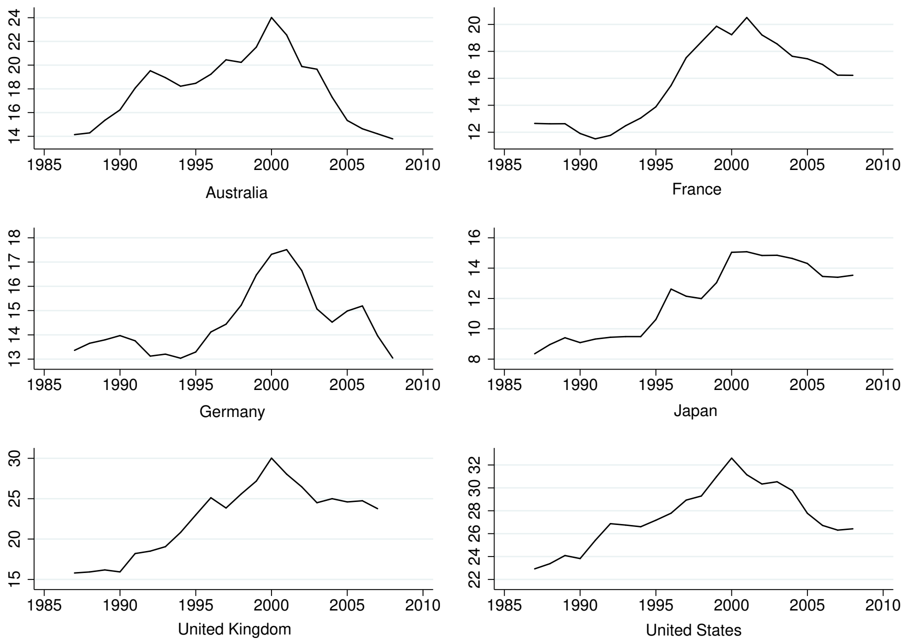


Figure 4.15 Investment in ICT in Australia, France, Germany, Japan, the United Kingdom, and the United States. Data source is OECD (2016).

the internet in households and firms, ICT investment rose during the 1990s, reaching its crest around the turn of the millennium. Around the time the dot-com bubble reached its climax (which, referring to the development of the NASDAQ, is often dated March, 2000), investment in communication and internet technologies began to decline. This is in line with Gordon (2015), who argues that most of the ICT applications with the potential to fundamentally change working conditions had been broadly implemented in the early 2000s. A similar picture is obtained when looking at the development of the aggregate ICT capital stock, which provides the benefit that the data is available for a longer time span (the data can be accessed via The Conference Board, 2015).

Figure (4.16) illustrates that the ICT capital stock grew at rates of approximately 17 percent per year during the internet boom that preceded the dot-com collapse. In the aftermath of the crisis, growth rates of ICT settled down to values substantially below 10 percent per year. The year 2010 even saw stagnation in the development of ICT capital in the United Kingdom. In the post-2010 period, there has been a slight recovery in the growth rate of ICT capital; however, this increase is still considerably below the historical progress around the early 2000s.

Today, ICT has not only reached a high degree of saturation in the business sector, but is also widespread among private households. Figure (4.17) shows the portion

Figure 4.16 The development of the ICT capital stock in the United Kingdom and the United States, growth rates. Data is from The Conference Board (2015).

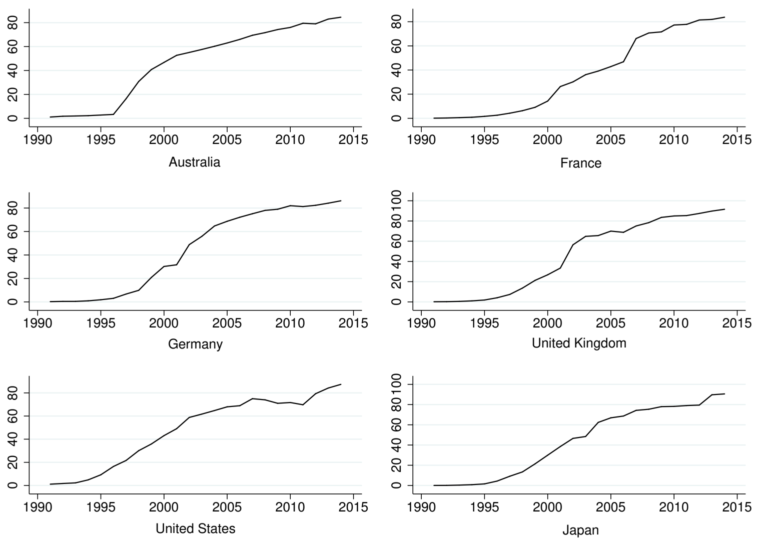


Figure 4.17 Internet users per 100 people in Australia, France, Germany, the United Kingdom, the United States, and Japan. Data is from World Bank (2014b).

of the population that has used the internet during the past 12 months, either via computer or mobile phone. In the mid-1990s, few people had access to the internet, which is reflected in less than 10 internet users per 100 inhabitants in each of the countries illustrated in Figure (4.17). Around the year 2000—the time of the dot-com collapse—internet usage of private households was still relatively low, ranging from under 20 percent in France to approximately 40 percent in Australia and the United States. However, between 2000 and 2010, the internet found its way to private users on a broad scale, resulting in a sharp increase in the fraction of the population that uses applications with a link to the world wide web. Around the year 2015, more than 90 out of every 100 people in Australia, France, Germany, the United Kingdom, the United States, and Japan are internet users.

In addition to the declining potential of ICT, a more general explanation of the growth crisis has been brought forward by Cowen (2011), who postulates the idea of the “Great Stagnation”. The theory emphasizes that during the last 100 years, the advanced economies have realized most of the productivity gains that could be easily attained with minimal effort. Today, most of the factors that drove historical welfare increases are exhausted, leading to falling rates of both innovation activity and economic growth.

While the consequences for long-run growth seem to be alarming given the rate of technological progress implied by the analyses of this chapter, the examination also shows that during the past 200 years, the world has never run out of new ideas. Figure (4.13) highlights that periods with below-average innovation activity have historically always been followed by periods featuring substantial technological advancements. For this reason, the analysis in this chapter implies that the current lack of radically new ideas may only contribute to a *temporary* decline in long-run growth rates.

In fact, new potential GPTs are already on the horizon. Nikulainen and Kulvik (2009) provide some evidence from Finland that nanotechnologies have the potential to be widely applicable and to influence economies similarly to recent GPTs. Youtie et al. (2008a,b) also conclude that nanotechnology may be a breakthrough innovation with long-run growth potential. Figure (4.13), however, illustrates that the effects of such a new technology emerge only after a significant time lag. Even if nanotechnological applications were already marketable—which is debatable—it would be years before the diffusion process exerted positive effects on growth.

4.4 Conclusions: Secular stagnation and growth crises in the advanced economies

This chapter provided evidence for a supply-side explanation of the long-run decline in growth rates observable in the advanced economies, which has recently become known as “secular stagnation”. When examining the empirical growth rate of the advanced economies—i.e. the 25 most affluent countries in terms of real per capita incomes—, a striking fact observable in nearly all of the countries is a substantial slowdown in growth that was initiated in the early 2000s. While this phenomenon has initially been ignored by the economics profession, there is now a rapidly growing literature examining the decline in income increases since the turn of the millennium. In 2014, Summers (2014) revived the term of secular stagnation with a particular emphasis on the demand-side. However, the advanced economies currently face a situation that has stronger similarities to the mirror image of the circumstances in the late 1930s that prompted Hansen (1939) to publish his theory on long-run stagnation. Aggregate supply grew at the fastest rates of any time in the history of the United Stated in the mid-1930s, whereas demand lagged considerably behind this development. In contrast, the current situation is primarily shaped by a significant decline in productivity growth.

In light of the inapplicability of the demand-side view for the current situation, this chapter provides a supply-side explanation for the decline in growth that is based on three elements. First, the empirical analysis of the determinants of economic growth reveals conditional convergence as a clear empirical regularity. Whereas this observation explains the disproportionally high growth rates of many of the advanced economies in the aftermath of World War II, it also provides a primary indication of why the transnational growth crisis may have occurred. During the last decades, many of the emerging markets have been able to approximate the income level of the developed economies. This development brought with it a substantial increase in desired savings, which contributed to an excess of savings over investment. Second, the analysis of technological progress implies that innovation activity has considerably declined since the turn of the millennium. This limits the potential of long-run growth by restricting the creation of both new and improved capital goods, production processes, and marketable products. As technology typically diffuses across countries, this development affects each of the advanced economies similarly. Third, the ability to make use of the technological frontier depends on the national level of human capital, which strongly differs across the developed countries. While the slowdown in technical change explains the slump in growth trends in the affluent economies, the disparity in the human capital level and the distribution of the potential for its future accumulation determine the relative position of the countries in the group of advanced economies.

In accordance with the work of Gordon (2015), the analysis implies that the positive effects of the ICT-revolution have vanished over the past decade. Most of the inventions that have replaced repetitive and tedious clerical work have been adopted, and most of the potential for office work, retailing, and business dynamism is exhausted. More recent ICT innovations are concerned with the exchange of information, but as further increases in the speed of communication are difficult to achieve given the already high level of electronic integration, they no longer exert a fundamental influence on productivity. The declining potential provided by the technological frontier has led to a transnational reduction in growth rates across the developed economies.

What are the implications provided by the analysis in this chapter with regard to future growth rates? While the results suggest that the current long-run growth potential is low, the analysis also shows that over the past 200 years, radically new technologies particularly emerged in periods where the rate of technical progress experienced a decline. Transferring the current situation to the development of growth rates in the next 5-10 years may thus be deceptive. Some applications are already on the verge of becoming new general purpose technologies, among which nanotechnology, biotechnology, and medical technology are the most promising examples. If their positive effects reach the economies on a broad front, growth rates similar to those realized between 1970 and 2000 may easily be possible after some time delay. Yet, the idea of a future phase of prosperity as experienced by Germany and other European countries in the 1950s and 1960s can still be seen as utopian at best.

4.A Appendix of Chapter (4)

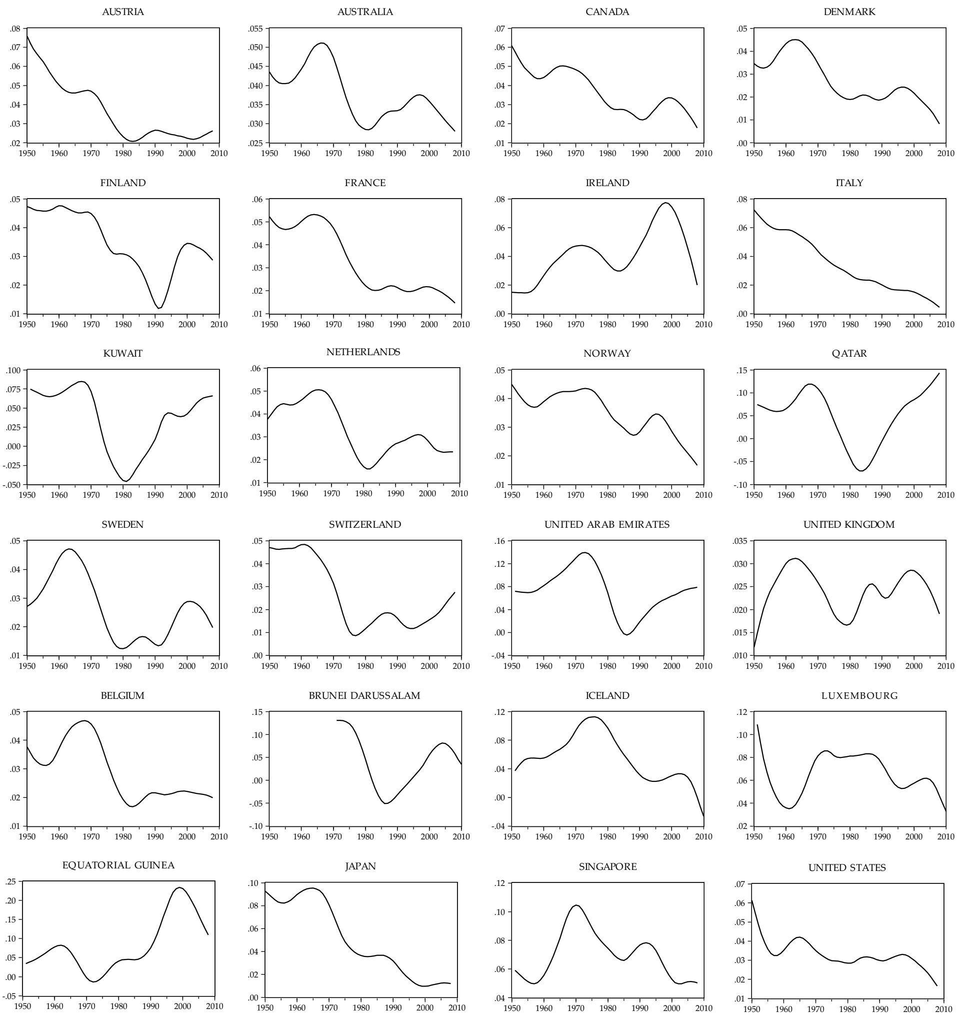


Figure B4-1 Trends of real per capita GDP growth in the top-25 countries, 1950–2010. Data source is Heston et al. (2012), trend is calculated using the Hodrick-Prescott filter.

Table A4-1 Data sources of the variables used in the regression analyses in Tables (4.1) and (4.2) and in the descriptive analysis of this chapter.

Variable	Data Source
$\mathrm{d}y/\mathrm{d}t$	Heston et al. (2012)
$y_{t-\tau}$	Heston et al. (2012)
SCHOOLY	Barro and Lee (2013)
LIFEEX	World Bank (2012c)
FERT	World Bank (2012c)
INVS	World Bank (2012c)
DEM	UTIP (2012), Vanhanen (2012)
OPEN	Heston et al. (2012)
HOF	Freedom House (2014)
GOVC	Heston et al. (2012)
SUBSAH	Own calculations, classification of World Bank (2012c)
LATINAM	Own calculations, classification of World Bank (2012c)
PAT	World Bank (2012c)
CIT	Hall et al. (2001)

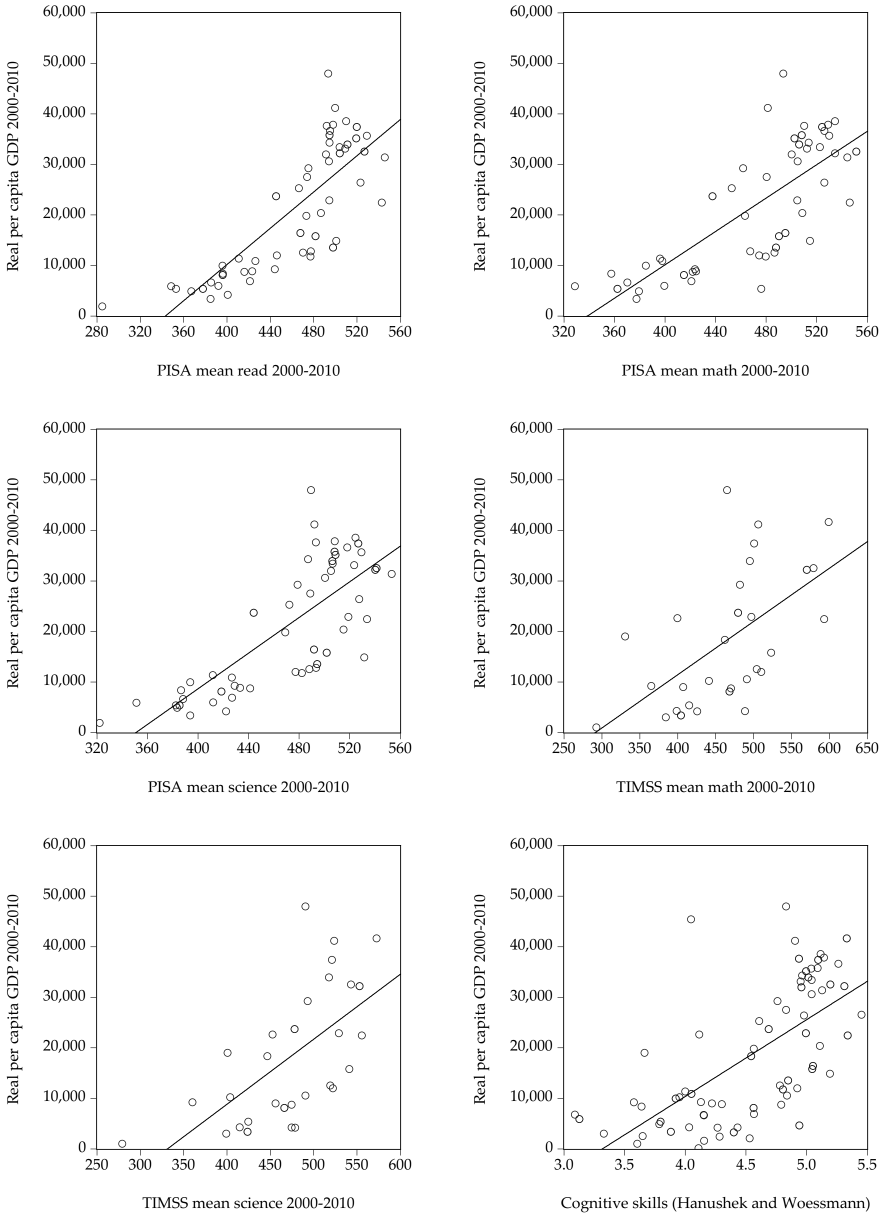


Figure B4-2 Schooling quality and growth. Data source is Heston et al. (2012), World Bank (2012a), Hanushek and Woessmann (2012), and own calculations. The graphs show the averages of the PISA and TIMSS waves conducted in the period 2000–2010 and the average value of real per capita GDP. The number of observations depends on the number of countries participating in the respective program: PISA ($N = 58$), TIMSS ($N = 41$), and the cognitive skills index ($N = 75$).

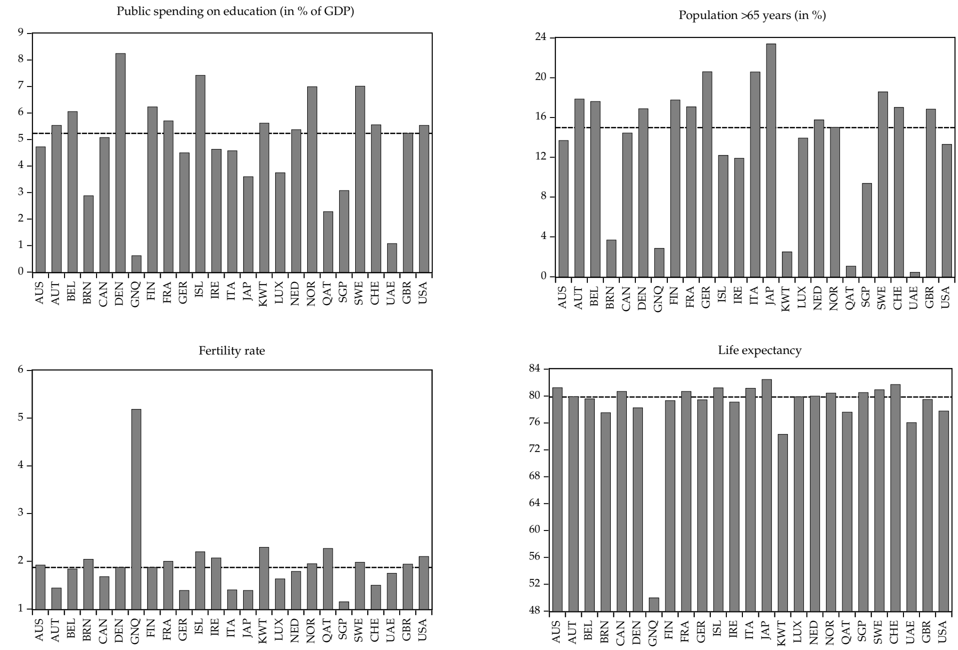


Figure B4-3 Human capital in the advanced economies. Public spending on education, fraction of population aged 65 and above, fertility rates, and life expectancy in the top 25 economies. Data source is World Bank (2012c).

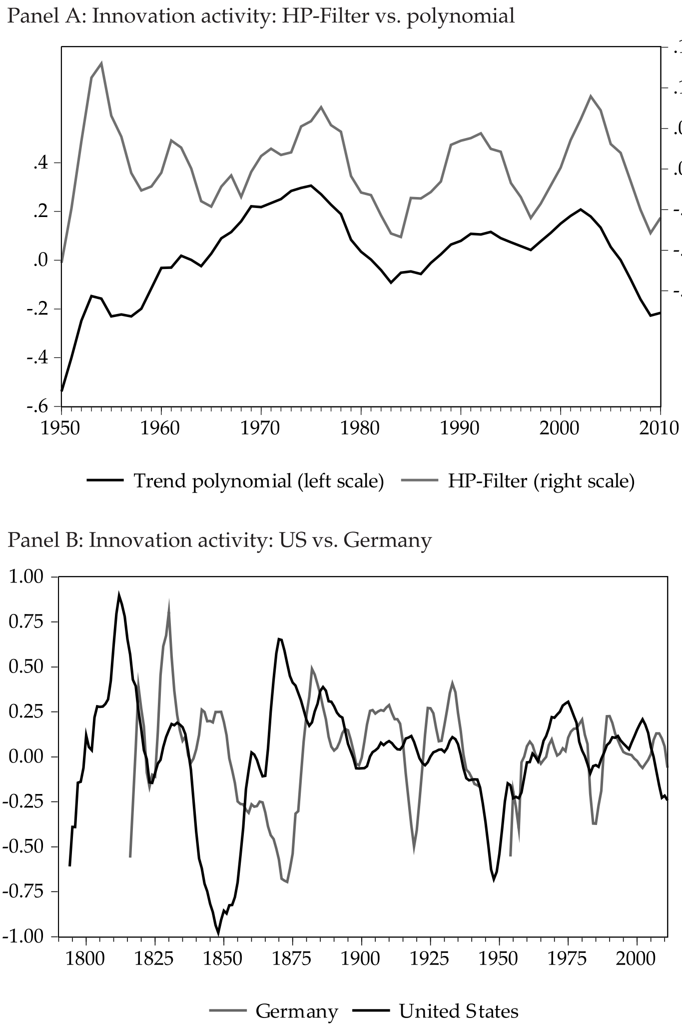


Figure B4-4 Innovation index (Λ_{it}) for the United States and in Germany. Data source is US Patent and Trademark Office (2013), Federico (1964), World Intellectual Property Organization (2013), and OECD (2013a).

Chapter 5

The Effect of Income Inequality on Growth, and the Role of Governmental Redistribution[64]

Background The previous chapters illustrated how decisions of individuals regarding investment in physical capital, innovation activity, and education influence the rate of long-run growth. In Chapter (2), the implicit assumption was that the initial wealth of the individuals is sufficient for them to cover the cost of the investment independently. Chapter (3) enlarged the analysis by illustrating that under some circumstances, provision of funds by the financial sector can help to soften the budget constraint and to facilitate the realization of investments if the wealth of individuals is not sufficient to bear the associate cost themselves.

While the previous chapters made no concrete assumption regarding the distribution of incomes, this chapter introduces the concept of income inequality into the methodological framework. In the context of the growth mechanisms discussed previously, inequality of incomes implies that some individuals have the ability to carry out their investment projects, while insufficient wealth prevents others from exploiting their full intellectual and entrepreneurial potential. The result is a decline in education and innovation activity, which translates to a lower rate of economic growth. In general, functioning capital markets can help to mitigate this problem. However, particularly in the case of developing economies, capital markets are often imperfect (see Galor and Zeira, 1993 and Galor and Moav, 2004). To prevent economic growth from being negatively affected from such situations, public redistribution may seem to be an adequate policy measure to equalize the investment opportunities among individuals. This, however, only holds if redistribution itself is neutral in the growth process. The present chapter is concerned with an empirical examination of the effect of inequality and its remedy in the form of public redistribution.

[64] This chapter is based on joint work with Philipp Scheuermeyer that appeared as Gründler and Scheuermeyer (2018).

5.1 Introduction

In his famous book *"Equity and Efficiency: The Big Tradeoff*", Okun (1975) points out that the trade-off between social justice and economic efficiency "plagues us in dozens of dimensions of social policy". Okun's notion led to the widespread belief that public redistribution via taxes and transfers creates disincentives and inefficiencies that Okun compares to a "leaky bucket", with money lost whenever transfers are made from the rich to the poor. However, empirical evidence for the existence of such a trade-off is rather ambiguous.

The literature at hand can be divided into two distinct groups. One branch examines the link between inequality and growth, while the other studies the growth effects of redistributive taxes and social transfers. This chapter follows a novel approach by simultaneously exploring the growth effects of both income inequality and effective public redistribution, with the latter computed as the difference between market and net income Gini coefficients. We find that a high level of inequality reduces GDP growth, but its remedy—redistribution via taxes and transfers—is detrimental to growth as well. Thus, the direct negative effect of redistribution offsets its indirect positive growth effect from reduced net inequality. Taken together, this means that at a given level of market inequality, the impact of redistribution on economic growth is insignificant. However, the growth effects of both inequality and redistribution depend on the development level of the economy. Whereas redistribution—on aggregate—fosters growth in developing countries, it seems to have a rather impedimental effect in advanced economies. To study these effects in greater detail, we explore the transmission channels through which inequality and redistribution affect economic development. In fact, recent studies on the inequality-growth nexus mainly focused on reduced form evidence, neglecting the mechanisms behind the identified effects. Our results suggest that higher inequality is negatively related to education and yields an increase in the fertility rate. Both effects are particularly prevalent in the presence of limited access to capital and can be mitigated by public education spending. Meanwhile, the direct negative effects of redistribution are mainly due to a decrease in investment and an increase in fertility.

How do these findings relate to earlier studies on the topic? Whereas cross-country analyses tend to find a negative relationship between income inequality and economic growth, the results have become ambiguous since the advent of panel data methods.[65] Particularly, Li and Zou (1998) and Forbes (2000) contradict previous findings by detecting a positive impact of inequality on economic growth. In contrast, Barro (2000) yields little indication of a uniform relationship between inequality and growth, as he finds a negative effect of inequality in developing countries and a positive effect in richer economies. Castelló-Climent (2010) confirms this interaction with the development level, but finds an overall negative growth effect of income and human capital inequality. Focusing on the use of consistently measured inequality data, Knowles (2005) finds a negative effect of Ginis from household expenditures, but not of Ginis from gross incomes. Voitchovsky (2005) enriches the debate by

[65] The empirical growth literature of the 1990s is comprehensively reviewed in Aghion et al. (1999).

looking at the shape of the income distribution. The study concludes that growth is promoted by inequality at the top end of the income distribution, but weakened by inequality at the bottom end. Finally, Halter et al. (2014) emphasize the time dimension of the inequality-growth relationship by showing that higher inequality fosters growth in the short term, but hampers growth in the medium to long run. Hence, one explanation for the inconclusiveness of the literature is that estimates based on time-series variations pick up positive short-run effects of inequality, whereas methods which also exploit cross-country variations capture its negative impact in the medium to long run.

The empirical evidence for the growth effects of redistributive fiscal policy is also divided. Using specific fiscal policy instruments to proxy the extent of redistribution—such as marginal tax rates or the amount of social spending—, earlier studies tend to find a negligible or slightly positive impact on growth (see, e.g., Perotti, 1996). In light of these findings, Lindert (2004) suggests that large welfare states have come up with methods to minimize the negative incentive effects and deadweight losses from taxes and social spending. In contrast, a study by Muinelo-Gallo and Roca-Sagalés (2013), which uses panel data from 21 high-income OECD countries, shows that distributive expenditures and direct taxes produce significant reductions in inequality, but also in GDP growth.

So far, a lack of meaningful and comparable data limits the exploration of the growth effects of inequality and redistribution. First, with regard to the relationship between inequality and growth, several studies (Knowles, 2005, Atkinson and Brandolini, 2009) highlight that mixing Ginis from different income definitions or applying simple transformations to make them more comparable is inappropriate but nevertheless a common approach in the literature. Meanwhile, attempts to work with consistently measured inequality data have so far been restricted to a very narrow selection of countries and years (Knowles, 2005, Voitchovsky, 2005), imposing the risk that findings are due to sample selection rather than different income definitions.[66] Second, regarding the effect of redistribution on growth, most studies use fiscal policy variables to measure the extent of public redistribution. Yet the size of taxes and transfers tells little about their progressivity, meaning that the redistributive impact of specific fiscal policy measures is unclear and not comparable across countries.

Recent advances in data availability allow us to address these issues by employing a set of inequality data that maximizes comparability for the broadest possible sample of countries and years (Solt, 2016). Applying a flexible missing data algorithm, the Standardized World Income Inequality Database (SWIID 6.1) provides consistent Ginis of net and market incomes for roughly 5,100 country-years. Covering data from 192 countries between the early 1960s and 2014, our regression sample thus enables investigation of the global relationship between inequality and growth, as well as of the effects at different development levels.

By replacing the ad-hoc fixed adjustments that have long been necessary to generate a large dataset for cross-country research, the SWIID alleviates a general

[66] This problem was already noted by Knowles (2005) and highlighted in a literature survey by Neves and Silva (2014).

trade-off between data comparability and coverage. Meanwhile, we also scrutinize our results based on a sub-sample of the most reliable observations. In addition, we are among the first to exploit the full potential of the SWIID by directly incorporating data uncertainty into our regression results via multiple estimation tools.

Above all, a clear distinction between inequality before and after taxes and transfers in the SWIID enables measurement of redistribution via calculation of the difference between market-income and net-income Gini coefficients. Thus, we regress growth on effective redistribution rather than relying on rough proxies of redistributive fiscal policies. Although it is commonly applied in sociology and public policy (see, e.g., Lupu and Pontusson, 2011; Van den Bosch and Cantillon, 2008), use of the "pre-post" approach for measuring redistribution via the difference between market and net inequality is quite novel in the empirical growth literature. Berg et al. (2014) utilize an early version of the SWIID to acquire data on effective redistribution. While the study finds little evidence for a significant growth effect of redistribution, it suggests that inequality is an impediment to economic growth. Thewissen (2014) calculates a measure of pre-post redistribution using data from the LIS and the OECD. Based on a panel of high-income countries, the study finds no robust influence of inequality and redistribution on economic performance, but indicates a positive relationship between top income shares and growth.

While subject to some studies based on cross-country data (e.g. Perotti, 1996; Deininger and Squire, 1998; Easterly, 2007; Castells-Quintana and Royuela, 2017), the transmission channels of inequality have been rather neglected in *panel data* studies, a point which is criticized by Galor (2009).[67] Meanwhile, the transmission mechanisms of redistribution are largely unexplored empirically. Hence, we are the first to simultaneously study the transmission channels of both inequality and redistribution via panel data econometrics, thereby accounting for unobserved heterogeneity in both the medium- and the long-run. Our results reveal that income inequality acts mainly via human capital accumulation, investment, and the fertility rate. Public redistribution, however, seems to deter investment and to boost the fertility rate. Holding these transmission variables constant, the negative effects of inequality and redistribution on growth vanish. Moreover, the negative impact of inequality on growth is reinforced by credit market imperfections, but attenuated by generous public spending on education. Finally, we provide evidence for the endogenous fiscal policy channel: An increase in market inequality enhances public redistribution, which is why a low level of market inequality is conducive to economic growth.

This chapter is organized as follows: Section 5.2 reviews the main theories on inequality, redistribution, and growth, laying the groundwork for the empirical investigations. Section 5.3 details our empirical specification. Section 2.5 describes

[67] A recent literature survey (Neves and Silva, 2014) identifies only three panel-data studies that examine the transmission channels of inequality on growth. All of these studies focus on single transmission channels. Drawing on *cross-sectional* data, Castells-Quintana and Royuela (2017) study multiple transmission mechanisms, finding that inequality may trigger both positive and negative effects. As the empirical strategy is based on a control function approach that focuses on the between-country variation, our results are not directly comparable. Yet with regard to the identified negative effect of inequality that accounts for 80% of the total effect estimated by Castells-Quintana and Royuela (2017), our results are complementary.

the data, focusing on our measures of inequality and redistribution. The chapter provides an overview of the extent of redistribution across countries and highlights the empirical relationship between inequality and redistribution. We report the baseline results in Section 2.6, followed by an extensive sensitivity analysis. Subsequently, Section (5.6) examines the aggregate effect of public redistribution and investigate its transmission channels. The empirical section closes with an examination of the effects of inequality and redistribution at different levels of development. Section 5.7 concludes.

5.2 The link between inequality, redistribution, and economic growth

Numerous explanations exist for the link between inequality and economic growth.[68] This section consolidates the theoretical approaches into five categories: differential saving rates, credit market imperfections, endogenous fertility, sociopolitical unrest, and the endogenous fiscal policy approach.

5.2.1 Differential saving rates versus credit market imperfections

The classical approach postulates that inequality stimulates growth: Assuming that the marginal propensity to save rises with the income level of individual households (see, e.g., Kaldor, 1955), a concentration of income at richer households increases aggregate saving, which is channeled into investments and thus conducive to growth (Bourguignon, 1981).

However, in the presence of credit constraints and investment indivisibilities an unequal distribution of wealth or income may just as well be detrimental to growth. The credit market imperfections approach, pioneered by Galor and Zeira (1993), suggests that inequality restrains some individuals from exploiting their intellectual potential when credit is not available to cover the direct or opportunity costs of schooling. As the maximum amount of human capital accumulation per person is limited and the returns to human capital are diminishing, an increase in inequality thus reduces both the average quantity and productivity of human capital. Naturally, a public education system that provides free and high quality schooling can mitigate these negative effects.

A similar argument applies to physical capital investment. Viewing people as potential entrepreneurs who face individual investment opportunities that are bound by decreasing marginal returns and credit market imperfections, the poor may not be able to realize their investment projects while the wealthy overinvest. A rise in inequality would thus reduce the average productivity of physical capital, whereas its quantity may be less affected.

[68] A review of the perspective of the new growth theories can be found in Aghion et al. (1999). Voitchovsky (2009) and Neves and Silva (2014) provide surveys of the more recent theoretical and empirical literature on inequality and growth.

Galor and Moav (2004) provide an intertemporal reconciliation between the *differential saving rates* and the *capital market imperfection* approaches in a unified growth theory. Whereas inequality supports growth by increasing aggregate saving and physical capital investment in early stages of development, inequality is detrimental to growth after human capital accumulation becomes the dominant driver of growth in more developed economies. In advanced economies, however, the effect of inequality eventually diminishes, as credit constraints become less binding.

5.2.2 Endogenous fertility

Initial inequality can be detrimental to growth due to a positive link between inequality and the fertility rate. This transmission channel is closely related to the human capital argument as decisions concerning human capital investment and family size are interrelated (Becker and Barro, 1988). Poor parents may lack the resources to invest in their children's education, particularly if they are excluded from capital markets. Thus their only chance to increase family income (or their old-age support) is to increase household size. In contrast, richer families may face relatively high opportunity costs of raising children. As a result it may be optimal for richer parents to have fewer children and to invest more in human capital, providing their offspring with the prospect of higher lifetime incomes.

Firstly, from this it follows that poor societies tend to have high fertility rates and low levels of education. Secondly, empirical evidence underlines that more inequality is associated with larger fertility differentials between educated and uneducated women (Kremer and Chen, 2002). Building upon this finding, De la Croix and Doepke (2003) emphasize the growth effects of fertility differentials. A mean-preserving spread in income distribution increases the number of poorly educated children from disadvantaged families relative to highly educated children from richer families. As the relative weight of the less educated increases, average human capital is diluted. Moreover, an increase in inequality also raises the total fertility rate, which imposes another negative effect on per capita income growth.[69]

5.2.3 Sociopolitical unrest

Inequality may also deter growth by causing an increase in political and social instability (e.g., Alesina and Perotti, 1996 and Alesina et al., 1996). By increasing risk, political instability exerts a negative effect on investment. Moreover, particularly if inequality is accompanied by low rates of social mobility, individuals may engage in criminal activities instead of work or education. By violating property rights, high crime rates may constitute an impediment to physical investment.

A related argument deals with crony capitalism and nepotism. In highly unequal societies a wealthy upper class may enjoy disproportionate political power. As a consequence, the rich may subvert political or legal institutions, engage in rent-seeking activities, and thus hinder GDP growth (e.g., Glaeser et al., 2003).

[69] See Galor and Zang (1997), Morand (1999), and Kremer and Chen (2002) for models of endogenous fertility arguing along similar lines of reasoning.

5.2.4 Endogenous fiscal policy: market inequality and redistribution

The previously described models are all related to the distribution of disposable income. However, another line of the literature focuses on the growth effects of market inequality and public redistribution. Perotti (1996) named the theory put forward by Bertola (1993), Alesina and Rodrik (1994), and Persson and Tabellini (1994) the *endogenous fiscal policy approach*, and divided it into two successive arguments: The first—called the *political mechanism*—states that an unequal distribution of market incomes creates a high demand for redistributive taxes and transfers via the political voting process (Meltzer and Richard, 1981). The second—the *economic mechanism*—stresses the negative incentive effects of redistribution for physical or human capital accumulation and labor effort.

Two limitations apply to this line of reasoning: First, by stimulating risk taking, entrepreneurship, and innovation a positive insurance effect of public redistribution might offset its negative incentive effect. Second, governments also engage in *indirect redistribution* by providing public goods. This may lead to an increase in social mobility and to an equalization of market incomes, which is not captured in standard measures of redistribution such as taxes and transfers.

5.2.5 Overview

In sum, the testable implications that we draw from theory are that the growth effect of inequality depends on (i) the degree of credit market imperfection, (ii) the public provision of education, and (iii) the development level. Inequality should exert a negative influence on growth via a decrease in human capital and an increase in the fertility rate, while its effect on physical capital may be ambiguous. Finally, a high level of market inequality should be related to a high level of public redistribution, which is most likely detrimental for growth. As many of the proposed transmission channels are offsetting, the net effect of inequality and redistribution remains an empirical question, which we examine in the following sections.

5.3 Empirical model and estimation technique

Our estimation strategy uses 5-year averages of all variables, addressing the long-term perspective of growth theory, the need to smooth short-term fluctuations, and the occurrence of gaps in the data. Employing the model structure developed in a number of recent empirical growth studies (Bond et al., 2001, Voitchovsky, 2005, Halter et al., 2014), the 5-year growth rate evolves as

$$\dot{y}_{it} = y_{it} - y_{it-1} = (\theta - 1)y_{it-1} + \lambda h_{it} + \gamma \Psi_{it} + \delta R_{it} + \beta \mathbf{X}_{it} + \eta_i + \xi_t + v_{it}, \quad (5.1)$$

where y_{it} is GDP per capita of country i ($i = 1, \ldots, N$) at 5-year period t ($t = 1, \ldots, T$), h_t denotes human capital endowment per person, and $\mathbf{X}_t$ comprises an array of control variables. In addition, η_i denotes country-specific effects, ξ_t is a time effect of period t, and $v_{it} \equiv u_{it} - \xi_t - \eta_i$ is the error term of the estimation. The marginal effects of our variables of interest—inequality Ψ_{it} and redistribution R_{it}—are captured by the coefficients γ and δ.

As redistribution and inequality depend on the political and institutional environment of the countries, the disregard of growth-promoting covariates in Equation (5.1) could lead to inconsistency in the estimated coefficients. For this reason, we employ a standard system specification which has proven to explain empirical growth patterns quite accurately in a number of earlier studies (see Barro, 2000, 2003, 2013a). However, many of the standard control variables in growth regressions also reflect the transmission channels of inequality and redistribution that we have summarized in Section 5.2. Therefore, a fully specified growth model only identifies the growth effect of inequality and redistribution *beyond* its effect via the standard transmission channels (Galor, 2009). To estimate the full growth effect of inequality and redistribution, we compare the results from the comprehensive growth model with reduced specifications that omit the transmission variables.

Both the differential savings approach and the capital market imperfection theory emphasize that much of the effect of inequality is channeled to growth via education and investment. For this reason, our specification includes average years of schooling (SCHOOLING) and the investment share (INVS). As an additional proxy for human capital, we account for the health level of the population via inclusion of the logarithm of life expectancy at birth, denoted with log(LIFEEX). To measure the effect of the political stability mechanism, we incorporate an index of rule of law and democracy (POLRIGHT) and the inflation rate (INFL) as a proxy for economic uncertainty. The endogenous fertility channel enters into the system via the logarithm of the fertility rate, denoted with log(FERT). Finally, the specification accounts for government consumption (GOVC) and openness (OPEN). The first is assumed to decrease the steady state level of output due to distortions caused in the private sector, while the latter may simultaneously boost growth and inequality due to technological spillovers and increased competition. Data for the control variables are from commonly used sources in growth regressions and are described in online appendix OA-1.

Controlling for the variables discussed above, we examine whether inequality Ψ_t and the amount of redistribution R_t affect the growth rate. Both variables are strongly interwoven: By simply including redistribution in the model, the estimated parameter captures both the effect from a lower level of inequality (which we expect to be positive) *and* the incentive effects from the redistributional measures employed to achieve the reduction in inequality (which we expect to be negative). The simultaneous inclusion of both variables enables us to isolate these contradicting effects.

Ideally, we would like to expunge the endogenous components from the data using an instrumentation strategy based on strong and valid external instruments in an IV or 2SLS setting. In this case, the empirical strategy would be able to identify causal relationships. However, the empirical literature thus far has not proposed any time-varying external instrument that is i) available for a large number of country-years and ii) fulfills the exclusion restriction.[70] In the absence of a valid external instrument, our analysis employs lagged regressors as internal instruments.

From the rich palette of dynamic panel estimators that exploit internal instruments, our analysis employs the system GMM estimator to empirically estimate Equation (5.1), which is described in detail in online appendix OA-2. System GMM has been shown to have better finite sample properties than other commonly used dynamic panel techniques, such as the difference GMM estimator (Arellano and Bond, 1991) or the Anderson and Hsiao (1982) estimator (see Blundell et al., 2000). In addition, this strategy helps to circumvent a dynamic panel bias, which occurs when including a lagged dependent variable in time-demeaning approaches such as within-group estimates (see Nickell, 1981). However, the validity of system GMM relies on some crucial assumptions, particularly the Arellano and Bover (1995) conditions. To detect possible violations of these assumptions, we conduct Hansen's J and Difference-in-Hansen tests along with each regression.[71] Moreover, to respond to the growing concern of weak instrumentation in empirical growth studies, we follow Bazzi and Clemens (2013) and Kraay (2015) by performing a number of weak instrument diagnostics to ensure consistency and to guarantee that the estimated parameters are unbiased.

The system GMM estimator uses lagged variables as internal instruments for endogenous regressors. To avoid arbitrary assumptions on exogeneity, we treat all independent variables as endogenous, as is common in the literature. This, however, potentially leads to a large number of instruments. Yet, a large set of instruments possibly overfits the instrumented variables, which may fail to expunge the endogenous components and may bias parameter estimates towards those from noninstrumenting estimators. To tackle this problem, we use a collapsed instrument matrix. Intuitively, the collapsed matrix contains one instrument for each lag distance and instrumenting variable. In addition, we follow Roodman (2009b) in combining this method with a reduced lag structure.[72]

70 In fact, Bazzi and Clemens (2013) stress that violations of the exclusion restriction in numerous growth regressions focusing on external instruments raise doubts about the obtained results. Among the few attempts to instrument inequality with external instruments, the most promising study is Easterly (2007), which proposes instrumenting inequality with agricultural endowment. However, this instrumentation strategy is not applicable when using panel data that contains both affluent and developing countries over a long time-span.

71 A more detailed description of the estimator in the context of the empirical application can be found in Bond et al. (2001) and Roodman (2009b).

72 When presenting our empirical results, we compare the growth effects of redistribution and inequality in reduced models with those in more comprehensive specifications. This comparison is essential for the analysis of the transmission channels but evokes the problem of a large difference in the number of regressors across models. A common setting of lag structures for each model would thus either result in invalidity of the over-identifying restrictions in the reduced models or proliferation of instruments in the comprehensive models (reflected in either very low or very large p-values of Hansen's J-test). One way to cope with this problem would be to change the lag structure across models, which would, however, impede comparison across the specifications. For this reason, we use second and third lags of the variables in our reduced setting. When the model is enlarged by the control variables, we collapse the instrument matrix based on all available lags but maintain the lag structure with respect to the variables included in the reduced model.

In principle, our specification can be estimated using one-step or two-step GMM. Whereas one-step GMM estimators use weight matrices independent of estimated parameters, the two-step variant weights the moment conditions by a consistent estimate of their covariance matrix. The two-step procedure is asymptotically more efficient (Bond et al., 2001), but the computed standard errors may be downward biased in small samples. We therefore rely on the Windmeijer (2005) finite sample corrected estimate of the variance, which yields a more accurate inference.

5.4 Data description and the link between inequality and redistribution

5.4.1 Data on inequality and computation of redistribution measures

Our main variables of interest are inequality (Ψ) and redistribution (R). To measure inequality, we use the Gini coefficient, which gauges personal income inequality be-tween households within a given country. In principle, the Gini can be calculated using market incomes ("market Gini") or disposable incomes ("net Gini"). Differences in these variables are the result of taxes and transfers. For this reason, our redistribution measure REDIST is calculated as

$$\text{REDIST}_{it} = \text{GINI(M)}_{it} - \text{GINI(N)}_{it}, \tag{5.2}$$

where GINI(M) is market inequality, and GINI(N) denotes inequality of disposable incomes. This measure is often referred to as the "pre-post-approach" in the sociological and public policy literature.

When working with cross-national income inequality data, researchers are confronted with a trade-off between the comparability and the coverage of observations (Solt, 2016, 2015). The Luxemburg Income Study (LIS) constitutes the gold standard of cross-nationally comparable inequality data, but the calculation of inequality measures using a uniform set of assumptions and definitions strongly restricts data availability to only 232 observations from 41 countries. The limited scope of countries and years included in the LIS impedes the application of system GMM and does not allow for the investigation of the effect of redistribution based on a large panel of countries. The incorporation of a larger number of observations, however, typically comes at the cost of sacrificing the benefits of comparability and harmonization. Atkinson and Brandolini (2001, 2009) review the pitfalls inherent in the use of secondary datasets and conclude that simple adjustments are not sufficient to generate comparable inequality measures that rest on common income definitions and reference units.

To ease this problem, the Standardized World Income Inequality Database (SWIID) compiled by Solt (2009, 2016) offers model-based multiple imputation estimates of the missing country-years in the LIS series. Next to the benchmark data from the LIS, the SWIID employs source data from a large number of cross-national inequality databases, national statistical offices, and scholarly articles, thereby making use of a maximum of possible information. Hence, the coverage of country-years for which harmonized data is available for both net *and* gross inequality far exceeds those of alternative cross-national inequality datasets. Currently, the SWIID covers 192 countries from 1960 to 2016 with estimates of net and market income inequality for 5,119 country-years. By calculating 5-year averages, we obtain a total of 1,052 country-years, yielding a regression sample of up to 969 observations. As we intend to investigate the effect of inequality and redistribution across different development levels, the large data coverage of the SWIID—particularly with regard to the scope of countries included—is decisive for the purpose of this chapter.

Despite of its large coverage, the SWIID maximizes comparability by closely following the advise from Atkinson and Brandolini (2001, 2009). Essentially, it does so by adopting a flexible missing data algorithm, which produces consistent measures of gross and net inequality, based on information from closely related observations.[73] Thus the SWIID replaces the more or less ad-hoc global fixed adjustments that have long been unavoidable to generate a sufficiently large dataset for cross-country research (as used in, for instance, Forbes, 2000 and Halter et al., 2014).[74] We use version 6.1 of the SWIID, which was published in October 2017. Whereas an earlier version of the dataset was critized by Jenkins (2015), Solt (2015) shows that most of this criticism is misplaced or has been solved since version 5.0 of the dataset.

To reflect data uncertainty, the SWIID reports 100 different imputations for every observation, which are generated via Monte Carlo simulations. Thus data uncertainty can be directly incorporated into the regression results via multiple imputation tools, or it can be ignored by averaging the imputations to generate one point estimate for each observation. As Section (5.5.3) shows that multiple imputation estimations hardly affect the estimated coefficients and standard errors, we primarily work with the point estimates.

Our standard redistribution variable REDIST is the difference between market and net Ginis, calculated from all available country-years in the SWIID. While this calculation allows for a large sample of data, caution is advised when interpreting this measure. Some of the Ginis of gross or net income inequality are estimates based on data from neighboring countries, which means that the difference between both measures of inequality contains little information about country specific redistribution. To address this problem, the SWIID reports a sub-sample of most reliable inequality data, for which a measure of redistribution (REDIST(S)) is explicitly provided. This sample is solely based on countries where survey data on net and gross incomes is available. Moreover, as historical data is often less reliable, it neglects observations from developing countries before 1985 and from advanced economies before 1975. Unfortunately, with only 408 country-years (5 year averages), the restricted sample is slightly less than half the size of the full sample, which is why we limit the use of the restricted sample to robustness checks and estimations that explicitly focus on

73 The online appendix (OA-3) gives a brief summary of the SWIID's standardization process, based on the extensive description in Solt (2016).

74 For example, it was long common to either ignore the difference between market income inequality and disposable income inequality or to simply assume that both measures differ by a fixed amount, irrespective of the obvious variations in the scope of the welfare state.

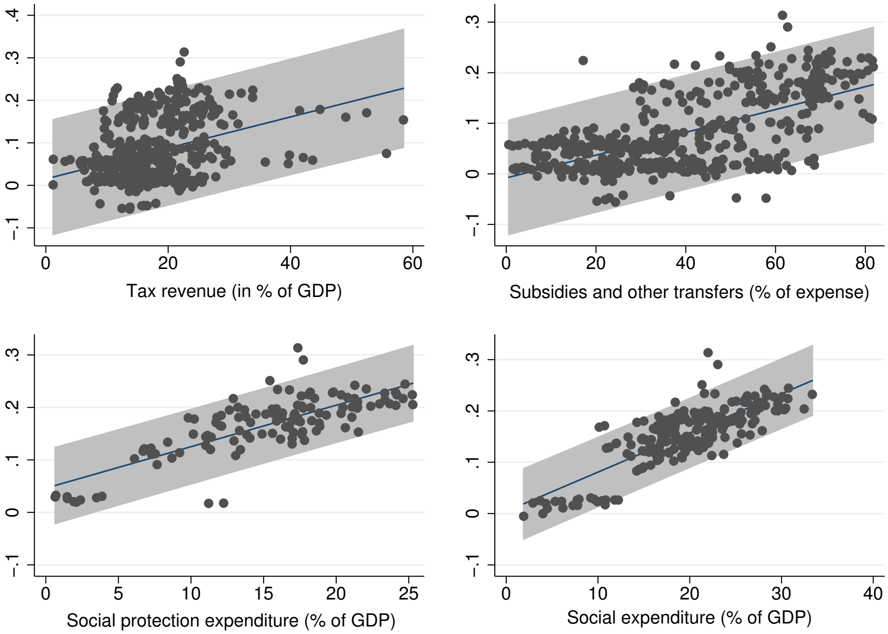


Figure 5.1 The relationship between REDIST (y-axis) and fiscal policy measures. The figure plots the pre-post redistribution variable REDIST from the SWIID against four proxies of distributive fiscal policy. The data on Tax revenue and Subsidies and other transfers are from the World Bank World Development Indicators. Social protection expenditure and Social expenditure are from the OECD National Accounts Database.

the effect of redistribution. Online appendix OA-4 provides a brief illustration on the extent of redistribution measured by REDIST and REDIST(S) across countries.

The pre-post approach of Equation (5.2) yields a measure of effective redistribution, illustrating the overall result of governmental redistribution via taxes and transfers, rather than the effort by which the result is achieved. Compared to earlier studies, this provides two advantages: First, as their redistributive impact is uncertain and varies considerably across countries, our analysis does not depend on rough redistribution measures such as marginal tax rates or social subsidies. Second, as pre-post redistribution data is (now) more widely available than data on redistributive fiscal policies, our study rests on a considerably expanded number of country-years.

Figure (5.1) plots our measure of effective redistribution against four proxies of governmental redistribution efforts, like social expenditures or tax revenues. It shows that there is a significant correlation between pre-post redistribution and distributive fiscal policy. Foremost, the correlations with social protection expenditure and social expenditure are strong (0.79% and 0.82%), suggesting that social spending is often effective in reducing inequality. Somewhat less pronounced is the correlation between redistribution and tax revenues (0.36%), which is not surprising, as the volume of tax revenues tells rather little about its usage or the progressiveness of taxation. Altogether, the strong but less than perfect correlations confirm that pre-post redistribution reflects the effect of redistributional policies, but it also suggests that the pre-post data contains additional information not captured by the fiscal policy measures (see, also Berg et al., 2014).

A potential drawback of the pre-post approach is that market inequality is not independent from the extent of public redistribution (Bergh, 2005). On the lower end of the income scale, a generous welfare system may boost gross inequality by encouraging low income earners to withdraw from the labor market and to live from transfers instead of market income. On the upper end, high income earners may be discouraged by taxes and thus reduce their labor supply, which lowers gross inequality. We follow Berg et al. (2014) by suggesting that the effect of redistribution on market inequality may not be substantial, as its effects on the lower and on the upper scale of the income distribution are offsetting.

5.4.2 The relationship between inequality and redistribution

The political economy mechanism of the endogenous fiscal policy channel suggests a positive relation between inequality of market incomes and redistribution. Empirical evidence on this channel, however, is rather ambiguous. Whereas earlier studies (e.g. Perotti, 1994, 1996) find a negative relationship between initial inequality and different proxies for redistribution, more recent studies conclude that societies with an unequal distribution of market incomes tend to redistribute more than others (see, e.g., Milanovic, 2000). One explanation for these contradicting results may be the lack of adequate measures for inequality and redistribution. Although the endogenous fiscal policy channel is triggered by the extent of market inequality, some earlier studies use net inequality to explain demand for redistribution. In addition, many studies rely on imperfect measures of redistribution, as the size of public transfers and taxes may be little indication of their redistributive impact.

Our dataset allows us to reconsider the endogenous fiscal policy channel by using market inequality and effective redistribution based on a large panel of countries. The data implies that lower levels of net inequality are the result of redistributional activities of the government, as the level of redistribution is strongly correlated with the extent of net inequality (-65% in the full sample and -74% in the restricted sample).

According to the endogenous fiscal policy channel, we would expect more redistribution in countries that feature a higher level of market inequality. Figure (5.2) illustrates the relationship between those variables in the restricted sample during the 2010–2014 period, pointing to a weak correlation between market inequality and redistribution. However, the figure reveals a distinct empirical pattern: countries that are located under the regression line almost entirely are from the developing world, whereas the sample of countries above the regression line consists primarily of advanced economies. Thus, the Meltzer-Richard effect seems to be much more prevalent in richer countries. An important conclusion of this observation is that

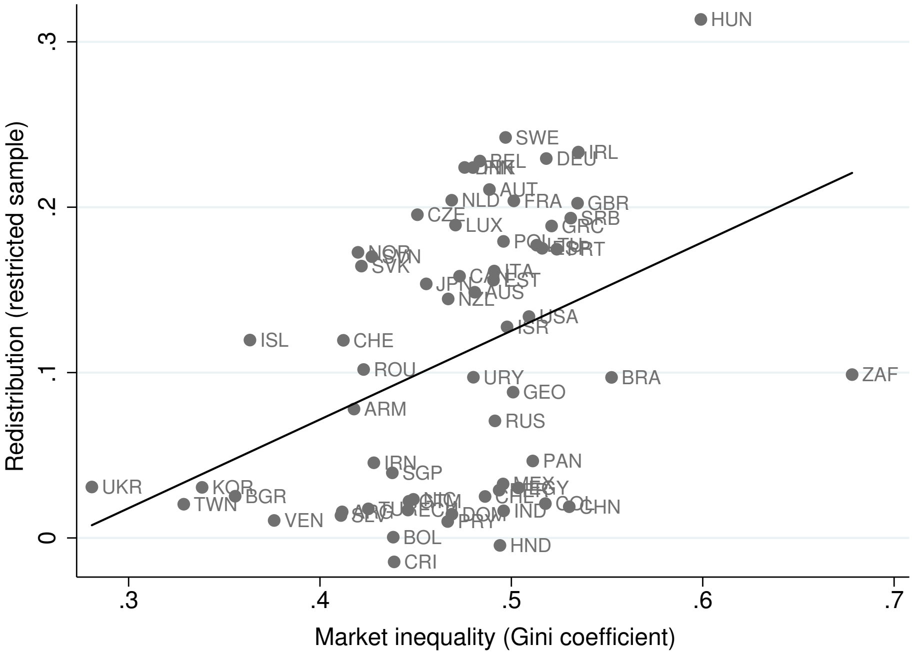


Figure 5.2 The relationship between market inequality and redistribution. The figure plots observations for each country in the most recent period (2010–2014). Data is from the restricted sample containing the most reliable data.

the effect of market inequality on redistribution must be examined while holding constant the development level of the economies.

Consider the simple reduced model

$$\text{REDIST(S)}_{it} = \alpha + \delta\text{GINI(M)}_{it} + \beta \log(\text{GDP}_{pc})_{it-1} + \eta_i + \xi_t + v_{it},$$

where the denotation of the variable is the same as in the previous section. Table (5.1) presents the results of the estimation of the model using Pooled OLS (POLS), Within-Group (WG), and 2SLS estimations. Whereas Column (1) neglects both η_i and ξ_t, Column (2) includes country fixed effects and Column (3) additionally incorporates period fixed-effects. Column (4) conducts 2SLS regressions with fixed-effects, where GINI(M) is instrumented with its lagged values.

The results support the political mechanism of the endogenous fiscal policy channel, as higher levels of market inequality are associated with higher amounts of redistribution in each of the estimations. Except for the two-way fixed-effects model displayed in Column (3), the development level is positively related to redistributive policies of the government. The income level may be a proxy of the deeper institutional causes that distinguish the countries in their level of redistribution. Due to higher transparency, more efficient institutions and less corruption, the opportunities for

Table 5.1 The relationship between market inequality and redistribution. Dependent variable is redistribution, REDIST(S).

	(1) POLS	(2) Within-Group	(3) Within-Group (time-dummies)	(4) Within-Group (2SLS)
GINI(M)	0.285***	0.297***	0.270***	0.277***
	(0.0807)	(0.0578)	(0.0571)	(0.0653)
log(GDP$_{pc}$)	0.0721***	0.0124***	-0.0110	0.0103**
	(0.00685)	(0.00427)	(0.00696)	(0.00425)
Constant	-0.721***	-0.151***	0.0664	
	(0.0769)	(0.0493)	(0.0709)	
Observations	399	399	399	350
Countries	65	65	65	64
R-squared	0.52	0.49	0.58	0.44

Notes: Table reports regressions of REDIST(S) on GINI(M) using Pooled OLS (Column 1), Within-Group without and with time-dummies (Columns 2 and 3), and 2SLS with country fixed effects (Column 4) estimations. Robust standard errors in parantheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

rent-seeking and crony capitalism decline during the development process. Likewise, less-developed countries tend to be less-democratic. If the voter cannot influence the political process, a higher level of inequality most likely does not yield a higher amount of redistribution.

5.5 Regression results

5.5.1 Baseline regressions

We now turn to the investigation of the growth effect of inequality and redistribution. Table (5.2) reports the results of our baseline system GMM growth estimations when the full sample of available data from the SWIID is used. Our regression sample covers a maximum of 969 observations from 164 countries. The time dimension includes 5-year averages from the initial period 1965–1969 to the period 2010–2014.[75]

Employing all available country-year observations, Column (1a) of Panel A shows a reduced specification of our growth model in which—aside from time dummies and country fixed-effects—the lagged level of per capita income is the only control variable. As mentioned previously, theory suggests that inequality exerts its influence on growth via several transmission channels. These channels involve standard growth determinants such as physical and human capital accumulation, fertility rates, and

[75] The number of included countries is lower than the number of countries for which inequality measures are available. For most countries, exclusion of the regressions sample is caused by unavailability of covariates, particularly in the comprehensive models. For some countries, the lag structure of the regression model prohibits inclusion, as inequality measures are only available for single periods (e.g. Nauru in 2005–2009, Myanmar in 2005–2009, Palau in 2005–2009, and Somalia in 2000–2005.)

Table 5.2 Baseline regressions, full sample. Dependent variable is real per capita GDP growth.

	(1a)	(1b)	(2)	(3)	(4)
	Panel A: Contemporaneous values of inequality and redistribution				
GINI(N)	-0.303***	-0.293***	-0.155***	-0.0781*	0.0660
	(0.0653)	(0.0723)	(0.0567)	(0.0461)	(0.0429)
REDIST	-0.276***	-0.262***	-0.207***	-0.129**	0.0464
	(0.0793)	(0.0785)	(0.0720)	(0.0594)	(0.0642)
L.log(GDP_{pc})	0.00298	0.00395	-0.0161***	-0.0214***	-0.0336***
	(0.00442)	(0.00490)	(0.00602)	(0.00475)	(0.00482)
INVS			0.132***	0.167***	0.0730***
			(0.0257)	(0.0213)	(0.0270)
SCHOOLING			0.00781***	0.00756***	0.00260
			(0.00203)	(0.00155)	(0.00176)
GOVC				-0.00970	-0.0358
				(0.0264)	(0.0249)
INFL				-0.00195	-0.00221
				(0.00165)	(0.00162)
OPEN				0.00481	0.00735
				(0.00795)	(0.00752)
POLRIGHT				0.000327	-0.000773
				(0.00148)	(0.00148)
log(LIFEEX)					0.0769***
					(0.0287)
log(FERT)					-0.0389***
					(0.00908)
Observations	969	813	813	813	813
Countries	164	132	132	132	132
Hansen p-val	0.079	0.181	0.240	0.406	0.888
Diff-Hansen	0.544	0.748	0.879	0.996	0.997
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.879	0.937	0.760	0.807	0.506
Instruments	88	88	108	142	162
Collapsed	No	No	Yes	Yes	Yes
	Panel B: Initial values of inequality and redistribution				
GINI(N)$(t-1)$	-0.344***	-0.241***	-0.103*	-0.0324	0.0488
	(0.0817)	(0.0804)	(0.0617)	(0.0582)	(0.0493)
REDIST$(t-1)$	-0.318***	-0.243***	-0.198***	-0.179**	-0.0347
	(0.0701)	(0.0792)	(0.0754)	(0.0856)	(0.0814)
Observations	865	737	737	737	737
Countries	163	132	132	132	132
Hansen p-val	0.011	0.041	0.090	0.544	0.891
Diff-Hansen	0.544	0.748	0.879	0.938	0.997
AR(1) p-val	0.000	0.001	0.000	0.000	0.000
AR(2) p-val	0.980	0.615	0.482	0.475	0.224
Instruments	88	88	108	142	162
Collapsed	No	No	Yes	Yes	Yes

Notes: Table reports two-step system GMM estimations with Windmeijer-corrected standard errors in parentheses. All regressions include period fixed effects. Hansen p-val gives the J-test for overidentifying restrictions. Diff-Hansen represents the Difference-in-Hansen statistic of the instrument subset for the level equation. AR(1) p-val and AR(2) p-val report the p-values of the AR(n) test. Instruments illustrates the number of instruments. Collapsed indicates whether the instrument matrix is collapsed to prevent instrument proliferation. The lag structure utilized in Panels A and B is identical. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

political stability. Thus, as pointed out by Galor (2009), the only way to identify the full growth effect of income inequality is excluding some of the usual controls.[76]

The results in Column (1a) suggest that both high net inequality, but also its cure in the form of public redistribution, are similarly bad for growth. The point estimate of the net Gini is negative and highly significant, suggesting that an increase of the Gini by ten percent points, i.e. roughly one standard deviation, lowers the annual growth rate by an average of 3.0 percentage points. The estimated parameter of redistribution is significantly negative as well and roughly the same size as the effect of inequality.[77]

Our sample varies somewhat when we include additional control variables. Thus, Columns (1b)–(4) rest on a sample that contains data for all of the control variables in order to enable a clear comparison between different regression models. The reduction in country-years in the common sample from 969 to 813 is primarily due to a loss of observations on both ends of the time dimensions, which similarly affects observations from advanced and developing economies. In this smaller sample, the effects of inequality and redistribution remain more or less unchanged in the reduced specification reported in Column (1b).

In Column (2), the investment share and the average years of school attainment are introduced into the model. Both variables are not only standard components of empirical growth models, but also—according to the theories of differential saving rates, credit-market imperfections, sociopolitical unrest, and endogenous fiscal policy—part of the transmission process from inequality to growth. Holding these transmission variables constant, we would expect this model to show that inequality has a smaller impact on growth. Indeed, the parameter estimate of the Gini declines to -0.155, which is about half of the marginal effect detected in Column (1b). In line with theory and previous empirical studies, the newly introduced controls are positive and significant.

When we introduce a number of additional control variables in Column (3), the effect of inequality shrinks further, but still remains significant. Finally, Column (4) controls for demographic factors. Some theoretical models suggest that fertility is endogenous to income inequality. Holding the fertility rate constant could thus eliminate another transmission channel. Indeed, in Column (4), the estimated effect of the Gini diminishes and becomes insignificant when fertility is held constant, which resembles the findings by Barro (2000) and De la Croix and Doepke (2003).[78] Similar to their results, the direct effect of fertility is negative and highly significant in our growth regression.

Panel B alters the analysis by investigating the effect of initial levels of inequality and redistribution, i.e. we lag both variables by one 5-year period. The empirical pattern found with respect to the parameters estimates of inequality and redistribution

76 By estimation of reduced models, we also avoid a potential "bad control" problem (see Angrist and Pischke, 2009).

77 The Wald test further implies that both effects are similar in size, as $H_0 : \delta = \gamma$ yields $p = 0.68$.

78 Our sample composition does not change from Column (3) to Column (4), which strengthens the evidence for the endogenous fertility channel.

reappear in this setting, although the negative effect of inequality already ceases to be significant in Column (3).

A simple comparison of the parameter estimates of inequality and redistribution in Columns (1a) or (1b) of Table (5.2) suggests that the positive growth effect from a lower level of net inequality is, on average, fully offset when the decline in inequality is achieved via taxes and transfers. However, the estimated coefficient of REDIST in Table (5.2) should be interpreted with caution. Whereas the maximum number of available observations is utilized here, the redistribution variable may be measured imprecisely in certain cases where estimates rest entirely on information from other countries. Hence, Table (5.3) replicates the baseline estimates but introduces REDIST(S) as redistribution variable, which is calculated from a subsample consisting of only the most reliable observations. In this case, our regression sample shrinks to a maximum of 399 observations from 65 countries. Regarding the reduced model of Column (1a), the estimated parameters of redistribution and net inequality shown in Panels A and B are similar to the results obtained from the full sample estimations but slightly smaller, particularly with respect to redistribution. As the table particularly drops observations from developing countries, this decline suggests that the effects of REDIST and GINI(N) are stronger pronounced in poorer economies (see Section 5.5.4). However, the results strongly resemble the findings of the baseline table in two essential points: i) the growth effects of redistribution and inequality are both negative and largely offset each other and ii) the decline in the parameter estimates across the specifications suggests that investment, schooling, and fertility are the main transmission variables of inequality to growth.

5.5.2 Tests for the validity and strength of the utilized instruments

Assessing the validity of our results, we refer to the test statistics given in the lower part of Tables (5.2) and (5.3). The first requirement is the absence of second-order serial correlation in the residuals, which does not pose any problem as the AR(2) p-value is always greater than 0.1. In addition, the p-values of Hansen's J-test reported in the comprehensive models of both tables suggest that the null of joint validity of all instruments cannot be rejected; and the Difference-in-Hansen tests emphasize the superiority of system GMM over difference GMM in each model. Yet, with p-values below 0.1, there could be some doubt about the validity of our instruments in the reduced models of Panel B and—to a much lesser extent–Panel A of Table (5.2). However, since Hansen's J-test is also a general test of structural specification, the rejection of the null in the reduced model may point to an omitted variable problem rather than indicating general invalidity of the instruments (see Roodman, 2009c). As we deliberately omit certain regressors to capture the full impact of inequality and to circumvent potential problems with "bad controls", a rejection of the reduced specification is not surprising.

In specifying our instrument matrix, we carefully attended to the concern of weak instruments. Bazzi and Clemens (2013) showed that some of the instrumental variables in widely-cited growth regressions may be weak, casting doubt on the concluded consequences for economic development. Kraay (2015) argues that this

Table 5.3 Baseline regressions, restricted sample. Dependent variable is real per capita GDP growth.

	(1a)	(1b)	(2)	(3)	(4)
	Panel A: Contemporaneous values of inequality and redistribution				
GINI(N)	-0.218***	-0.181***	-0.0918*	-0.0519	0.0409
	(0.0602)	(0.0638)	(0.0525)	(0.0554)	(0.0740)
REDIST(S)	-0.125**	-0.0946*	-0.0651	0.0231	0.0127
	(0.0543)	(0.0544)	(0.0715)	(0.0778)	(0.0624)
L.log(GDP$_{pc}$)	-0.0182**	-0.0165**	-0.0286***	-0.0328***	-0.0290***
	(0.00801)	(0.00754)	(0.00660)	(0.00776)	(0.00541)
INVS			0.158***	0.163***	0.0791*
			(0.0383)	(0.0406)	(0.0460)
SCHOOLING			0.00516**	0.00659***	0.00427**
			(0.00219)	(0.00239)	(0.00180)
GOVC				-0.0782*	-0.0957***
				(0.0444)	(0.0340)
INFL				-0.00114	-0.00117
				(0.000998)	(0.000955)
OPEN				-0.00224	0.00269
				(0.00582)	(0.00519)
POLRIGHT				-0.00229	-0.00106
				(0.00224)	(0.00177)
log(FERT)					-0.0373***
					(0.0116)
log(LIFEEX)					0.0113
					(0.0703)
Observations	399	389	389	389	389
Countries	65	63	63	63	63
Hansen p-val	0.287	0.372	0.222	0.275	0.636
Diff-Hansen	0.836	0.976	0.968	0.730	0.951
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.194	0.178	0.147	0.116	0.148
Instruments	69	69	56	68	74
Collapsed	No	No	Yes	Yes	Yes
	Panel B: Initial values of inequality and redistribution				
GINI(N)	-0.231***	-0.188***	-0.118**	-0.0692	-0.0000289
	(0.0497)	(0.0585)	(0.0532)	(0.0487)	(0.0672)
REDIST(S)	-0.203***	-0.158**	-0.127	-0.0477	-0.0355
	(0.0679)	(0.0667)	(0.0857)	(0.0834)	(0.0686)
Observations	343	334	334	334	334
Countries	65	63	63	63	63
Hansen p-val	0.083	0.106	0.100	0.207	0.530
Diff-Hansen	0.510	0.749	0.248	0.788	0.974
AR(1) p-val	0.001	0.001	0.001	0.000	0.000
AR(2) p-val	0.415	0.451	0.351	0.442	0.152
Instruments	62	62	55	69	77
Collapsed	No	No	Yes	Yes	Yes

Notes: Table reports two-step system GMM estimations with Windmeijer-corrected standard errors in parentheses. All regressions include period fixed effects. Hansen p-val gives the J-test for overidentifying restrictions. Diff-Hansen represents the Difference-in-Hansen statistic of the instrument subset for the level equation. AR(1) p-val and AR(2) p-val report the p-values of the AR(n) test. Instruments illustrates the number of instruments. Collapsed indicates whether the instrument matrix is collapsed to prevent instrument proliferation. The lag structure utilized in Panels A and B is identical. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

problem is particularly severe with respect to empirical investigations on the effect of inequality on growth. Table (A5-1) in the appendix follows the suggestion of Bazzi and Clemens (2013) to open the "black box" of GMM by providing two tests proposed by Sanderson and Windmeijer (2016) (henceforth SW). The first is a weak instrument F-test that builds on Angrist and Pischke (2009) but allows for separate diagnostics for each endogenous regressor. The second is the SW χ^2 test for under-identification, which is also reported separately for each regressor. In general, these tests are designed for weak-instrumentation diagnostics of external instruments in traditional 2SLS settings. However, there have been some attempts to transfer these tests to dynamic panel GMM settings (Bun and Windmeijer, 2010; Newey and Windmeijer, 2009) via construction of the exact GMM instrument matrix for both the difference and the levels equation of the system GMM estimator, which can in turn be used to carry-out the standard 2SLS regressions and test.

With respect to our redistribution variable, the SW tests of weak instrumentation show that in each model specification, the relative IV bias is less than 30 percent in both the level and the difference equation, and often even (much) smaller than 10 percent. The test also points to a general instrument strength of our inequality variable, albeit to a slightly lesser extent. The SW χ^2 test shows that underidentification is not a problem in either the levels or the difference equation with respect to both GINI(N) and REDIST. As additional weak instrument diagnostics, we replicate a battery of tests conducted by Kraay (2015), including weak-instrument-robust tests on (joint) significance of the endogenous regressors, as well as weak-instrument-robust confidence intervals. Both the AR-test developed by Anderson and Rubin (1949) and the K-test proposed by Kleibergen (2005) demonstrate the significance of the model specification and our variables of interest.[79]

Weak-instrument-robust confidence intervals are computed based on the conditional likelihood ratio test (CLR) developed by Moreira (2003) and compared with the corresponding intervals suggested by the Wald test.[80] The computed CLR intervals are robust to weak instruments in the sense that they have the correct size in cases when instruments are weak as well as when they are not. We focus on Column (1a) of Table (5.2) and (5.3), where the results suggest a significantly negative effect of both inequality and redistribution. The confidence intervals derived from the CLR test are (slightly) wider than the Wald confidence intervals. However, the weak-instrument-robust intervals are entirely in the negative parameter space, suggesting that even in a potential presence of weak instrumentation, the effects of both variables would still be negative.

[79] In order to obtain test statistics on the significance of GINI(N) and REDIST via the AR-test and the K-test, we use the reduced-model of Column (1a) and assume that the remaining variables are strongly identified.

[80] The results are not directly comparable to the effects identified in Tables (5.2) and (5.3) as the CLR statistic can only be inverted to obtain weak-instrument-robust confidence intervals in the single-endogenous-regressor case (Finlay and Magnusson, 2009; Finlay et al., 2016), prompting us to re-specify the empirical model so that the remaining variables are treated as exogenous. In the case of multiple endogenous regressors, such inference can only be carried out using projection-based confidence intervals that may be computed by a grid search. These intervals, however, are conservative, meaning that they have asymptotic size less than or equal to nominal size.

5.5.3 Sensitivity analysis of the baseline results

The SWIID reports 100 different imputations for every observation in order to reflect the uncertainty that goes in hand with the generation of consistent series of inequality data. As it is common in the literature (see, e.g Berg et al., 2014 and Acemoglu et al., 2015), we have thus far averaged this data to generate point estimates, which we can handle with regular regression techniques. In this section, we account for a potential bias in the estimation caused by data uncertainty. Specifically, we follow the the suggestion of Solt (2016) by running multiple imputation regressions. Essentially, the multiple imputation routine estimates repeated regressions for each of the 100 imputations of GINI(N) and REDIST and then pools the results according to the combination rules of Rubin (1987). The estimated standard errors are thus adjusted for the variability between the imputations and hence usually larger than those received by averaged data.[81]

Table (5.4) presents the estimated coefficients of GINI(N) and REDIST resulting from multiple imputation regressions. The reported regression models exactly resemble the baseline system GMM specifications from Table (5.2), except that they are estimated with the full set of imputations instead of the averaged data. The results are comparable to the baseline outcomes, but the standard errors of GINI(N) and REDIST are slightly larger, leading to slightly less significant effects in Columns (2) and (3). Altogether, however, our baseline results are robust to incorporation of the sampling error of the SWIID imputations, which is why we can safely proceed with the standard system GMM estimator.

In the online appendix, we further examine the stability of our results. First, as different estimation techniques may yield different implications for the growth effect of inequality (see Neves and Silva, 2014), Table (OT-2) in the online appendix presents the results from alternative estimation strategies. These techniques include system GMM techniques in which the instruments are replaced by their principal components to reduce the instrument count (Bai and Ng, 2010; Kapetanios and Marcellino, 2010), difference GMM, 2SLS, and simple OLS regressions. The negative effect of inequality and redistribution is visible in each of the reduced specifications.[82]

Second, to assess the level of significance of the redistribution effect, Table (OT-3) in the online appendix documents the aggregate growth effect of public redistribution

---

81 See Brownstone and Valletta (2001) and Jenkins (2015) for a more detailed summary of the multiple imputations technique.

82 Note that difference GMM results in a decline in the number of observations from 969 to 805. The reason is that the estimator requires having at least three consecutive observations for each of the regressors, thereby magnifying gaps in our sample. In addition, the Difference-in-Hanson statistics reported in Table (5.2) emphasize that the extra moment conditions of system GMM are valid, resulting in substantial efficiency losses when using first-difference GMM. Contrary to some earlier studies that rely on time-series variation (e.g. Li and Zou, 1998 and Forbes, 2000), our first-difference GMM estimates support the negative effect of net inequality on growth found in our baseline estimates. The reason for the deviation is twofold. First, earlier studies are based on a substantially lower number of countries. As we might expect the negative effect of net inequality to be more pronounced in poor countries, the neglect of data from the developing world yields a bias in the estimation. Second, previous studies on the topic largely ignore the incomparability problems that arise when using different data compilations, which is heavily criticized by Solt (2016).

Table 5.4 Baseline regressions using multiple imputations (MI) estimates (MI $= 100$). Dependent variable is real per capita GDP growth.

	(1a)	(1b)	(2)	(3)	(4)
	Panel A: Contemporaneous values of inequality and redistribution				
$\text{GINI(N)}_{\text{MI}}$	-0.316***	-0.307***	-0.171**	-0.106**	0.0359
	(0.0774)	(0.0862)	(0.0670)	(0.0535)	(0.0512)
$\text{REDIST}_{\text{MI}}$	-0.208**	-0.230**	-0.161**	-0.131*	0.0165
	(0.100)	(0.103)	(0.0785)	(0.0735)	(0.0799)
Observations	969	813	813	813	813
Countries	164	132	132	132	132
MI F-Stat	5.489	5.407	9.109	11.72	14.75
MI F p-value	0.000	0.000	0.000	0.000	0.000
Average RVI	0.131	0.155	0.129	0.105	0.076
Largest FMI	0.353	0.373	0.342	0.375	0.323
Imputations	100	100	100	100	100
Instruments	88	88	108	142	162
Collapsed	No	No	Yes	Yes	Yes
	Panel B: Initial values of inequality and redistribution				
$\text{GINI(N)}_{\text{MI}}(t-1)$	-0.319***	-0.235***	-0.0380	0.00415	0.0644
	(0.0774)	(0.0814)	(0.0625)	(0.0625)	(0.870)
$\text{REDIST}_{\text{MI}}(t-1$	-0.250***	-0.200**	-0.0957	-0.0855	-0.00800
	(0.0905)	(0.0987)	(0.0771)	(0.0925)	(0.577)
Observations	865	737	737	737	737
Countries	163	132	132	132	132
MI F-Stat	4.516	3.611	7.189	9.148	0.929
MI F p-value	0.000	0.000	0.000	0.000	0.538
Average RVI	0.155	0.170	0.655	0.097	0.024
Largest FMI	0.404	0.464	0.884	0.286	0.006
Imputations	100	100	100	100	100
Instruments	88	88	108	142	160
Collapsed	No	No	Yes	Yes	Yes

Notes: Table reports multiple imputations two-step system GMM estimations with Windmeijer-corrected standard errors in parentheses. Control variables and specifications of Panels A and B are identical to those applied in the corresponding columns in Table (5.2). All regressions include period fixed effects. MI F-Stat gives the F-statistic of the multiple imputation estimations, MI F p-value reports the referring p-values. Average RVI documents the average relative variance increase due to nonresponse, largest FMI reports the largest fraction of missing information. Instruments illustrates the number of instruments. Instruments are the second lag of the explanatory variables in levels for the difference equation and the first lag in differences for the level equation. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

in the restricted sample of high quality data. Leaving net inequality open, the estimated parameter of redistribution captures both the *direct* incentive effect of redistributive taxes and transfers plus the *indirect* effect resulting from the change in net inequality. The Gini of market inequality, GINI(M), which is possibly affected by some feedback effects of redistribution, is kept constant in this case. The results support our previous outcomes by detecting no significant effect of redistribution in either direction.

Finally, the baseline results are stable when we estimate the growth effect of relative redistribution, i.e. the ratio of REDIST to GINI(M) (see Table (OT-4) in the online appendix).

5.5.4 Different development levels

The basic regression results suggest that inequality and growth are negatively related. However, this conclusion is based on the whole sample, whereas previous studies suggest that the effect of inequality on growth varies across different development levels (see Barro, 2000, Galor and Moav, 2004, and Castelló-Climent, 2010).

Table 5.5 The impact of inequality and redistribution for different levels of development, estimated via interaction terms. Dependent variable is real per capita GDP growth.

	(1a)	(1b)	(2)	(3)	(4)
	Panel A: The effect of inequality at different development levels				
GINI(N)	-2.707***	-2.653***	-2.072***	-2.101***	-1.215***
	(0.227)	(0.278)	(0.200)	(0.229)	(0.238)
GINI×L.log(GDP$_{pc}$)	0.295***	0.289***	0.225***	0.227***	0.139***
	(0.0247)	(0.0323)	(0.0219)	(0.0254)	(0.0256)
L.log(GDP$_{pc}$)	-0.113***	-0.112***	-0.0960***	-0.0975***	-0.0771***
	(0.0103)	(0.0141)	(0.00936)	(0.00956)	(0.00924)
Observations	969	813	813	813	813
Countries	164	132	132	132	132
Hansen p-val	0.185	0.302	0.318	0.145	0.514
Diff-Hansen	0.627	0.916	0.653	0.869	0.940
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.286	0.235	0.463	0.345	0.102
Instruments	114	89	118	109	142
Joint p-val	0.000	0.000	0.000	0.000	0.000
	Panel B: The effect of redistribution at different development levels				
REDIST	1.364***	1.275***	0.644*	0.585	-0.0101
	(0.441)	(0.401)	(0.373)	(0.361)	(0.342)
REDIST×L.log(GDP$_{pc}$)	-0.144***	-0.136***	-0.0722*	-0.0649*	-0.00195
	(0.0456)	(0.0427)	(0.0395)	(0.0393)	(0.0345)
L.log(GDP$_{pc}$)	0.0137**	0.0131**	-0.0138**	-0.0172***	-0.0353***
	(0.00533)	(0.00555)	(0.00653)	(0.00628)	(0.00499)
Observations	969	813	813	813	813
Countries	164	132	132	132	132
Hansen p-val	0.090	0.230	0.095	0.276	0.436
Diff-Hansen	0.471	0.230	0.794	0.976	0.969
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.988	0.953	0.892	0.798	0.465
Instruments	114	88	117	108	141
Joint p-val	0.007	0.006	0.109	0.250	0.777

Notes: Table reports two-step system GMM estimations with Windmeijer-corrected standard errors in parentheses. Covariates are identical to those in the corresponding model specification of the baseline results reported in Table (5.2). All regressions include period fixed effects. Hansen p-val represents the J-test for overidentifying restrictions. Diff-Hansen gives the Difference-in-Hansen statistic of the instrument subset for the level equation. AR(1) p-val and AR(2) p-val report the p-values of the AR(n) test. Instruments illustrates the number of instruments. Joint p-val shows the p-values on the Wald test for joint significance of inequality (Panel A) and redistribution (Panel B) and the respective product with the moderator variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The upper panel of Figure (5.3) illustrates the marginal growth effect of the Gini coefficient for different development levels and the associated 90 percent confidence interval.[83] The underlying model is Column (1a) of Panel A of Table (5.5), where we introduce an interaction term between the Gini coefficient and initial incomes, denoted by GINI$\times$L.log(GDP$_{pc}$). This inclusion allows for investigation of the impact of inequality without relying on fixed threshold values to distinguish between development levels. We conduct the analysis identically to the baseline specification; however, for reasons of lucidity, Table (5.5) only reports the interacting variables, as there are virtually no changes in the effects of the covariates.

The figure illustrates that the marginal effect of net inequality on growth is significantly negative in poor economies. Yet the impact of an unequal distribution of incomes weakens as the economies develop and eventually turns positive once incomes exceed 18,000 USD.

The lower panel of Figure (5.3) illustrates the results from a similar analysis conducted in Panel B of Table (5.5) concerning the influence of redistribution across different levels of development by inclusion of REDIST$\times$L.log(GDP$_{pc}$). The figure highlights that redistribution contributes positively to economic growth in earlier stages of development. However, this positive effect vanishes once the economies reach an average income level of approximately 15,000 USD. Overall, our findings emphasize the need to distinguish between the development level when aiming to evaluate the effect of redistributive policies.

5.6 Empirical investigation of the transmission channels

In the previous regressions the effects of inequality and redistribution diminish when we control for investment, schooling, and fertility. This could be evidence that inequality exerts its influence on growth via transmission channels that act specifically through these variables. Holding constant the transmission variables would thus remove the growth effect of inequality, which means that the reduced model would be preferable. Yet the direction of the effect is unclear. Whereas the theoretical models discussed in Section 5.2 predict a causal effect of inequality and redistribution on the transmission variables, the reverse causation may be plausible as well.

Taking this problem into account, Table (5.6) directly examines how inequality and redistribution affect the suspected transmission variables. Each of the major transmission variables—investment, schooling and fertility—is regressed on the reduced empirical specification that we employed in the previous sections to explain economic growth. This approach provides the advantage of good comparability among the transmission regressions and our main growth regressions.[84] All regressions are based on the common sample with 325 country-years to maximize comparability of the results between the different model specifications. The variables of interest are lagged by one period to measure the effects of inequality and redistribution on the transmission variables, rather than the reverse.

83 The figures illustrating interaction effects with continuous modifying variables are based on the algorithm suggested by Brambor et al. (2006).

84 Although system GMM is designed for dynamic models, it does not require the dependent variable to appear on the right hand side (see Roodman, 2009b).

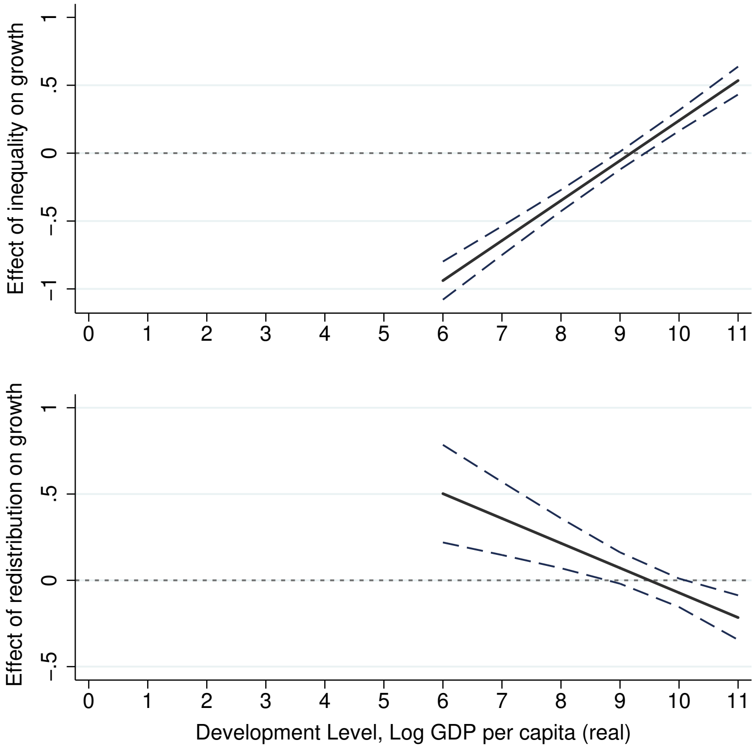


Figure 5.3 The marginal effect of inequality (upper panel) and redistribution (lower panel) on growth at different levels of development: values are calculated using the results of the growth regressions in Column (1a) of Panels A and B of Table (5.5). The black solid line plots the marginal effect of inequality and redistribution at various levels of development. Surrounding dashed lines represent the 90% confidence intervals.

The first column of Table (5.6) shows that both inequality and redistribution are negatively related to investment, which could be due to the sociopolitical unrest or credit market imperfections channel. Moreover, the incentive effects of redistribution seem to matter for investment decisions, which is not surprising as progressive taxes lower the return on investment.

The results from the schooling and fertility regressions in Columns (2) and (3) are in line with our expectations from theory and the reduced-form estimates. Whereas inequality has a positive effect on the fertility rate, it negatively affects school attainment. Redistribution is insignificantly related to schooling, but significantly increases fertility.

The findings of Column (1)–(3) explain why the effects of inequality and redistribution disappear in the more comprehensive models of our baseline growth

Table 5.6 Transmission channels of inequality, restricted sample. Dependent variables are investment shares (INVS), fertility (FERT), and schooling (SCHOOLING).

	(1) INVS	(2) SCHOOLING	(3) FERT	(4) INVS	(5) SCHOOLING	(6) FERT
log(GDP$_{pc}$)	0.0692*	-0.472	-0.0441	0.0762**	-0.357	-0.342
	(0.0372)	(0.397)	(0.337)	(0.0359)	(0.432)	(0.258)
GINI(N) $(t-1)$	-0.807**	-12.60***	11.66***	-0.585	-17.67***	10.62***
	(0.380)	(3.754)	(3.500)	(0.376)	(4.947)	(1.699)
REDIST(S) $(t-1)$	-0.936***	8.030	7.673**	-0.692**	4.414	4.124*
	(0.333)	(5.387)	(3.597)	(0.316)	(4.252)	(2.282)
CREDIT				-0.00129	-0.0189*	0.0244***
				(0.00134)	(0.0111)	(0.00543)
GINI$\times$CREDIT				0.00393	0.0706*	-0.0622***
				(0.00387)	(0.0394)	(0.0144)
Observations	325	325	325	325	325	325
Countries	62	62	62	62	62	62
Hansen p-val	0.158	0.267	0.167	0.337	0.128	0.185
Diff-Hansen	0.932	0.267	0.247	0.936	0.655	0.672
AR(1) p-val	0.056	0.013	0.014	0.009	0.001	0.008
AR(2) p-val	0.153	0.725	0.383	0.165	0.292	0.294
Instruments	33	33	33	53	53	53
Joint p-val				0.214	0.004	0.000

Notes: Table reports two-step system GMM estimations with Windmeijer-corrected standard errors in parentheses. All regressions include period fixed effects. Hansen p-val gives the J-test for overidentifying restrictions. Diff-Hansen represents the Difference-in-Hansen statistic of the instrument subset for the level equation. AR(1) p-val and AR(2) p-val report the p-values of the AR(n) test. Instruments illustrates the number of instruments. Joint p-val shows the p-values on the Wald test for joint significance of GINI(N) and its product with the respective moderator variable. Joint p-val shows the p-values of the Wald test for joint significance of GINI(N) and its product with CREDIT for all interaction models. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

regressions. Table (A5-2) illustrates how the separate introduction of each of the transmission variables affects the estimated coefficients of inequality and redistribution. To avoid that different sample compositions yield changes in the point estimates, all estimations are based on a common regression sample. Whereas inequality and redistribution exert sizable negative effects in the reduced model, the effect of inequality declines by roughly 40% when the model controls for investment (Column 2) and school attainment (Column 3). Finally, the coefficients of both inequality and redistribution become insignificant after holding constant the fertility rate, which reflects the positive effects that GINI(N) and REDIST exert on fertility in Table (5.6).

As poorer families could finance human or physical capital investments via credit, the impact of inequality may depend on the state of financial development. Empirically, such a conditional effect can be examined by the introduction of an interaction term into the model. Ideally, we would want to introduce an interaction term between the Gini and a moderator variable that directly measures the degree of imperfections in capital markets. As such a variable does not exist in the cross-country context, the ratio of private credit to GDP (CREDIT) serves as a proxy for credit availability.[85]

[85] We instrument the credit ratio and the interaction term with their lagged values, as they are possibly endogenous to growth. The data source of CREDIT is World Bank (2017).

Indeed, the results from Columns (4)–(6) reveal that the net Gini and its interaction term with CREDIT are individually and jointly significant in both the schooling and the fertility regression, but insignificant in the investment regression.[86] In contrast to the model displayed in Column (1), the effect of inequality on investment becomes insignificant if the model accounts for credit availability. Yet the negative effect of redistribution on investment persists. Columns (5) and (6) imply that the negative effect of inequality on schooling as well as the positive effect of inequality on fertility are considerably stronger the lower the availability of credit. Poor families seem to choose a higher quantity of children if they are unable to finance their children's education due to credit market restrictions. Hence, the data provide evidence for the endogenous fertility and the credit market imperfection channel.

In Table (5.7) we test whether the conditional relationship between inequality and the transmission variables also applies to the effect of inequality on growth (Panel A). Therefore, we introduce the interaction term between inequality and credit availability in the baseline models of Table (5.2). In the reduced model reported in Columns (1a) and (1b), both the Gini and the interaction term with the credit to GDP ratio are highly significant, individually and jointly. Based on the results from this regression, the solid upwards-sloping line in Figure (5.4) plots the marginal growth effect of inequality across different levels of CREDIT. As indicated by the dashed 90 percent confidence bands, the marginal effect of inequality is negative at low values of CREDIT, but becomes insignificant at a credit to GDP ratio of roughly 60 percent, which is located around the 75th percentile of our sample. However, only at very high levels of CREDIT the effect of inequality turns significantly positive. The critical value lies at a credit to GDP ratio of about 130 percent, which is located above the 90th percentile of the sample.

Our regressions of the transmission variables suggest that much of the negative influence of inequality on growth results from forgone investments in human capital. In addition, some of the most productive investment opportunities (in regard to human or physical capital) may be replaced by less productive alternatives. Yet we can only control for the quantity of investments, and not for their average productivity. This might be one reason why the interaction effect shrinks, but still remains significant when we control for the investment share and the average years of schooling in Columns (2) and (3). Resembling the baseline results, the effect of inequality only vanishes when the fertility rate is held constant (Column 4), so that another element of the credit market imperfection channel is eliminated.

Finally, the growth effect of inequality could depend on the volume of public spending on education, which could ease the access to education for the poor. Indeed, the negative marginal effect of inequality on growth seems to be stronger if public education spending is low. Figure (5.4) plots the marginal effect of inequality based on a regression model that includes an interaction term between the net Gini and the ratio of public education spending to GDP (PSE).[87] When education spending

[86] See the p-values on the Wald tests of joint significance, given in the last line of Table (5.6).

[87] The data source of PSE is World Bank (2017).

Table 5.7 Interaction between the Gini coefficient, credit, and public spending on education. Dependent variable is real per capita GDP growth.

	(1a)	(1b)	(2)	(3)	(4)
	Panel A: Baseline results with credit availability				
GINI(N)	-0.252***	-0.212***	-0.135*	-0.0997	0.0741
	(0.0653)	(0.0655)	(0.0695)	(0.0628)	(0.0652)
REDIST	-0.109*	-0.125**	-0.0948	-0.0774	0.0651
	(0.0598)	(0.0576)	(0.0607)	(0.0745)	(0.0647)
CREDIT	-0.00128***	-0.00100***	-0.000614**	-0.000630***	-0.000154
	(0.000335)	(0.000270)	(0.000245)	(0.000242)	(0.000236)
GINI×CREDIT	0.00310***	0.00236***	0.00145*	0.00151**	0.000148
	(0.000966)	(0.000788)	(0.000747)	(0.000748)	(0.000674)
	(0.00398)	(0.00412)	(0.00477)	(0.00447)	(0.00512)
Observations	917	787	787	787	787
Countries	160	131	131	131	131
Hansen p-val	0.178	0.293	0.218	0.422	0.838
Diff-Hansen	0.406	0.555	0.798	0.905	0.984
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.928	0.880	0.703	0.767	0.501
Instruments	100	100	104	138	151
Joint p-val	0.000	0.003	0.115	0.117	0.061
	Panel B: Baseline results with public spending on education				
GINI(N)	-0.608***	-0.411***	-0.183*	-0.192*	0.0380
	(0.152)	(0.121)	(0.107)	(0.106)	(0.0827)
REDIST	-0.162*	-0.103	-0.0721	-0.111	-0.0194
	(0.0878)	(0.0660)	(0.0592)	(0.0832)	(0.0582)
PSE	-3.131***	-2.162**	-1.213	-1.295*	-0.146
	(1.020)	(0.852)	(0.753)	(0.780)	(0.542)
GINI×PSE	6.347**	4.107*	3.404*	3.430	-0.0729
	(2.546)	(2.155)	(1.855)	(2.166)	(1.357)
Observations	738	661	661	661	661
Countries	150	127	127	127	127
Hansen p-val	0.348	0.330	0.322	0.420	0.766
Diff-Hansen	0.789	0.867	0.562	0.681	0.975
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.949	0.790	0.748	0.656	0.737
Instruments	88	124	119	134	140
Joint p-val	0.000	0.024	0.178	0.186	0.749

Notes: Table reports two-step system GMM estimations with Windmeijer-corrected standard errors in parentheses. All regressions include time dummies. Covariates are identical to those in the corresponding model specification of the baseline results reported in Table (5.2). Hansen p-val gives the J-test for overidentifying restrictions. Diff-Hansen represents the Difference-in-Hansen statistic of the instrument subset for the level equation. AR(1) p-val and AR(2) p-val report the p-values of the AR(n) test. Instruments illustrates the number of instruments. Joint p-val shows the p-values on the Wald test for joint significance of GINI(N) and its product with the respective moderator variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

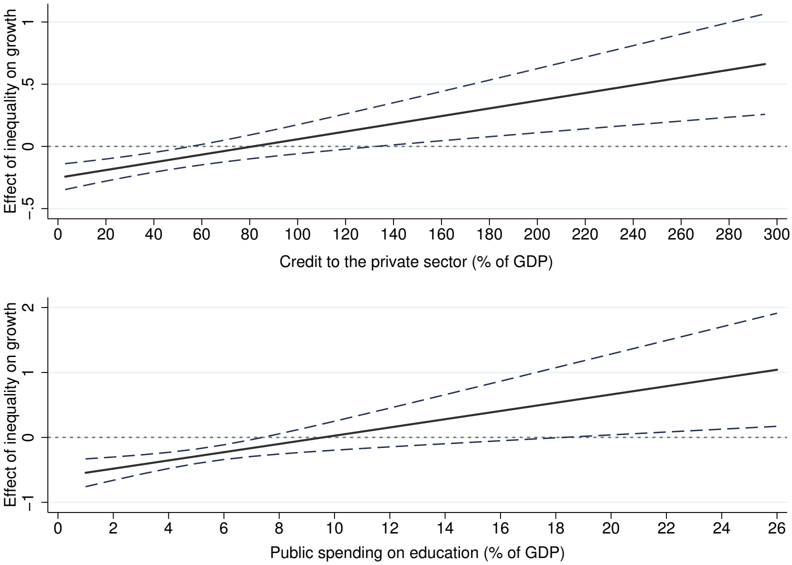


Figure 5.4 The conditional effects of inequality. The graph illustrates the marginal effect of inequality on growth across different levels of credit availability (upper panel) and different levels of public education spending. The graphs are computed based on the estimates of Column (1a) in Panels A and B of Table (5.7). The solid line plots the marginal effect of inequality. Surrounding dashed lines represent the 90% confidence intervals.

increases, the negative effect of inequality diminishes, becoming insignificant once a level of roughly 7 percent is passed.

In summary, this section shows that inequality exerts its influence on growth by reducing the average level of human capital and increasing the fertility rate, particularly in countries where credit availability is low. The negative effect of inequality on physical capital investments can be mitigated if access to capital markets is guaranteed, but the impact of public redistribution via taxes and transfers is negative regardless of credit availability. In addition, redistribution raises the fertility rate. Finally, a highly developed public education system seems to mitigate the negative effect of inequality on growth. These findings also explain the change in the effect of inequality and redistribution across different levels of development documented in Section (5.5.4): in poorer countries, public schooling systems and capital markets are considerably less developed than in advanced economies. Consequently, the growth effect of inequality is significantly negative and redistribution may be a policy measure to overcome the roots of this negative effect. As the economies become richer, the transmission mechanisms become increasingly irrelevant, which is why the negative effect of inequality diminishes.

5.7 Conclusions

Based on a current set of harmonized worldwide data, this chapter finds that income inequality has a robust negative effect on growth when the transmission variables of inequality are left open. By showing that less equal societies tend to have a less educated population and higher fertility rates, in particular when credit availability is low, the results support the credit market imperfections and the endogenous fertility channel. Moreover, inequality hampers physical capital investment if credit availability is limited.

In line with the political economy mechanism of the endogenous fiscal policy channel, this chapter finds that a higher level of market inequality predicts more public redistribution. Moreover, redistribution by taxes and transfers seems to directly harm economic growth when net inequality is held constant. We find evidence that this may be due to an impairment of physical capital investment and an increase in the fertility rate.

When estimating the aggregate growth effect of redistribution—its direct negative effect combined with its indirect positive effect resulting from lower net inequality—our results suggest that both effects are offsetting. Thus, at a given level of market inequality, redistribution seems to be a free lunch. Nonetheless, the most growth friendly environment is a low level of net inequality that stems from an equable distribution of market incomes, but not from redistributive taxes and transfers.

Finally, the findings show that the growth effects of inequality and redistribution vary with the development level. A negative impact of inequality prevails in developing and middle-income countries, where the negative potential for inequality is severe due to capital market imperfections and an insufficient provision of public goods. In high income countries, where opportunities are on average distributed more equally, no significant correlation between inequality and growth occurs. Likewise, the analysis reveals that redistribution by taxes and transfers is beneficial for growth in poor countries, but rather harmful in rich economies.

A relatively new branch of the literature decomposes overall inequality into several categories, particularly distinguishing between inequality of opportunity (IO) and inequality of effort or outcomes (Marrero and Rodriguez, 2013; Roemer and Trannoy, 2016; Marrero et al., 2016). This literature argues that IO is harmful to growth, while inequality of effort may be growth-enhancing due to incentive effects. Our results provide further evidence in this direction by identifying a strong negative effect of inequality in lower-developed nations with a higher average IO, whereas this effect vanishes in affluent countries where inequality of effort is (much) more prevalent (see Ferreira and Gignoux, 2011; Roemer and Trannoy, 2016). Our analysis also provides the implication that public spending on education as well as financial development may mitigate the negative growth-impulses of IO.

Two paths for future research remain: First, it is still possible that a low level of education and a high fertility rate are the *cause* rather than the *effect* of inequality, which is why more research is necessary in order to fully rule out that results are driven by feedback effects. Second, as the pre-post approach measures effective redistribution, it does not provide insights on the growth effects of specific redistributive policies. Future research should identify and analyze the policy instruments by which redistribution is achieved, and, in doing so, determine how it can be accomplished most efficiently.

5.A Appendix of Chapter (5)

Table A5-1 Weak instrument diagnostic of the baseline results.

	Model Specification of Baseline Table				
	(1a)	(1b)	(2)	(3)	(4)
	Panel A: Levels-Equation				
Weak IV tests					
Sanderson-Windmeijer F-Stat (GINI(N))	20.43	16.31	19.14	15.32	17.74
Sanderson-Windmeijer F-Stat (REDIST)	22.62	15.37	20.55	17.44	17.37
Stock-Yogo maximal IV relativ bias $\leq$ 30%	4.35	4.35	4.35	4.25	4.23
Under-identification tests					
Sanderson-Windmeijer χ^2 p-val (GINI(N))	0.000	0.000	0.000	0.000	0.000
Sanderson-Windmeijer χ^2 p-val (REDIST)	0.000	0.000	0.000	0.000	0.000
	Panel B: Difference-Equation				
Weak IV tests					
Sanderson-Windmeijer F-Stat (GINI(N))	6.38	16.00	16.71	30.42	181.33
Sanderson-Windmeijer F-Stat (REDIST)	4.35	8.71	14.61	19.60	121.81
Stock-Yogo maximal IV relativ bias $\leq$ 30%	4.08	4.08	4.00	3.88	3.92
Under-identification tests					
Sanderson-Windmeijer χ^2 p-val (GINI(N))	0.000	0.000	0.000	0.000	0.000
Sanderson-Windmeijer χ^2 p-val (REDIST)	0.000	0.000	0.000	0.000	0.000
	Panel C: Weak-instrument-robust tests				
AR-test p-val (Anderson and Rubin, 1949)					
Model	0.000	0.000	0.000	0.000	0.219
GINI(N)	0.000	0.000	0.000	0.000	0.156
REDIST	0.000	0.000	0.000	0.004	0.146
K-test p-val (Kleibergen, 2005)					
Model	0.000	0.000	0.000	0.000	0.139
GINI(N)	0.000	0.000	0.000	0.000	0.186
REDIST	0.000	0.000	0.000	0.000	0.150

	Panel D: Weak-instrument-robust intervals		
	CLR (Moreira, 2003)	Wald	Level
Column (1a) of Table (5.2) (all observations)			
GINI(N)	[-1.631; -0.265]	[-1.609; -0.105]	90%
REDIST	[-0.326; -0.057]	[-0.229; -0.011]	90%
Column (1a) of Table (5.3) (REDIST(S) sample)			
GINI(N)	[-0.419; -0.260]	[-0.464; -0.119]	90%
REDIST	[-0.931; -0.030]	[-0.473; 0.105]	90%

Notes: Table reports weak instrument diagnostics. The Sanderson-Windmeijer tests are computed as described in Sanderson and Windmeijer (2016). The F-test extends the weak instrument test for individual regressors proposed by Angrist and Pischke (2009). Benchmark values refer to Stock and Yogo (2005). AR p-val reports the p-value of the Anderson and Rubin (1949) test, while the K-test refers to the test described by Kleibergen (2005). Weak-instrument robust intervals are computed following the conditional likelihood ratio test of Moreira (2003). In order to make the calculation of the intervals computationally feasible, the intervals are constructed for the one-endogenous regressor case. To facilitate comparison, we also report the corresponding intervals of the Wald test.

Table A5-2 The change in the parameter estimate of inequality dependent upon inclusion of the transmission variables. Dependent variable is real per capita GDP growth.

	(1)	(2)	(3)	(4)	(5)
	Panel A: Contemporaneous values of inequality and redistribution				
GINI(N)	-0.181***	-0.115**	-0.111**	0.120	0.139*
	(0.0638)	(0.0483)	(0.0516)	(0.0767)	(0.0749)
REDIST(S)	-0.0946*	0.00632	-0.0864	-0.0893	-0.0348
	(0.0544)	(0.0498)	(0.0544)	(0.0573)	(0.0534)
L.log(GDP$_{pc}$)	-0.0165**	-0.0303***	-0.0168**	-0.00490	-0.0150**
	(0.00754)	(0.00581)	(0.00742)	(0.00781)	(0.00666)
SCHOOLING			0.00340*		0.00245
			(0.00176)		(0.00206)
INVS		0.204***			0.125***
		(0.0331)			(0.0311)
log(FERT)				-0.0673***	-0.0501***
				(0.0110)	(0.0104)
Observations	389	389	389	389	389
Countries	63	63	63	63	63
Hansen p-val	0.372	0.591	0.450	0.529	0.808
Diff-Hansen	0.976	0.999	0.991	0.972	0.998
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.178	0.100	0.155	0.101	0.088
Instruments	69	72	72	72	78
	Panel B: Initial values of inequality and redistribution				
GINI(N)$(t-1)$	-0.188***	-0.103**	-0.119**	-0.00338	0.0595
	(0.0585)	(0.0467)	(0.0585)	(0.0617)	(0.0614)
REDIST(S)$(t-1)$	-0.158**	-0.0157	-0.111*	-0.111**	-0.1000
	(0.0667)	(0.0691)	(0.0605)	(0.0554)	(0.0724)
Observations	334	334	334	334	334
Countries	63	63	63	63	63
Hansen p-val	0.106	0.139	0.153	0.154	0.475
Diff-Hansen	0.510	0.555	0.747	0.801	0.948
AR(1) p-val	0.000	0.000	0.000	0.000	0.000
AR(2) p-val	0.451	0.451	0.435	0.776	0.108
Instruments	62	65	65	65	71

Notes: Table reports two-step system GMM estimations with Windmeijer-corrected standard errors in parentheses. All regressions include period fixed effects. Hansen p-val gives the J-test for overidentifying restrictions. Diff-Hansen represents the Difference-in-Hansen statistic of the instrument subset for the level equation. AR(1) p-val and AR(2) p-val report the p-values of the AR(n) test. Instruments illustrates the number of instruments. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A5-3: Data sources and descriptive statistics of the variables used in the regressions

The growth rate of real per capita GDP as well as the initial level of GDP, the investment share (INVS), the degree of openness (OPEN), and government consumption (GOVC) are from PWT 9.0 as published by Feenstra et al. (2015). Average years of schooling (SCHOOLING) are collected from Barro (2013b) and include the years of primary, secondary, and tertiary education that individuals of age 15 and older have received during their educational training. POLRIGHT denotes an index of democracy and rule of law d with $d \in (1, 7)$, provided by Freedom House (2017). As the variable is coded inversely—i.e. lower numbers are associated with higher rates of democracy—we recode the variable to obtain POLRIGHT $= 8 - d$ to make sure that the coefficient in the estimation illustrates the impact of an increase in democracy, rather than the reverse. We further use fertility rates (FERT), inflation rates (INFL) and data on life expectancy at birth (LIFEEX) as reported by World Bank (2017). Table (A5-3) provides an overview of the data used in our empirical models, their means, maxima, minima, and standard deviations.

Table A5-3 Descriptive statistics of variables used in the regressions.

Variable	Explanation	N	Mean	Std. Dev.	Min	Max
$\dot{y}$	Real per capita GDP growth	1630	0.020	0.040	-0.303	0.392
log(GDP$_{pc}$)	Log of real per capita GDP	1632	8.706	1.227	5.408	12.331
GINI(N)	Gini of net incomes	1108	0.374	0.100	0.169	0.676
GINI(M)	Gini of market incomes	1052	0.452	0.068	0.222	0.681
REDIST	Redistribution	1052	0.071	0.069	-0.080	0.314
REDIST(S)	Redistribution (restricted sample)	408	0.106	0.076	-0.016	0.314
INVS	Investment	1632	0.206	0.106	0.011	1.36
SCHOOLING	Average years of school attainment	1551	5.853	3.146	0.040	13.18
GOVC	Government Spending	1632	0.200	0.103	0.019	0.912
INFL	Inflation	1560	37.001	269.004	-6.630	6962.8
OPEN	Openness	1632	0.482	0.469	0.002	7.100
POLRIGHT	Index of Political Rights	1559	4.050	2.173	1.000	7.000
Log(LIFEEX)	Log of life expectancy	1943	4.123	0.200	3.059	4.420
log(FERT)	Log fertility rate	1949	1.286	0.549	0.132	2.177
CREDIT	Credit to the private sector (% of GDP)	1479	0.378	0.375	0.009	2.951
PSE	Public spending on education	971	.045	0.020	0.008	0.264

A5-4: Usage of difference GMM and system GMM to estimate Equation (5.1)

Our preferred econometric strategy used to estimate the marginal impacts of the variables included in Equation (1) is system GMM. To give a brief intuition on its basic assumptions and properties, first rewrite Equation (1) as

$$y_{it} = \theta y_{it-1} + \lambda h_{it} + \gamma \Psi_{it} + \delta R_{it} + \beta \mathbf{X}_{it} + \eta_i + \xi_t + v_{it}. \tag{5.3}$$

This equation, in principle, can easily be estimated by OLS. However, when working with macroeconomic data, unobserved heterogeneity η_i often yields biases if not accounted accurately for. A simple way to overcome this problem would be to use a within-group estimator or a first-difference approach such as Anderson and

Hsiao (1982). However, whereas the former suffers from a Nickell (1981) bias when conducting dynamic panel estimations, first-difference transformations neglect the cross-sectional information in the data and magnify gaps in unbalanced panels. As a result, efficiency gains are possible when estimating the model in a Generalized Method of Moments (GMM) context.

A common approach to account for both unobserved heterogeneity and endogeneity in models with lagged dependent variables is the GMM estimator proposed by Arellano and Bond (1991).[88] Define that $\Delta k \equiv (k_{it} - k_{it-1})$ and $\Delta_2 k \equiv (k_{it-1} - k_{it-2})$, the basic idea of this approach is to adjust (5.3) to

$$\Delta y = \theta \Delta_2 y + \lambda \Delta h + \gamma \Delta \Psi + \delta \Delta R + \beta \Delta \mathbf{X} + \Delta \xi + \Delta v \tag{5.4}$$

and then to use sufficiently lagged values of y_{it}, h_{it}, Ψ_{it}, R_{it}, and $\mathbf{X}_{it}$ as instruments for the first-differences. However, differencing Equation (5.3) discards the information in the equation in levels. This drawback is particularly severe in the context of inequality studies, as most of the variation in inequality data stems from the cross section rather than the time-dimension. Moreover, Blundell and Bond (1998) and Bond et al. (2001) show that the difference GMM estimator can be poorly behaved if time-series are persistent or if the relative variance of the fixed effects η_i is high. The reason is that lagged levels in these cases provide only weak instruments for subsequent first-differences, resulting in a large finite sample bias.

System GMM proposed by Arellano and Bover (1995) and Blundell and Bond (1998) provides a tool to circumvent this bias if one is willing to assume a mild stationary restriction on the initial conditions of the underlying data generating process.[89] In this case, additional orthogonality conditions for the level equation in (5.3) can be exploited, using lagged values of Δk and $\Delta_2 k$ as instruments. In doing so, system GMM maintains some of the cross-sectional information in levels and exploits the information in the data more efficiently.

Satisfying the Arellano and Bover (1995) assumptions, system GMM has been shown to have better finite sample properties than difference GMM (see Blundell et al., 2000). To detect possible violations of these assumptions, we conduct Difference-in-Hansen tests to assess the validity of the additional moment restrictions for each of the system GMM regressions.[90] The system GMM estimator uses lagged variables as internal instruments for endogenous regressors. To avoid arbitrary assumptions on exogeneity, we treat all independent variables as endogenous, as is common in the literature. This, however, potentially leads to a large number of instruments. Yet, a large set of instruments possibly overfits the instrumented variables, which may fail to expunge the endogenous components and may bias parameter estimates towards those from noninstrumenting estimators. To tackle this problem of “instrument proliferation”, we use a collapsed instrument matrix. Intuitively, the collapsed matrix

88 In the case of the growth-inequality nexus, two examples are Forbes (2000) and Panizza (2002).

89 The assumption on the initial condition is $\mathrm{E}(\eta_i \Delta y_{i2}) = 0$, which holds when the process is mean stationary, i.e. $y_{i1} = \eta_i/(1-\theta) + v_i$ with $\mathrm{E}(v_i) = \mathrm{E}(v_i \eta_i) = 0$.

90 A more detailed description of the estimator in the context of the empirical application can be found in Bond et al. (2001) and Roodman (2009b).

contains one instrument for each lag distance and instrumenting variable. In addition, we follow Roodman (2009b) in combining this method with a reduced lag structure.

When presenting our empirical results, we compare the growth effects of redistribution and inequality in reduced models with those in more comprehensive specifications. This comparison is essential for the analysis of the transmission channels but evokes the problem of a large difference in the number of regressors across models. A common setting of lag structures for each model would thus either yield in invalidity of the over-identifying restrictions in the reduced models or proliferation of instruments in the comprehensive models (reflected in either very low or very large p-values of Hansen's J-test). One way to cope with this problem would be to change the lag structure across models, which would, however, impede comparison across models. For this reason, we use the second lags of the variables in levels as instruments for the difference equation and first lags of the differentiated variables for the level equation in our reduced setting. When the model is enlarged by the control variables, we collapse the instrument matrix, maintaining the lag structure of the reduced model.

In principle, our specification can be estimated using one-step or two-step GMM. Whereas one-step GMM estimators use weight matrices independent of estimated parameters, the two-step variant weights the moment conditions by a consistent estimate of their covariance matrix. The two-step procedure is asymptotically more efficient (Bond et al., 2001), but the computed standard errors may be downward biased in small samples. We therefore rely on the Windmeijer (2005) finite sample corrected estimate of the variance, which yields a more accurate inference.

As a robustness check, we use a second approach to reduce the instrument count, which is based on principal component analyses (PCA) of the utilized instruments. This technique futrher allows for exploitation of information from a larger lag number (Bai and Ng, 2010; Kapetanios and Marcellino, 2010). The PCA variant of our empirical specification replicates our baseline regressions. This approach is based on all components with eigenvalues at least 1.

A5-5: Standardization Procedure in the SWIID

Our preferred measures of income inequality and redistribution stem from the Standardized World Income Inequality Database (SWIID, Version 6.1, released in October 2017) generated by Solt (2009, 2016). The SWIID is based on the UN World Income Inequality Database (WIID), and several other cross-country inequality datasets, data provided by national statistical offices, and scholarly articles. As the source data is not directly comparable, Solt (2016) provides an algorithm to transform and adjust the original data, achieving estimates of net and market inequality comparable to those of the LIS Key Figures. A very rough overview of the standardization procedure can be given as follows: (1) The data is sorted into categories by welfare definitions and by equivalence scale. (2) Ginis of net and market inequality on the basis of household adult equivalent income from the Luxembourg Income Study (LIS) are added as a baseline, generating a dataset in which each country-year observation has data entries in at least one of thirteen categories. (3) Ratios between the variables in different categories are estimated as a function of country-decade, country, region and development status through various regression models. In further steps eleven series of estimates, comparable with the LIS net-income data, are calculated and combined into a single variable. (4) Possible measurement errors are corrected by using five-year weighted moving averages on all data points except those taken from the LIS and certain time periods. To fully reflect data uncertainty, the SWIID reports 100 different imputations for every observation, which are generated via Monte Carlo simulations.

A5-6: Redistribution in the world, REDIST and REDIST(S)
Figure (B5-1) illustrates the histogram of REDIST and REDIST(S) using 5-year averages, as in our empirical specification. When considering all country-years available in the SWIID, the mean value of redistribution is 6.56 percentage points. The standard deviation of 6.44, however, indicates that there are some major differences in the extent of redistribution across countries. The most expansive social system in the sample reduces market inequality by 26.07 percentage points, whereas some policies even yield an increase in inequality.

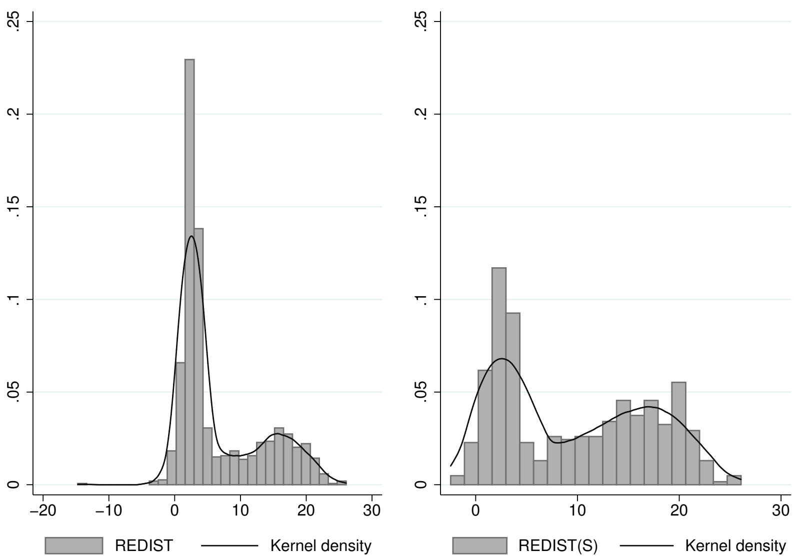


Figure B5-1 The distribution of REDIST and REDIST(S) across countries. REDIST: N=1,128, skewness=1.043, kurtosis=2.847. REDIST(S): N=453, skewness=0.268, kurtosis=1.627. Kernel is Epanechnikov.

The data also highlights that there are substantial differences in the amount of redistribution between countries at different stages of development. Using the classification of the World Bank, the mean value of redistribution in the sample of high-income countries is 12.09 percentage points and substantially exceeds the mean redistribution level in low-income countries (3.62). As REDIST(S) is composed of a larger fraction of rich economies, the picture changes slightly when considering the subsample of redistribution data that includes only the most reliable observations. The mean value increases to 9.64, but the bimodal distribution is preserved. Whereas the sample now includes a higher frequency of observations with high levels of redistribution, REDIST(S) contains less data points in which inequality is enhanced by political intervention.

Chapter 6

The Influence of Democracy on Long-Run Development[91]

Background The previous chapters described how individual investment decisions—in physical capital, innovation activity, or education—contribute to economic development. These growth mechanisms, however, can only exert their influence if individuals have the opportunity to carry out their desired investments. Put differently, in order for the previously illustrated growth drivers to operate, the political framework of the country requires a form of government that provides individual and economic freedom, freedom of contract, property rights, and legal liability. This framework is particularly likely to be realized in democratic countries. Although the definitions of democracy in political science vary, there is a broad consensus that it consists of at least four key elements: (1) A political system for choosing and replacing the administration via free and fair election, (2) the active participation of citizens in the political process, (3) the guarantee of human rights for all inhabitants, and (4) a rule of law which applies equally to all citizens (see Diamond, 2008).

The empirical literature on the effect of democracy on growth is diverse: while some studies find tentative evidence for a positive relationship between political rights and income increases, others stress that the link between these variables is either insignificant or even negative. In accordance with this ambiguity, the democracy dummy and the rule of law index included in some of the empirical investigations conducted in the previous chapters only point to a weak positive growth-effect of democracy. This chapter argues that the ambiguous effect of democracy is the result of its imprecise measurement. This chapter shows that the most severe pitfall of traditional indicators of democratization is their strategy to aggregate the underlying variables. To overcome this shortcoming and, more generally, to provide a new methodology to conduct classifications and compute composite measures in the social sciences, this chapter illustrates how machine learning techniques can be used to overcome the problem of arbitrary aggregation functions. Specifically, the underlying method uses Support Vector Machines (SVM) to compute a democracy indicator that is continuously on the $(0, 1)$ interval, enabling a very detailed and sensitive measurement of democracy for 185 countries in the period between 1981 and 2011.

[91] This chapter is based on joint work with Tommy Krieger and has been published as Gründler and Krieger (2016).

In the second step, the chapter analyzes the effect of democracy on growth based on the new measurement of democracy.

The Support Vector classification technique has generated a number of promising results in classification applications in various branches of science; however, until now little effort has been made to apply this method to economic problems. For this reason, the mathematical appendix in Chapter (A) provides a very detailed description of the theory underlying machine learning, pattern recognition algorithms, and SVMs. Support Vector algorithms can be used for estimation of real-valued functions (see Section (A.2) in the mathematical appendix) and indicator functions (see Section (A.3)). The method introduced in this chapter makes extensive use of both variants.

6.1 Introduction

Today, the belief in democracy and its positive effects on freedom, liberty, and wealth is widespread among citizens of different countries. Covering preferences of the vast majority of the world's citizens, the World Value Survey (2014) finds that 79 percent of the global population wish to live in a country that is governed democratically.[92] This preference is not only prevalent in countries with long democratic traditions (United States: 78.7 percent, Sweden: 91.9), but can also be found in Islamic states (Pakistan: 78.3, Malaysia: 86.6), African nations (Rwanda: 74.1, Zimbabwe: 86.1), South America (Chile: 83.4, Ecuador: 84.2), and Asia (China: 80.6, South Korea: 86.0). Beginning in 2011, the unfulfilled desire for democracy in the Arab World (Egypt: 93.6, Yemen: 76.3) culminated in a wave of protests, riots, and demonstrations that spread through the nations of the Arab League and the surrounding area. Driven by a fatigue with authoritarian rule, the desire for improvement of economic opportunities was one major trigger for the uprisings (see Campante and Chor, 2012).

While most of the citizens around the world seem to be quite confident that democracy brings with it an improvement in living standards, academics in the fields of political science and economics could not disagree more about the effect of democratization on economic growth. Gerring et al. (2005) summarize the academic literature by concluding that *"the net effect of democracy on growth over the last five decades is negative or null"*. More recently, some studies point to a positive effect of democracy on the income level (e.g. Acemoglu et al., 2014 and Madsen et al., 2015), whereas other surveys still find no positive contribution (e.g. Murtin and Wacziarg, 2014).

The analysis in this chapter provides evidence of a robust positive influence of democracy on economic growth. The examination implies that the ambiguity in the recent literature can first and foremost be traced back to the composition of existing democracy indicators. Available indices suffer from substantial weaknesses in conceptualization, particularly with regard to the strategy employed to aggregate

[92] See question V140 of the World Value Survey's 6th Wave, conducted between 2010 and 2014: *"How important is it for you to live in a country that is governed democratically? On this scale where 1 means it is not at all important and 10 means absolutely important what position would you choose?"* The above number refers to all respondents that respond to the question with a value of 7 or larger.

the underlying secondary data. As a result, existing indicators do not react with sufficient sensitivity to political events and regime changes.

This problem is amplified by the specification of the applied estimation techniques. A large number of recent studies eliminate unobserved heterogeneity via Within-Group estimations or difference GMM. However, while the first method yields a considerable dynamic panel bias in the context of growth regressions with lagged dependent variables (Nickell, 1981), the latter is accompanied by dramatic efficiency losses if additional orthogonality restrictions can be exploited (see Blundell and Bond, 1998). Even more severe, when estimating empirical models using transformations that remove the information in the equation in levels, it is particularly necessary to have democracy indicators that react very sensitively to political events and regime changes. Otherwise, relying on the limited within-country information in the panel is likely to yield ambiguous results concerning the growth effect of democratization.

This chapter addresses both challenges. First, a novel approach to measure democracy is introduced which is based on machine learning algorithms for pattern recognition. These algorithms operate on the basis of a model drawn from example inputs which is used to make data-driven predictions or decisions. The particular advantage gained via application of such methods is that they give computers the ability to learn without being explicitly programmed. Whereas the machine learning toolbox provides numerous promising instruments, Support Vector Machines (SVM) in particular have recently produced striking results in various branches of science. Practical applications include categorization of cancer cells (Guyon et al., 2002), classification of hyperspectral data in geophysics (Gualtieri, 2009) and identification of biomarkers of neurological and psychiatric disease (Orrú et al., 2012). The idea behind these approaches is transferred to the problem of democracy measurement in order to obtain an index which may be referred to as the *Support Vector Machines Democracy Indicator* (SVMDI). The indicator is continuous on the interval from 0 to 1, thereby considerably enhancing the level of detail. The most important improvement, however, is that the aggregation of the characterizing variables is not arbitrary, as the SVM algorithm puts the problem of learning—i.e. the classification of country-years—into the context of an nonlinear optimization problem. The SVMDI is available for 185 countries in the period from 1981 to 2011, covering countries representative of over 99 percent of the global population.

In the second step, the effect of the SVMDI on economic growth is analyzed in a system GMM framework which considers the econometric challenges described above. The findings indicate a robust positive relationship between the SVMDI measure and economic growth. This result remains stable when changing the estimation technique to some of the recently applied strategies found in the literature. In particular, accounting for waves of democratization via instrumental variable regressions using regional and cultural democratization trends as external instruments strongly supports the baseline outcomes.

To demonstrate the superiority of the SVM classification procedure, this chapter also provides an extensive comparative analysis of the results obtained by SVMDI and alternative democracy indicators. Given the inability of hitherto existing democracy indicators to react with sufficient sensitivity to political developments, the SVMDI is the only indicator that suggests a positive effect on growth in models that rely on the within variation of countries. This implies that even small steps in the transition process towards democracy are important for increases in living standards. However, when using the system GMM framework of the baseline estimations, the positive association between democracy and growth emerges as a clear empirical pattern, even when relying on rough measures of democratization.

Finally, the chapter investigates the transmission channels through which democracy triggers income increases. It turns out that democracy exerts its influence via better education, higher investment shares, and lower fertility rates. In contrast, there is little evidence of a redistribution-enhancing effect, which may explain why the analysis does not detect nonlinear effects of democracy in comprehensive model specifications.

The chapter is organized as follows. Section (6.2) discusses the ambiguity in terms of the effect of democracy on growth in recent studies. Section (6.3) critically analyzes the most commonly used traditional democracy indicators. Section (6.4) briefly introduces the theory underlying machine learning, which is derived in detail in Sections (A.2) and (A.3) in the mathematical appendix. In the next step, the section describes the application of these methods in constructing the SVMDI algorithm. This Section further provides an overview of the democracy level and its historical trends in the world and compares the SVMDI to alternative indicators. Section (6.5) is concerned with the estimation strategy and the presentation of the empirical results. In Section (6.6), the transmission channels of democracy are examined. Section (6.7) concludes.

6.2 The ambiguous effect of democracy in recent studies

The effect of democracy on growth is strongly ambiguous in recent studies, both theoretically and empirically. On the theoretical side, it has been argued that democratization may benefit growth, most importantly via better provision of public goods and education (Saint-Paul and Verdier, 1993, Benabou, 1996, and Lizzeri and Persico, 2004) or by imposing constraints on kleptocratic dictators and preventing political groups from monopolizing lucrative economic opportunities (Acemoglu et al., 2008 and Acemoglu and Robinson, 2012). In addition, Alesina et al. (1996) emphasize that increased political stability enhances national and foreign investment. Feng (1997) illustrates that democracy reduces the likelihood of regime changes, which indirectly benefits growth. However, a large body of literature emphasizes the possible negative effects of democratization, mainly as a result of a higher level of redistribution, which is assumed to reduce growth (see, for instance, Alesina and Rodrik, 1994 and Persson and Tabellini, 1994). In addition, Olson (1982) argues that sufficient organization of interest groups can lead to stagnation in democracies.

Empirically, cross-sectional analyses conducted by Barro (1996) and Tavares and Wacziarg (2001) suggest a (slightly) negative effect of democracy on growth. The investigation of Barro (1996) also provides evidence for a nonlinear relationship between the variables, where an increase in political rights at low levels of democratization benefits growth, but triggers a negative effect if a critical threshold of democratization is exceeded. Barro (2003) confirms the nonlinear effect using panel data, where other panel data analyses yield quite ambiguous results. Rodrik and Wacziarg (2005) find no significant effect of democratic transition on growth in the long-run, but emphasize short-run benefits and a decline in economic volatility. Likewise, Apolte (2011) reports ambiguous effects of democracy on prosperity in transition countries, tentatively arguing that basic constitutional rights and constraints on the government may be conducive to growth. Burkhart and Lewis-Beck (1994), Giavazzi and Tabellini (2005) and Murtin and Wacziarg (2014) also find no robust indication of a positive relationship running from democracy to growth. Using semi-parametric methods, Persson and Tabellini (2008) report an average negative effect of departure from democracy on growth. Persson and Tabellini (2009) analyze the effect of democratic capital, measured by a nation's historical experience with democracy and by the incidence of democracy in its neighborhood. Whereas the results imply that democratic capital stimulates growth, Acemoglu et al. (2014) argue that the formidable challenge in this case is the difficulty of disentangling the impact of unobserved heterogeneity from the effect of democratic capital. Gerring et al. (2005) apply a similar approach, concluding that democratization facilitates income increases. Providing a dichotomous index of democracy, Acemoglu et al. (2014) find that the degree of democracy is positively correlated with future GDP per capita. The authors also use regional waves of democratization in an IV approach to account for possible problems caused by endogeneity. A similar approach is conducted by Madsen et al. (2015), who use the strength of democracy in linguistically comparable countries as an external instrument. Both approaches find a positive link between democracy and the level of incomes.

A different branch of literature is concerned with the reverse effect, i.e. the causal relationship of economic growth to democracy. This literature goes back to Lipset (1959), who finds a strong and positive correlation between the level of income per capita and the likelihood of transition to democracy. Recent surveys, however, provide ambiguous results. While Acemoglu et al. (2008, 2009) suggest that growth does not contribute to the process of democratization, Murtin and Wacziarg (2014) endorse Lipset's modernisation theory.

6.3 Recent democracy indicators

The traditional way to create a democracy indicator follows three steps. First, it is necessary to choose a definition of democracy. Second, a number of instruments must be designed that are able to describe the properties of the theoretical concept. Finally, a suitable manner for combining the selected variables must be found for computation of the democracy index.

In practical applications, however, a large number of problems arise in each of these steps. The first issue concerns the nature of democracy. With no generally accepted definition at hand, the interpretations range from minimal approaches primarily focusing on the election process (see, e.g., Dahl, 1971) to concepts that additionally incorporate human rights and social inequality (see, e.g., Rawls, 1971). As a result of this variety, the indicators deviate considerably in their underlying instruments. For instance, the popular index of Vanhanen (2000) only utilizes two dimensions—participation and competitiveness in elections—to characterize a democracy. The advantage of such a minimal concept is that data can be collected for a large number of countries and years, yielding a democracy indicator that covers a broad sample of observations. However, researchers employing democracy data need to acknowledge the cost-benefit trade-off and must ensure that any substantial analytical conclusion drawn in their investigation is consistent with the underlying data concept. In the case of the Vanhanen-index, the allure of large data coverage comes at a high cost. First, instrumentation of participation and competition via (respectively) voter turnout and the percentage of votes going to the largest party constitute, at best, poor measures of the corresponding attribute (for a detailed discussion, see Munck and Verkuilen, 2002). Second, the aggregate index is obtained by simply multiplying the two attributes, where Vanhanen (2000) does not offer any theoretical justification for the arbitrary assumption that equal weight ought to be assigned to the attributes.

A similar minimal concept is used in the index of Boix et al. (2013) which defines a country-year as *democratic* if it meets three conditions in terms of contestation and participation.[93] The obvious drawback of this approach, one inherent to each dichotomous indicator of democracy $d_{\{0,1\}}$, is the lack of detail. In particular, the implicit assumption in empirical cross-country analyses is that each country with $d_{\{0,1\}} = 1$ is equally weighted in the computation of estimates. With regard to the Boix et al. (2013) measure for the year 2010, this implies classifying Pakistan, Bangladesh, Mali, Liberia, Sierra Leone, Zambia, and Lesotho as having the same extent of democratization as the United States, Germany, Canada, and the United Kingdom. In addition, the data sources underlying the classification of countries change over time (see Boix et al., 2013), yielding inconsistency in the indicator across time periods.

Two measures of democracy have achieved a particularly high degree of popularity. These are the Polity IV score provided by Marshall et al. (2014) and the rating compiled by Freedom House (2014). What is common to both approaches is that they are neither dichotomous, nor continuous.[94] For Polity IV and Freedom House the range of possible values runs from -10 to 10 and from 1 to 7, respectively. Although they differ in their purpose, both indices are quite similar in construction, building on the evaluation of country experts who classify nations along a set of predefined criteria. In both cases, however, the aggregation strategy is fraught with problems. The Freedom House (2014) index aggregates scores for two attributes—political rights and civil liberty—by simply adding up the values of their respective underlying

[93] These conditions are: (1) The executive is elected in popular elections and is responsible to voters, (2) the legislature or the executive are elected in free and fair elections, and (3) the majority of adult men have the right to vote.

[94] As Cheibub et al. (2010) emphasize, due to the discrepancies in their components, both Freedom House and Polity IV cannot be interpreted as cardinal measures or ordinal rankings. In fact, the measures are categorial, whereby the categories are not precise.

components. With regard to each of the two attributes, there is a bewilderingly long list of components that are all added with equal weight without any theoretical justification of the aggregation strategy. Meanwhile, with regard to the content of the underlying components, equal weighting seems particularly inadequate in this case.[95] The failure to provide a reasonable aggregation rule is compounded by a number of conceptual and measurement problems that are discussed in detail in Munck and Verkuilen (2002) and Cheibub et al. (2010). Arbitrariness of the aggregation rule is also a fundamental deficiency of the Policy IV score (for a detailed discussion, see Treier and Jackman, 2008).

More recently, some scholars have attempted to achieve more reliable measures by synthesizing existing democracy indicators. For instance, Acemoglu et al. (2014) propose an approach based on four established indices to obtain a dichotomous indicator. According to the applied heuristic, a country-year is classified as *democratic* ($d_{\{0,1\}} = 1$) if the rating of Freedom House (2014) is *free* or *partly free* and the Polity IV score provided by Marshall et al. (2014) is greater than zero. To address the issue that for certain observations only one of the underlying indicators is available, Acemoglu et al. (2014) use two additional indices (Boix et al., 2013 and Cheibub et al., 2010) to classify the country-years in question.[96] As in the case of the Boix et al. (2013) measure, the main drawback of this method is that it enables only a binary classification of democracy, which does not allow for a nuanced distinction between different countries. For instance, when referring to the observation in 2012, the measure of Acemoglu et al. (2014) implies that the young and fragile Tunisian democracy has the same quality as the established democracies of Canada and the United States. Employing this measure in cross-country analyses would imply assigning Tunisia and Canada the same value, which impedes sound interpretation of statistical estimations intended to evaluate the effect of democracy. Furthermore, a dichotomous indicator contradicts the broad consensus that cultivation of a democracy is a process which occurs over a longer period of time. Treating each country-year as equally (non)democratic neglects information about the process of democratization and results in a severe upwards bias in empirical estimations (Doucouliagos and Ulubaşoğlu, 2008).

Pemstein et al. (2010) propose another, more technical method to combine established indices. The basic idea underlying this concept is to synthesize ten available democracy indicators via a Bayesian latent variable approach to obtain the *Unified Democracy Score* (UDS). A formidable challenge presented by the inclusion of such a large number of indicators is that of dealing with the fact that the indicators differ substantially in the number of evaluated countries and periods. For instance, the Polity IV score is available continuously for the time-period from 1945 to present, while other indices are available only for very few periods. Nevertheless, the approach

[95] For instance, it is likely erroneous to consider the decentralization of power to be as important for democracy as the actual power exercised by elected representatives (Munck and Verkuilen, 2002).

[96] As the Polity IV index only evaluates countries with at least 500,000 inhabitants, such conflicts particularly arise when considering mini-states. In addition, the data of Freedom House (2014) only reaches back to 1973, while the measure of Acemoglu et al. (2014) includes the period between 1960 and 2010.

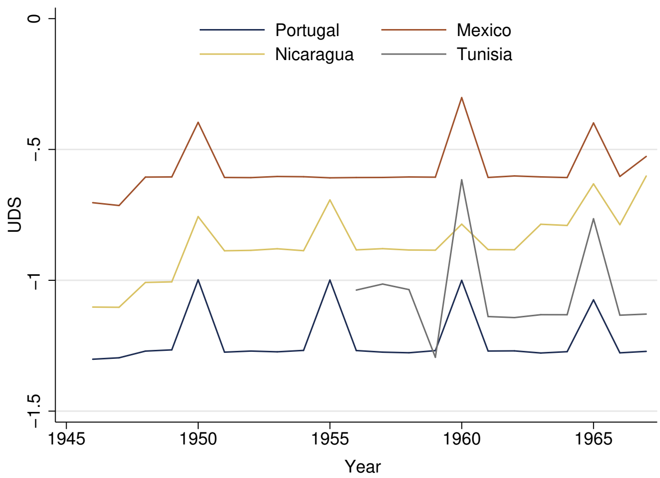


Figure 6.1 Inconsistency in the democracy indicator of Pemstein et al. (2010). Unified Democracy Scores (UDS) of Mexico, Nicaragua, Portugal, and Tunisia.

of Pemstein et al. (2010) includes all available information for each country-year, whereby the number of included secondary indicators varies from observation to observation. This, however, yields severe inconsistency in the UDS over time.[97] In fact, a large number of the included national series imply relatively constant democracy scores over time, only to be interrupted by a peak occurring almost every five years when analyzing the time period between 1950 and 1980. This peak is due to the index of Bollen (2001), which is only included in the UDS in the years 1950, 1955, 1960, 1965, and 1980. The resulting inconsistency is illustrated in Figure (6.1).

A very similar bias that affects the UDS of a considerable number of countries occurs in the early 1970s, the time period when the Freedom House (2014) ratings were initially published. Finally, Gugiu and Centellas (2013) emphasize that the UDS can hardly discern between countries that are not on opposite ends of the democracy spectrum, as the reported confidence intervals overlap.

The drawbacks discussed above may stand exemplary for the majority of the existing democracy indicators. There is an extensive literature discussing the advantages and disadvantages of the democracy indicators at hand (e.g. Munck and Verkuilen, 2002, Cheibub et al., 2010, Gugiu and Centellas, 2013), in which the consensus has been reached that existing indices suffer from a variety of conceptualization issues. Points of criticism include the low level of detail, the utilization of unfounded scaling, the disproportionate influence of expert knowledge, subjectivity and arbitrariness in

[97] Although for some country years the UDS was produced by drawing on information from ten democracy indicators, the majority of observations rely on an average of six underlying indicators which deviate in their composition for different country-years. This restricts comparison of UDS scores across countries and over time.

the conceptualization, the selection of the instruments, and the theoretical concept of democracy. Above all, however, the main concern is the fairly low level of sophistication with regard to the aggregation process and the way in which the underlying components are weighted.

6.4 Measuring democracy using Support Vector Machines

6.4.1 Motivation

Compared to other macroeconomic series—such as, for instance, the inflation rate, the unemployment rate, or the growth rate—the quantification of democracy is considerably more challenging, since there is neither a commonly accepted definition of democracy, nor a natural unit or scale by which it can be measured. The literature at hand has, however, arrived at the consensus that it is preferable to measure the degree of democratization rather than quantifying the stock of democracy, where the usage of scales with *a priori* chosen lower and upper bounds as benchmarks for the lowest (*fully autocratic*) and highest (*fully democratic*) possible degree is common. Mindful of this preference, traditional democracy indicators attempt to determine a number of requirements which a country has to fulfill to reach a certain degree of democratization, as opposed to trying to observe democracy *directly*.[98] More formally, the degree of democratization $d_{i,t} \in \mathcal{D} \subseteq \mathbb{R}$ of country i in period t can be expressed as a function $\mathfrak{F}\colon \mathcal{X} \subseteq \mathbb{R}^m \to \mathcal{D} \subseteq \mathbb{R}$ of the extent to which the country-year meets the selected conditions. Subsequently, these conditions are denoted with $\mathbf{x}_{i,t} = \left(x_{i,t}^1, ..., x_{i,t}^m\right)' \in \mathcal{X} \subseteq \mathbb{R}^m$, where m is the number of requirements, i.e.

$$d_{i,t} = \mathfrak{F}\left(x_{i,t}^1, ..., x_{i,t}^m\right) \quad \forall (i,t). \tag{6.1}$$

A basic property of the frequently used scales is that its range of values can be normalized to the $(0,1)$ interval without a loss of essential information.[99] Hence, the analysis subsequently focuses on the case where the output space is normalized, i.e. $\mathcal{D} = [0,1]$. This yields the advantage that each absolute change of the indicator can directly be interpreted as the change in the degree of democratization.

As emphasized in Section (6.3), the low degree of sophistication with respect to the aggregation function $\mathfrak{F}(\cdot)$ is undoubtedly a substantial methodological weak spot of existing democracy indicators. Hence, finding a suitable strategy to detect the unknown function $\mathfrak{F}(\cdot)$ without arbitrary assumptions is an essential step to improve the quality of democracy indicators. By using Support Vector Machines, the procedure introduced in this chapter transfers the problem of aggregation into a nonlinear optimization context, estimating the most appropriate function $\mathfrak{F}(\cdot)$. In fact, machine learning algorithms and Support Vector Machines are explicitly designed for

[98] In this context, democracy is frequently interpreted as a latent variable (see, e.g., Pemstein et al., 2010).

[99] For instance, the Polity IV Index originally ranges from -10 to 10 (Marshall et al., 2014). It is possible to obtain a normalized score of each country-year P_{it} with the same information via computation of $\frac{P_{it}+10}{20}$.

problems where the functional form is unknown and where researchers do not have any reasonable description of the functional relationship between the inputs and the desired response (see Steinwart and Christmann, 2008). To construct a statistical learning machine with Support Vectors (SV), two essential requirements must be met. First, the algorithm needs a set of input characteristics that are available for all observations in the sample. Second, there is need for a limited number of observations with known output, based on which the algorithm can learn (see Steinwart and Christmann, 2008).

Intuitively, the approach documented in this chapter first identifies country-years that can be indisputably categorized as *highly democratic* or *highly autocratic* and uses them as observations with known output. Based on these *a priori* labeled observations, an aggregation function $\mathfrak{F}(\cdot)$ is computed which uses eleven attributes to conduct a classification of the degree of democracy. These attributes include different aspects of political participation, political competition, and civil rights. Finally, the procedure yields a continuous measurement of democracy, which can be referred to as the *Support Vector Machines Democracy Indicator (SVMDI)*. This indicator can be interpreted as the degree of democratization based on a continuous scale reaching from 0 to 1.

6.4.2 Machine learning and Support Vector Machines

The field of machine learning studies algorithms that operate on the basis of a model drawn from example inputs that is then used to make data-driven predictions or decisions (see, e.g., Bishop, 2006). The enormous advantage gained through application of such methods is that of providing computers with the ability to learn without being explicitly programmed (Samuels, 1959). Largely developed at AT&T Bell Laboratories, the Support Vector Machines (SVM) algorithm as a subfield of machine learning was designed to have a firm orientation towards real-world application. Hence, utilization of SVM has achieved very promising results in various branches of sciences. Practical applications include categorization of cancer cells (Guyon et al., 2002), classification of hyperspectral data in geophysics (Gualtieri, 2009), and identification of biomarkers of neurological and psychiatric disease (Orrú et al., 2012). In addition, the algorithm has been used to categorize texts (Joachims, 2002) and to analyze hand written characters (Cortes and Vapnik, 1995).[100]

The machine learning toolbox consists of a wide range of different algorithms. In this application, two common methods of SV regression and SV classification are utilized. While the regression tool is essential to obtain the desired function $\mathfrak{f}(\mathbf{x}_i)$, SV classification is used to conduct validity tests of the selections. In the following, this section provides a brief introduction on how to use Support Vector Machines for regressions, while the concept of SV classification is closely related. It bears underscoring, however, that the mathematical literature on machine learning has

[100] Only little effort has thus far been made to apply the SVM algorithm in the field of economics; up until now its application has been restricted to financial topics and stock markets. For instance, Kim (2003) and Tay and Cao (2001) use SVM for financial time-series forecasting and Shin et al. (2005) apply the method in a bankruptcy prediction model.

developed considerably over time, which is why the following description concentrates primarily on its basic ideas. To provide a broader introduction in the field of machine learning, pattern recognition, and Support Vectors, the mathematical Appendix (A) illustrates these methods in greater detail. In particular, Chapter (A.2) is concerned with utility of Support Vector Machines for regressions, while Chapter (A.3) elucidates on application of SVM for classification purposes. For readers with a broader interested in the mathematical and computational issues of SVM, the inspiring work of Vapnik (1998), Smola and Schölkopf (2004), and Steinwart and Christmann (2008) is recommended.

To provide a brief introduction, the problem to be solved by the SV regression tool can be described as follows: Given a certain data set $\mathcal{F} = \{(\mathbf{x}_1, y_1); ...; (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i \subset \mathcal{X} \in \mathbb{R}^m$ and $y_i \in \mathbb{R}$, the aim is to find a function $\mathfrak{f}\colon \mathcal{X} \subset \mathbb{R}^m \to \mathbb{R}$ with the property

$$\mathfrak{f}(\mathbf{x}_i) = y_i \quad \forall\, i = 1, ..., n. \tag{6.2}$$

However, due to measurement errors and unobserved characteristics, achieving a perfect fit is generally not feasible. For this reason, the aim of SV regression is to compute a function $\hat{\mathfrak{f}}\colon \mathcal{X} \subset \mathbb{R}^m \to \mathbb{R}$ which approximates the "true" function $\mathfrak{f}\colon \mathcal{X} \subset \mathbb{R}^m \to \mathbb{R}$ such that

1. the deviation between $\hat{\mathfrak{f}}(\mathbf{x}_i)$ and y_i does not exceed a given level ε for each observation i, and (simultaneously)
2. the shape of $\hat{\mathfrak{f}}(\cdot)$ is as flat as possible (Smola and Schölkopf, 2004).

Largely influenced by the Generalized Portrait algorithm (Vapnik and Lerner, 1963, Vapnik and Chervonenkis, 1964), the basic idea of SV regressions is to find a hyperplane in $\mathcal{X}$ that satisfies these two requirements. However, the functional flexibility of hyperplanes typically limits the possibility of obtaining precise approximations for all observations in the sample. As a result, the first condition is violated in most cases. To resolve this issue, Boser et al. (1992) suggest using a higher dimensional space $\mathcal{H}$ instead of $\mathcal{X}$—called *feature space*—where shifting of the data is accomplished via a nonlinear *feature map* $\Phi(\cdot) : \mathcal{X} \to \mathcal{H}$ that is chosen a priori.

This procedure, however, gives rise to the question of how to treat the high-dimensional space $\mathcal{H}$, since an appropriate map $\Phi(\cdot)$ is typically unknown. In addition, this approach can easily become computationally infeasible with respect to polynomial features of higher order or higher dimensionalilty. Boser et al. (1992) propose a method to overcome this problem, which has become known as *the kernel trick*, largely building on the idea initially introduced by Aizerman et al. (1964b). The approach circumvents direct construction of the hyperplane based on the data in $\mathcal{H}$ and relies instead on the dot products of the Support Vectors (Vapnik, 1998). This method is feasible if there exists an admissible SV *kernel* $k(\mathbf{x}, \mathbf{x}_i)$ that satisfies a certain number of conditions.[101] In the application of this chapter, the algorithm

[101] See, in particular, the Theorem of Mercer (1909) and the Theorems of Schoenberg (1942) and Burges (1999). For a detailed overview and discussion, see Smola and Schölkopf (2004) and Section (A.2.5) in the mathematical appendix.

uses the Gaussian Radial Basis Function (RBF) as a kernel, with the result that the corresponding feature space $\mathcal{H}$ becomes a Hilbert space of infinite dimension.

In this way, the optimal SV regression function can be calculated via

$$\hat{\mathfrak{f}}(\mathbf{x}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\ k(\mathbf{x}, \mathbf{x}_i) + b, \tag{6.3}$$

where b denotes the intercept, and the Lagrange multipliers $\alpha = (\alpha_1, \ldots, \alpha_n)'$ and $\alpha^* = (\alpha_1^*, \ldots, \alpha_n^*)'$ are computed by solving the optimization problem (Smola and Schölkopf, 2004)

$$\max_{\alpha,\alpha^*} -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(x_i, x_j) - \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0 \quad \text{and} \quad \alpha_i, \alpha_i^* \in [0, C],$$

with given cost parameter C and fixed margin ε.

6.4.3 The SVMDI algorithm

This section is concerned with the transfer of the general SV regression approach to the problem of quantifying the degree of democratization. In the first step, a set of input attributes ($\mathbf{x}$) must be specified that captures the character of democracies and that are available for each observation in the sample. This selection is based on a broad concept of democracy, i.e. the analysis does not exclusively focus on the core elements of *political participation* and *political competition* (as, for instance, proposed by Dahl, 1971), but also includes *civil liberty* and *independence of non-government institutions*, such as the judiciary and the press. In this sense, the approach follows a large body of theoretical literature which argues that democracy requires more than just a free general election process (see, e.g., Rawls, 1971 and Diamond et al., 1990).

In his seminal work, Vanhanen (2000) suggests quantifying the degree of political competition based on the share of power concentrated in the largest political party in the last general election. How to measure this share is, however, ambiguous. In fact, it seems plausible to either use the share obtained at the ballot box or to rely on the share of seats in the parliament. Both shares may be relatively similar in the majority of cases, but differ considerably with respect to some country-years. For instance, the 2002 Turkish general election saw 34.28 percent of all valid votes go to the *Justice and Development Party* (AKP) chaired by Recep Tayyip Erdogan. This, however, led to their acquiring 66 percent of all seats in parliament (Carr, 2014). Analyzing the bibliography of Vanhanen (2000) unfortunately brings to light severe data inconsistencies in the competition dimension over time and across countries, which provides a strong argument to refrain from utilization of this database. Instead, the necessary information is collected from a number of other primary sources that distinguish between both shares, most notably Nohlen et al. (1999, 2001) and Nohlen (2005, 2010). To obtain additional information concerning the degree of political competition, the analysis further rests on the ratio of votes and the ratio of parliamentary seats between the strongest and second strongest parties. In total, four variables to characterize the competition dimension of democracy are incorporated.[102]

The second attribute is *political participation*, which is included based on voter turnout (Vanhanen, 2000), the rating of political freedom provided by Freedom House (2014), and an indicator of political oppression and violence computed by Gibney et al. (2013).[103] In addition, *independence of judiciary* is reflected by the INJUD series in the database of Cingranelli et al. (2014), while the *freedom of press* indicator is obtained from Freedom House (2014). Finally, the quality of *civil liberties* is evaluated by two expert-based ratings provided by Freedom House (2014) and Cingranelli et al. (2014). The attribute obtained from the Cingranelli et al. (2014) compilation is based on the mean value of five scores regarding essential human rights.

In light of the critique traditional democracy indicators are confronted with (see Section (6.3)), selection of these variables is conducted cautiously with respect to the crucial issue of data quality. Whenever feasible, the analysis avoids inclusion of aggregated data by drawing on the original series. In addition, by incorporation of various series from different sources, their individual caveats are counterbalanced, at least to some extent. Yet it bears underscoring that the accuracy of the SVMDI rests on the attributes that are used as input variables.

In the second step, a subset of country-years $\mathcal{L} \subset \mathcal{F}$ is selected that consists of elements that can unambiguously be categorized as either highly *democratic* or highly *autocratic*. This selection lays the foundation for the SV algorithm, so that preliminary degrees of democratization must be assigned to both groups of labeled observations. At this point, the specification follows the seminal work of Ragin (2000, 2008) who suggests 0.05 (0.95) for *highly autocratic* (*highly democratic*) observations as appropriate benchmarks.[104] In order to compile $\mathcal{L}$, the algorithm follows Acemoglu et al. (2014) by using the Polity IV score (Marshall et al., 2014) as decision criterion. A country-year is labeled as democratic only if the Polity IV index assumes its highest

[102] The secondary datasets used include African Election Database (2014), Carr (2014), IPU (2014), IDEA (2014), Nohlen et al. (1999, 2001), Nohlen (2005, 2010), and World Bank (2014a).

[103] By using different sources and variables, the algorithm tries to counterbalance the methodological shortcomings of the input measurements, conceding that this strategy does not remove all issues. However, no better data is available for the large number of observations included in the sample.

[104] Undoubtedly, the setting of the thresholds is to some extent arbitrary, which is why the algorithm is conducted with several varying thresholds, ranging from 0 to 0.1 (0.9 to 1). The results turn out to be relatively unaffected by the particular choice. In each case, the algorithm is able to detect substantial differences in the degree of democratization in both sub-samples that have received a preliminary label. This issue is illustrated in Section (6.4.5) based on the example of Mongolia. Note also that the empirical results in terms of the democracy-growth nexus explored in this chapter are not sensitive to different thresholds.

possible value of 10. At the other end of the spectrum, countries are classified as *autocratic* if the Polity IV indicator is -7 or below.[105]

Subsequently, a random generator selects t_{demo} and t_{auto} elements of $\mathcal{L}$ and consolidates them into the set $\mathcal{T}_\zeta$. To avoid arbitrary assumptions, both parameters are chosen by a uniformly distributed random number generator. The algorithm proceeds (step 4) by conducting SV regression based on the observations in $\mathcal{T}_\zeta$, yielding a nonlinear function $\mathfrak{F}_{\mathcal{T}_\zeta} : \mathcal{X} \subset \mathbb{R}^{11} \to [0, 1]$.[106] For computation of $\mathfrak{F}_{\mathcal{T}_\zeta}(\cdot)$, the Gaussian Radial Basis Function (RBF) kernel is used, which possesses a number of advantageous properties (for a detailed illustration, see Section (A.2.9) in the mathematical appendix). In addition, the RBF has provided the most promising results in the robustness check. In the fifth step, the estimated aggregation function $\mathfrak{F}_{\mathcal{T}_\zeta}(\cdot)$ is used to assign a degree of democratization $d_{it} \in [0, 1]$ to all country-years included in the sample $\mathcal{F}$.[107] To prevent a potential selection bias, $\zeta = 1, \dots, 2,000$ iterations of the process from step 3 to 5 are computed. This bootstrapping procedure ensures numerical robustness with respect to our parameter selection, accounts for potential measurement errors in the underlying data, and enables the estimation of confidence intervals.[108] The *Support Vector Machines Democracy Index* (SVMDI) is the average value over the 2,000 iterations for each country-year, yielding a continuous measurement of democracy that ranges from 0 to 1.

For a given country-year $\langle i, t \rangle$, the SVMDI indicates the degree of democratization with respect to the liberal concept of democracy and the benchmark country-years included in the classification. Dependent on availability of the underlying data, the SVMDI is computable for 185 countries in the period from 1981 to 2011. To account for a potential bias due to inexact quantification, potential measurement errors in the underlying data, and omitted variables, the algorithm also computes confidence intervals for the SVMDI point estimates. The lower (upper) bound of these intervals corresponds with the 5th (95th) percentile of the simulated distribution of the point estimate that is compiled for each country-year based on the 2,000 iterations.

---

[105] Since the selection of the Polity IV index as a criterion seems arbitrary, the robustness of the classification is checked by using several other criteria based on other democracy indicators. These changes yield little differences in the resulting indicator, which is hardly surprising given the high concordance of established democracy indicators with respect to the top and the bottom of the global democracy distribution. As an additional internal validity check, the outcome of the algorithm is compared to SV classifications based on the input variables to examine if there are differences in the initial labels compared to those obtained by usage of the Polity IV indicator. This analysis reveals very little indication for any mislabeled country-years.

[106] To ensure that the estimated function can reach all values between 0 and 1, the margin parameter is set to $\varepsilon = 0.05$.

[107] From a theoretical perspective, it is possible that the predicted values are above the upper bound 1 or below the lower bound 0. To avoid such cases, an additional restriction in the implementation is included that ensures the scores to range between 0 and 1.

[108] The combination of bootstrapping and SVM is frequently used in the literature (see, e.g., Alonso-Atienza et al., 2012, Jain et al., 2014 and Wang and Ma, 2012).

6.4.4 Democracy in the world

How democratic are the countries in the world? Figure (6.2) maps the SVMDI derived in the previous section in the post-2010 period. This presents a very heterogeneous picture: while countries in Europe, Oceania, North America, and—to a large extent—in South America possess high SVMDI scores, a substantial part of the nations in Africa and Asia are considerably less democratic.

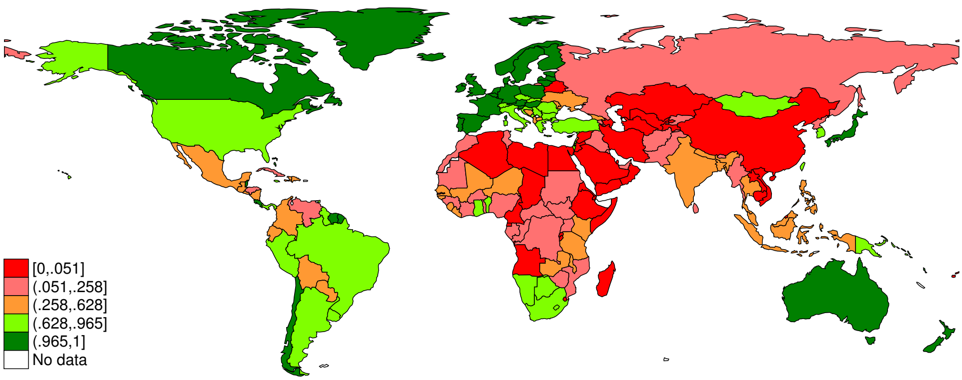


Figure 6.2 Democracy in the world (SVMDI), post-2010 period. Classes refer to the quintiles of the distribution of the SVMDI indicator.

An interesting pattern revealed by Figure (6.2) is that the degree of democratization shows clear tendencies towards regional concentration. If a country is (non-)democratic, there is a high probability that the same applies to its neighboring countries as well. There are three remarkable exceptions to that general rule: surrounded by countries with very low SVMDI scores, Mongolia (SVMDI: 0.8068), Ghana (0.9302), and—to a lesser extent—Benin (0.6413) succeeded in establishing democratic structures. Overall, the figure suggests that the extent of democratization is strongly polarized.

This polarization becomes particularly apparent when considering the distribution of the SVMDI measure, illustrated in Figure (6.3). The data suggests a bimodal distribution, where the first mode is located at a very low level of democracy, and the second mode lies at a substantially higher degree of democracy. This pattern is typical when examining the degree of democracy across countries and occurs in a similar manner when analyzing alternate measures. The reason is that there exist a substantial number of countries with an SVMDI index close to zero. These countries include nations where civil war is prevalent—e.g. Syria (0.0337), Afghanistan (0.0934), and Sudan (0.0601)—and countries with absolute monarchies, such as Swaziland (0.0069), Qatar (0.0305), and Brunei (0.0259). On the other hand, there are numerous countries where strong democratic institutions have been established, particularly in Europe, North America, Oceania and in some parts of Latin America. Figure (6.3) also demonstrates that the trend towards democratization emerges as a clear empirical pattern in the SVMDI data. Whereas the relative fraction

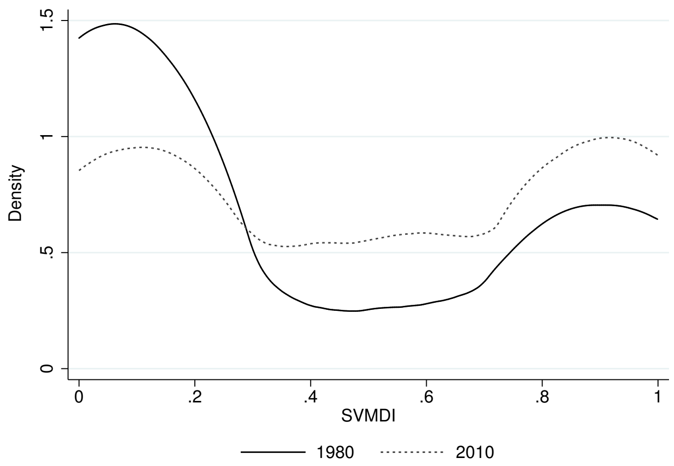


Figure 6.3 Democracy in the World (SVMDI), kernel density estimates 1980–2010. Kernel is Epanechnikov.

of non-democratic nations was extraordinarily high in the 1980–1984 period, the data approximates a more uniform distribution in the post-2010 period, which is characterized by a substantially higher number of democratic countries and a lower number of nations with poor SVMDI scores.

Figure (6.4) exemplary plots the SVMDI scores and the confidence intervals for Serbia, South Korea, Venezuela, and Argentina over the entire period from 1981 to 2011.[109] The figure highlights the considerable progress in democratization during the 1980s and the early 1990s, which later became known as “Democracy’s Third Wave” (see, for instance, Huntington, 1991, 2012). Beginning in Latin America in the early 1980s, the Third Wave washed over to Asia Pacific countries and reached its crest in Eastern Europe after the collapse of the Soviet Union. This development is clearly visible in the SVMDI data. Particularly noteworthy is the substantial progress achieved in South Korea and Argentina, both of which were classified as highly autocratic in the early 1980s. Similar movements towards democracy can be observed in Eastern Europe after the fall of the Iron Curtain in 1989.

The Serbian path to democracy, however, was more tortuous than those of its Baltic and East-Central European neighbors. Only following the resolution of the armed conflicts in Bosnia and Herzegovina (1992–1995) and Kosovo (1998–99) was an increase in political rights and democratization initiated (see McFaul, 2002). Meanwhile, democracy has not yet been cultivated in full, a fact which is clearly reflected in the SVMDI of the country (Greenberg, 2014).

A further issue that has gained increasing attention is the fear of a potential “reverse” wave occurring in Latin America due to the importance of autocracy and

[109] Note that the Serbian SVMDI is composed of the scores of SFR Yugoslavia (1981–91), FR Yugoslavia (92–02), Serbia and Montenegro (03–05), and Serbia (06–11).

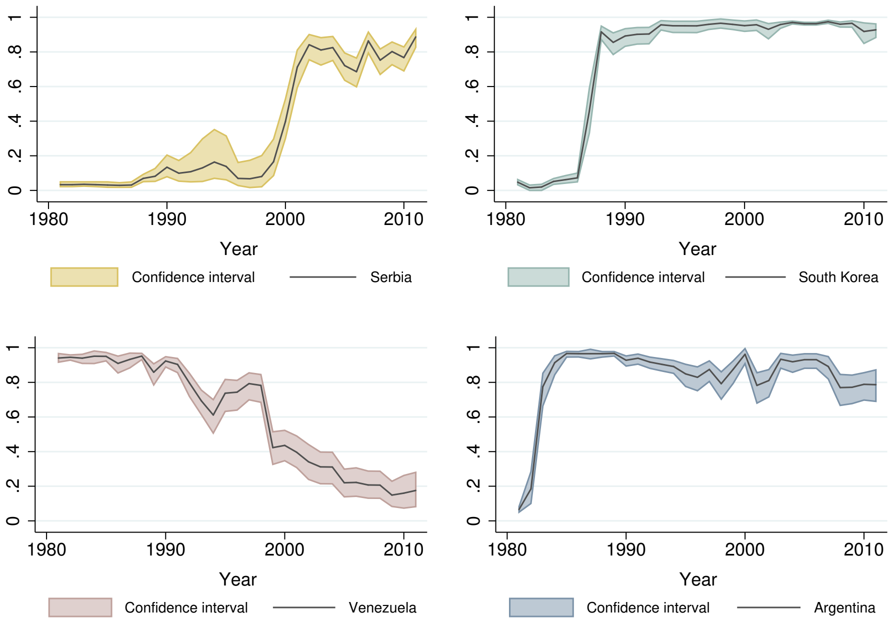


Figure 6.4 The path of democratization. SVMDI scores and confidence intervals of Serbia, South Korea, Venezuela and Argentina, whole period (1981–2011).

military in the region's political culture as well as the strong institutional position of its armed forces. Such a movement has already been ushered in in Venezuela and Paraguay (see Zagorski, 2003). As Venezuela was part of the Third Wave in the early 1980s, the SVMDI implies that it succeeded in establishing democratic structures at that time. However, the relapse was not long in coming. The data suggest a clear tendency towards autocracy which was initiated in the early 1990s and constantly promoted during the presidency of Hugo Chávez (Levitsky and Murillo, 2008).

Similar tendencies arose in Argentina during the presidency of Carlos Menem (1989–1999). While Argentina's democracy in the mid 1980s was more stable than previous regimes, it failed to establish enduring democratic institutions (Levitsky and Murillo, 2008). In fact, President Menem increasingly limited both the power of the congress and the independence of the Supreme Court (Larkins, 1998), resulting in Argentina's movement towards a *delegetive democracy* shaped by weak control mechanisms between different state agencies (O'Donell, 1994). With the continuation of presidential dominance and centralization of power (Elias, 2015, Levitsky and Murillo, 2008), Argentina's political and economic institutions remained strikingly weak under the presidency of Néstor Kirchner (2003–2007) and his wife Christina Fernández de Kirchner (2007–2015).

6.4.5 Relation to existing democracy indicators

One huge advantage of the SVMDI algorithm is that aggregation of the underlying attributes is much less arbitrary compared with traditional indicators, as it relies on much weaker assumptions. In particular, unification of attributes is conducted via a nonlinear optimization problem rather than via crude aggregation rules or the implicit assumption of equal weights. In addition, combining information from existing democracy indicators compensates for weaknesses in conceptualization as well as for potential measurement errors in the underlying secondary data. A direct result of these methodical improvements is a substantial increase in the level of detail in comparison with established approaches.

To demonstrate the superiority of the SVMDI algorithm, Figure (6.5) plots the democracy levels of Jamaica, Nicaragua, Venezuela, and Mongolia as gauged by SVMDI and several other indicators. Note that all indices are normalized to to the $(0, 1)$ interval in order to ensure sufficient comparability of the measurements.[110]

First, consider the case of Jamaica. What is striking in terms of the classification of the Jamaican democracy is the huge divergence between the trends observed in the early 1980s by the SVMDI and those identified by alternative measures. While the Polity scores and the Freedom House (2014) ratings do not change notably, the SVMDI score experiences a sharp decline in the year 1983. Given the political situation in that year, the result suggested by the SVMDI algorithm is much more plausible. In 1983, the "People's National Party"—until that time the largest opposition group in the parliament—boycotted the election, which resulted in the incumbent "Jamaica Labor Party" winning all seats in the parliament (Figueros, 1985). In fact, whereas 54 of 60 seats were completely unopposed, voting took place in six seats due to participation of minor parties. However, nationwide voter turnout was only 2.7 percent, which was the lowest value in the history of the country and the only time that it was below 50 percent (Wüst, 2005). From that time until 1989, Jamaica was a de facto one-party state. Such a situation, however, should factor negatively into a democracy measure, as political pluralism in parliament is an important aspect of democracy, even in minimal concepts such as that proposed by Dahl (1971). Without the control and criticism provided by a parliamentary opposition, the ruling party is able to exercise power without supervision. In fact, the rule of Edward Seaga, Prime Minister of Jamaica from 1980 to 1989, became increasingly authoritarian, which led to widespread public protest during the election in 1989 (Wüst, 2005).

The case of Nicaragua highlights a typical pattern of the Vanhanen (2000) index, which in the overwhelming majority of observations only changes (slightly) after elections have taken place. In Nicaragua, elections are held every five years. While the Vanhanen-index implies an increase in democracy in each electoral year, it remains unaltered during the interim period. In particular, with the exception of a minor decline in 2011, the index provides no indication for a decrease in the degree of democracy during the entire period. Likewise, the Polity score (Marshall et al.,

110 It is crucial to emphasize that the superiority of the SVMDI score in describing recent political developments is not limited to the illustrated countries, but can be observed with respect to the overwhelming majority of country-years included in the data.

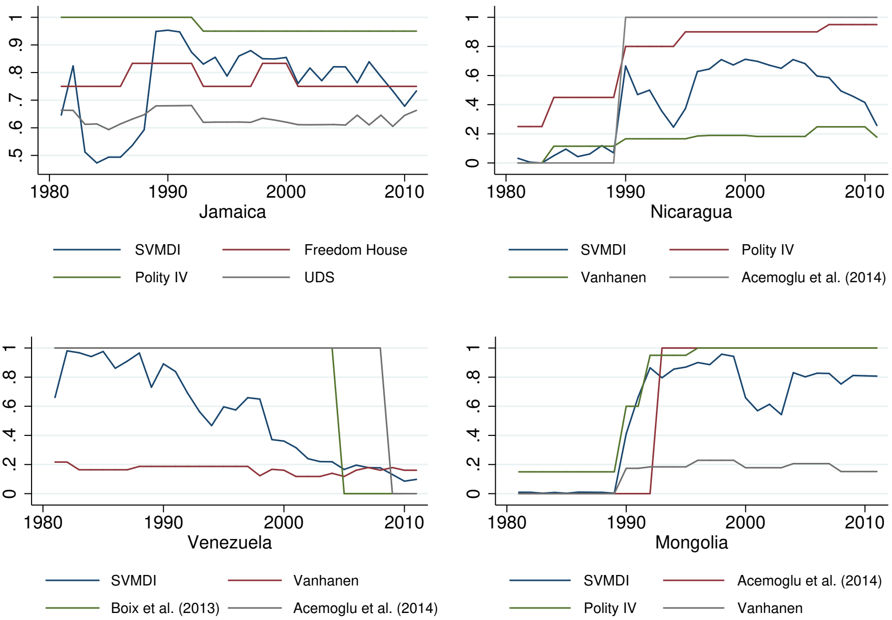


Figure 6.5 Democracy in Jamaica, Nicaragua, Venezuela, and Mongolia. SVMDI and traditional democracy indicators, 1980–2011.

2014) implies a similar period of flourishing democracy without any indication of an interruption. The dichotomous indicator of Acemoglu et al. (2014) changes only once, in 1990, the year when the first competitive election in the country took place (Williams, 1990). Failing to account for the consensus that Nicaragua's democracy is far from being in full bloom (Walker, 2009), the indicator suggests strong democratic structures in the country. In contrast, the SVMDI displays a continuous loss of democracy since 2006, the year when Daniel Ortega came into his second presidency after years as a member of the opposition. Due to the increasingly autocratic governance of President Ortega—including, for instance, growing oppression of critical journalists and opposition members, as well as controversial constitutional amendments (Anderson and Dodd, 2009, McConnell, 2014)—a decreasing trend is much more justifiable than a constant or even increasing level.

The third nation illustrated in Figure (6.5) is Venezuela. As highlighted in Figure (6.4) in the previous section, democratization in Venezuela experienced a decline during the past decades. This phenomenon is intensely discussed in the literature as a "reverse wave" of democracy (see, e.g., Huntington, 2012). However, the breakdown in Venezuelan democracy is captured quite differently by traditional democracy indicators. Whereas the indices from Boix et al. (2013) and Acemoglu et al. (2014) attest to a thriving democracy until the end of the 2000s, the index of Vanhanen

(2000) remains at a constant (low) level of roughly 0.20 over the whole period between 1981 and 2011, indicating no notable decline in democracy at all. The SVMDI, however, illustrates that the antidemocratic trend in Venezuela had already begun during the 1990s, which is much more reflective of the existing literature (see, e.g., Zagorski, 2003 and Levitsky and Murillo, 2008).

The last country depicted in Figure (6.5) is Mongolia. The figure highlights that the SVMDI algorithm is able to detect differences between country-years which have originally obtained a label in step two, i.e. observations that are elements of $\mathcal{L}$. Although Mongolia received a label of 1 for the period between 1999 and 2011, the figure clearly shows that the degree of democracy has changed considerably during this time.[111] What is striking about the figure is the sharp decline in the SVMDI of Mongolia in 2000. In this particular year, the ex-communist Mongolian People's Revolutionary Party (MPRP) won 72 of 76 seats, resulting in Mongolia's shift towards a one-party system (Severinghaus, 2001). Such a development, however, stands in contrast to some established definitions of democracy that typically require a multiple-party system. In fact, political competition is a central issue in theoretical and empirical concepts relating to democracy (see, for instance, Dahl, 1971, Vanhanen, 2000, Huntington, 2012). As the vote in the 2004 Mongolian parliamentary election was evenly split between the MPRP and the Motherland Democratic Coalition, Mongolia's SVMDI experienced a renewed increase. When relying on traditional indicators—such as Polity IV and the measures of Vanhanen (2000) and Acemoglu et al. (2014)—no changes in democratization are observable.

6.5 The empirical effect of democracy on growth

6.5.1 Estimation strategy

What is the relationship between democracy measured by SVMDI and long-run economic development? This section is concerned with an examination of this question, using a framework similar to that of Chapter (5) which is built on 5-year averages of all variables. Averaging the data is necessary due to the long-term perspective of growth theory, the need to disentangle short-term fluctuations and long-term effects, and the occurrence of gaps in the data for some of the covariates. Considering additive linkage of the variables, the basic dynamic panel specification is[112]

$$y_{it} = \theta y_{it-1} + \lambda h_{it} + \beta \mathbf{X}_{it} + \gamma d_{it} + \eta_i + \xi_t + v_{it}, \tag{6.4}$$

[111] In order to make these slight differences computable, the algorithm only uses a subset $\mathcal{T}_\zeta \subset \mathcal{L}$ with $|\mathcal{T}_\zeta| \ll |\mathcal{L}|$ to estimate $\mathfrak{F}_{\mathcal{T}_\xi}(\cdot)$ in iteration ζ. This procedure enables detection of possible differences between country-years that have been classified with $d_{\{0,1\}} = 1$ (democratic) in the second step.

[112] This specification is obtained by following the model structure developed in a number of recent empirical investigations where the growth rate is modeled to evolve as $y_{it} - y_{it-1} = (\theta - 1)y_{it-1} + \lambda h_{it} + \beta \mathbf{X}_{it} + \gamma d_{it} + \eta_i + \xi_t + v_{it}$ (see, e.g., Bond et al., 2001, Voitchovsky, 2005, and Halter et al., 2014).

where y_{it} is the log of initial per capita GDP in i at 5-year period t, h_{it} is human capital endowment, d_{it} is the democracy index, and $\mathbf{X}_{it}$ includes the covariates of the regression. As in the previous chapters, the selection of the covariates is based on the standard framework of Barro (2003, 2013a). These variables include the logarithmic value of real per capita GDP in $(t-1)$ to account for conditional convergence, denoted by log(GDP$_{pc}$); the investment share (INVS); government consumption (GOVC); the inflation rate (INFL); the degree of openness (OPEN); and the log of the fertility rate, log(FERT). Human capital enters into the equation using average years of schooling (SCHOOLY) and log(LIFEEX), the log of life expectancy at birth, to proxy education and health, respectively.[113] As with the investigations in the previous chapters, the analysis does not include measures of physical capital, as their calculation relies on arbitrary assumptions regarding depreciation and the initial value. Again, the idea here is to follow Barro (2003, 2013a) in assuming that higher levels of log(GDP$_{pc}$) and h_{it} reflect higher levels of capital endowment.

Equation (6.4) also captures country-specific effects η_i and time effects of period t, denoted by ξ_t, in order to account for the various institutional aspects of the countries. The term $v_{it} \equiv u_{it} - \xi_t - \eta_i$ denotes the idiosyncratic error of the model.

A common and widely-used approach to account for both unobserved heterogeneity and endogeneity is application of the estimator proposed by Arellano and Bond (1991). Define for reasons of lucidity that $\nabla k \equiv (k_{it} - k_{it-1})$ and $\nabla_2 k \equiv (k_{it-1} - k_{it-2})$, the basic idea of this approach is to adjust (6.4) to

$$\nabla y = \theta \nabla_2 y + \lambda \nabla h + \gamma \nabla d + \beta \nabla \mathbf{X} + \nabla \xi + \nabla v \tag{6.5}$$

and then use sufficiently lagged values of y_{it}, h_{it}, d_{it}, and $\mathbf{X}_{it}$ as instruments for the first-differences. However, first differencing Equation (6.4) removes the information in the equation in levels. This drawback is substantial when analyzing the long-run relationship between political rights and growth, as the variation in democracy data stems to a large extent from the cross section rather than the time-dimension. This particularly holds for hitherto existing democracy indicators. Blundell and Bond (1998) and Bond et al. (2001) show that the standard first-difference GMM estimator can be poorly behaved if time-series are persistent or if the relative variance of the fixed effects η_i is high. The reason is that lagged levels in these cases provide only weak instruments for subsequent first-differences, resulting in a large finite sample bias. In addition, difference GMM magnifies gaps in unbalanced panels as it requires at least three consecutive lags for each of the variables. This requirement results in an asynchronous loss of observations because data availability is typically more limited in developing countries. However, it is particularly necessary to include observations from developing economies, as these country-years contain information regarding the growth effect of regime change in transition economies.

System GMM proposed by Arellano and Bover (1995) and Blundell and Bond (1998) provides a tool to circumvent the previously described biases, if one is willing

[113] The data used in the regression stem from commonly used data sources in empirical growth research. Log(GDP$_{pc}$), INVS, GOVC, OPEN and INFL are from PWT 8.0 as documented in Feenstra et al. (2013), SCHOOLY is from Barro (2013b), log(LIFEEX) and log(FERT) are from World Bank (2014b).

to assume a mild stationary restriction on the initial conditions of the underlying data generating process.[114] In this case, additional orthogonality conditions for the level equation in (6.4) can be exploited, using lagged values of ∇k and $\nabla_2 k$ as instruments. By these means, system GMM maintains some of the cross-sectional information in levels and exploits the information in the data more efficiently. Satisfying the Arellano and Bover (1995) conditions, system GMM has been shown to have better finite sample properties (see Blundell et al., 2000). To detect possible violations of these assumptions, the documentation of each regression reports Difference-in-Hansen tests.

When introducing $\mathbf{\Theta}'_{it} \equiv [y_{it}\ h_{it}\ d_{it}\ \mathbf{X}'_{it}]$, the moment conditions used for the regressions in first-differences can be written as

$$\mathrm{E}[(v_{it} - v_{it-1})\mathbf{\Theta}_{it-s}] = 0 \text{ for } t \geq 3,\ 2 \leq s \leq 3,$$

and the additional moment conditions for the regression in levels are given by

$$\mathrm{E}[(v_{it} + \eta_i)(\mathbf{\Theta}_{it-1} - \mathbf{\Theta}_{it-2})] = 0 \text{ for } t \geq 3,$$

which implies that the instrument matrix is restricted to lag 3. Roodman (2009c) illustrates the need to introduce such a restriction, as otherwise the problem of "instrument proliferation" may lead to severe biases. In principle, the specification can be estimated using one-step or two-step GMM. Whereas one-step GMM estimators use weight matrices independent of estimated parameters, the two-step variant weights the moment conditions by a consistent estimate of their covariance matrix. Bond et al. (2001) show that the two-step estimation is asymptotically more efficient. Yet it is well known that standard errors of two-step GMM are severely downward biased in small samples. The analysis therefore relies on the Windmeijer (2005) finite sample corrected estimate of the variance, which yields a more accurate inference.

6.5.2 Baseline results

Panel A of Table (6.1) reports the results of the baseline regressions. The first column illustrates the effect of democracy measured by the SVMDI in a restricted model where the only covariate is the initial income level. The advantage of examining the effect of democracy in a very reduced specification is that the estimated parameter captures the full growth effect of democracy, leaving all possible transmission channels open. In addition, this estimation enables the investigation of SVMDI in a broad sample of 160 countries. The subsequent columns examine the effect of the SVMDI when additional controls are introduced; however, limited data availability for the covariates yields a decline in the number of countries included in the estimation. Panels B and C use exactly the same specifications as Panel A, but examine the influence of *initial* democracy in $(t-1)$ as well as nonlinear effects of democracy.

The result in Column (1) of Panel A provides a clear indication that democracy and income increases are positively and significantly related. The column rejects the

[114] The assumption on the initial condition is $\mathrm{E}(\eta_i \nabla y_{i2}) = 0$, which holds when the process is mean stationary, i.e. $y_{i1} = \eta_i/(1-\theta) + v_i$ with $\mathrm{E}(v_i) = \mathrm{E}(v_i \eta_i) = 0$.

Table 6.1 The effect of SVMDI on growth, dependent variable is real per capita GDP growth.

	(1)	(2)	(3)	(4)
	Panel A: Baseline regression results			
Log(GDP$_{pc}$)	0.00479	-0.00839**	-0.0180***	-0.0197***
	(0.00492)	(0.00342)	(0.00349)	(0.00309)
SVMDI	0.0264***	0.0242***	0.0149**	0.00294
	(0.00941)	(0.00840)	(0.00750)	(0.00684)
INVS		0.120***	0.0467	0.0445
		(0.0346)	(0.0310)	(0.0323)
SCHOOLY		0.00225	0.00214*	0.000111
		(0.00199)	(0.00123)	(0.00129)
Log(LIFEEX)			0.102***	0.0635***
			(0.0222)	(0.0206)
GOVC			-0.0112	-0.0168
			(0.0304)	(0.0291)
INFL			-0.00126*	-0.00110
			(0.000651)	(0.000680)
OPEN			0.00625*	0.00268
			(0.00331)	(0.00356)
Log(FERT)				-0.0333***
				(0.00643)
	Panel B: The effect of initial democratization			
SVMDI($t-1$)	0.0223*	0.0293**	0.0193*	0.00974
	(0.0134)	(0.0123)	(0.0108)	(0.00684)
	Panel C: Non-linear effect of democracy			
SVMDI	0.121***	0.0189	0.00944	0.00512
	(0.0431)	(0.0391)	(0.0227)	(0.0199)
SVMDI SQUARED	-0.107**	0.00297	0.00428	-0.000991
	(0.0446)	(0.0421)	(0.0235)	(0.0197)
SLM p-val	0.0284	1.000	1.000	1.000
Observations	1048	857	775	775
Countries	160	129	128	128
Hansen p-val	0.0000928	0.0262	0.878	0.991
Diff-in-Hansen	0.109	0.691	1.000	1.000
AR(1) p-val	0.0416	0.0777	0.116	0.119
AR(2) p-val	0.367	0.273	0.335	0.327
Instruments	40	78	154	173

Notes: Table reports two-step system GMM estimations. All estimations use period fixed effects and Windmeijer-corrections, robust standard errors in parentheses. The instrument matrix is restricted to lag 3. Test statistics refer to Panel A. Hansen p-val. gives the p-value of Hansen's J-test, AR(1) p-val. and AR(2) p-val. report the p-values of the AR(1) and AR(2) test. Diff-in-Hansen reports the p-value of the C statistic of the difference in the p-values of the restricted and the unrestricted model. The unrestricted model ignores the Arellano and Bover (1995) conditions. $*p < .10, **p < .05, ***p < .01$

hypothesis of convergence, reflecting the well-known argument in empirical growth research that convergence can only be detected when holding constant a number of variables that distinguish the countries (see, for instance, Barro and Sala-i-Martín, 1992). For this reason, the subsequent columns gradually introduce a number of standard controls in empirical growth regressions. The motivation for including additional controls is twofold. First, Hansen's p-value points to an omitted variable problem in the reduced regression in Column (1), which may result in a bias in the estimated parameter. Second, the analysis aims to investigate the mechanism through which democracy affects incomes by introducing potential transmission channels of democracy, as suggested by Tavares and Wacziarg (2001).

When introducing the investment share and the average years of schooling in Column (2), *conditional* convergence in the form of a negative relationship between initial incomes and growth can be observed. What is remarkable in this estimation is the robustness of the effect of SVMDI, which remains significantly positive and maintains its magnitude. Column (3) incorporates life expectancy at birth, government consumption, the inflation rate, and the openness of countries. The effect of democracy remains positive and significant, but the marginal effect shrinks slightly. The latter observation is in line with the findings of Doucouliagos and Ulubaşoğlu (2008), who show that inclusion of these additional covariates reduces the marginal impact of democracy on growth. Investigating bivariate correlations between SVMDI and the newly introduced covariates, the data implies that democracies tend to have higher life expectancies (correlation: 53 percent) and a lower probability of hyperinflation (-31 percent). Each of these effects stimulates growth, which is why the column suggests a lower marginal impact of SVMDI. Finally, when introducing the fertility rate, the effect of democracy becomes insignificant. As democracies tend to have substantially lower fertility rates (correlation: -60 percent), the fertility channel appears to be a crucial transmission mechanism of democracy on growth. In countries where non-democratic structures are prevalent, the trade-off between the quantity and the education of the children is often resolved in favor of having more offspring. In light of binding budget constraints, families may consider this a substitute for missing social security systems.

The test statistics given in the lower part of Table (5.2) highlight the high degree of validity of the results. The AR(2) p-value illustrates that there is no second-order serial correlation in the residuals. In addition, once additional controls are introduced in Columns (2)–(4), the p-values of the corresponding J-tests suggest that an omitted variable bias becomes increasingly unlikely. Finally, the Difference-in-Hansen statistics underline the validity of the instrument subsets used for the level-equation, implying superiority of system GMM over difference GMM.

Overall, there is a clear indication of a positive effect of democracy measured by SVMDI on the growth rate. This effect remains positive and significant in Panel B, which investigates the impact of the initial democratization level via inclusion of SVMDI in $(t-1)$. Whereas its marginal effect in the reduced specification in Column (1) slightly declines from 0.0264 to 0.0223, the influence of initial democracy tends to be marginally stronger than current democracy in the subsequent regressions. As in Panel A, the effect of democratization vanishes once additional controls are introduced that account for the transmission channels of democracy, particularly the fertility rate.

Some authors have stressed a non-linear relationship between democracy and growth, arguing that democracy enhances income increases at low levels of political freedom but depresses growth once a moderate level has been attained (see, e.g., Barro, 1996). In dictatorships, an increase in political rights may be growth enhancing due to the advantages arising from limitations on governmental power, increases in contractual freedom, and reductions in foreign trade barriers. At high levels of democracy, however, a further increase may eventually be an impediment to growth due to increases in redistributive efforts. Panel C deals with the examination of a possible nonlinear effect of democracy by inclusion of the squared SVMDI score. Whereas Column (1) provides indication of a parabolic influence of democracy on growth, the effect vanishes when additional covariates are incorporated. The Sasabuchi-Lind-Mehlum (SLM) test of Lind and Mehlum (2010) also indicates the presence of an inverted-U relationship in the reduced model, but does not detect a similar pattern in the more comprehensive specifications.

6.5.3 Sensitivity analysis I: Different estimation techniques

While the findings so far highlight a strong positive relationship between democracy and growth, it is important to explore whether the results are sensitive to the specified estimation strategy. Table (6.2) provides the results of two adjustments of Table (6.1). The first adjustment is first-difference GMM as proposed by Arellano and Bond (1991), and the second method uses Within-Group estimations. Both methods have been applied in recent studies concerning the effect of democracy on income increases (e.g. in Acemoglu et al., 2014, Rodrik and Wacziarg, 2005 and Gerring et al., 2005). The table reports three variants of each technique. The first specification is the reduced model of Column (1) of Table (5.2), while the second and third columns refer to the more comprehensive models reported in Columns (3) and (4) of Table (5.2). The columns are labeled in accordance with the variant of the baseline table that is used for specification.

Overall, the effect of democratization is remarkably stable across the regressions conducted in Table (6.2), strongly resembling the findings of the baseline estimations in significance and magnitude. One exception is the effect of SVMDI in the reduced model reported in Column (1), where Hansen's J-test again suggests an omitted variable problem. In addition, the Difference-in-Hansen test reported in Table (6.1) indicates that the additional moment conditions used in the system GMM estimation are valid, implying substantial efficiency losses when utilizing difference GMM. Note also that the number of observations declines from 1,048 to 888, as difference GMM requires observations for at least three consecutive periods. This technique reflects variations over time and eliminates the information in the equation in levels. Thus, when conducting difference GMM estimations, it can be expected that the main effect of democracy appears via the transition of non-democracies to democracies. Differencing the data, however, mainly yields losses of precisely the observations that contain information on the long-run effect of democracy, i.e. observations from developing economies during the transition process.

When introducing additional controls in Columns (3) and (4), the positive and significant effect of SVMDI found in the baseline model reappears. This is a strong indication that democracy exerts its influence via a number of transmission channels which have opposing effects on growth. If the specification does not control for the

Table 6.2 The effect of SVMDI on growth, different estimation techniques. Dependent variable is real per capita GDP growth.

	First-difference GMM (Arellano-Bond)			Within-Group (WG)		
	(1)	(3)	(4)	(1)	(3)	(4)
Log(GDP$_{pc}$)	-0.139***	-0.0781***	-0.0756***	-0.0329***	-0.0589***	-0.0579***
	(0.0341)	(0.0131)	(0.0136)	(0.00636)	(0.00873)	(0.00849)
SVMDI	-0.00214	0.0325**	0.0258*	0.0279***	0.0134**	0.00881
	(0.0407)	(0.0134)	(0.0133)	(0.00584)	(0.00616)	(0.00600)
INVS		0.0816**	0.0784**		0.0808**	0.0709**
		(0.0360)	(0.0358)		(0.0325)	(0.0322)
SCHOOLY		0.00292	-0.00343		0.00813***	0.00300*
		(0.00468)	(0.00595)		(0.00170)	(0.00176)
Log(LIFEEX)		0.0218	0.00948		0.133***	0.121***
		(0.0475)	(0.0432)		(0.0245)	(0.0231)
GOVC		0.0269	0.0290		-0.00852	-0.00502
		(0.0320)	(0.0328)		(0.0213)	(0.0212)
INFL		-0.000960	-0.000678		-0.000731	-0.000721
		(0.000636)	(0.000489)		(0.000549)	(0.000543)
OPEN		0.00288	0.00346		-0.00107	-0.000940
		(0.00460)	(0.00569)		(0.00400)	(0.00387)
Log(FERT)			-0.0278			-0.0405***
			(0.0194)			(0.00866)
Observations	888	647	647	1048	775	775
Countries	160	128	128	160	128	128
Hansen p-val	0.00841	0.211	0.263			
AR(1) p-val	0.0582	0.113	0.115			
AR(2) p-val	0.0590	0.221	0.230			
Instruments	27	99	111			

Notes: Table reports first-difference GMM (Arellano-Bond) and Within-Group (WG) estimations. Robust standard errors in parentheses. WG uses cluster robust standard errors. The instrument matrix in Columns (1)–(3) is restricted to lag 3. Hansen p-val. gives the p-value of Hansen's J-test, AR(1) p-val. and AR(2) p-val. report the p-values of the AR(1) and AR(2) test. $*p < .10$, $**p < .05$, $***p < .01$

effects of these variables, the estimated parameter of SVMDI captures the neutralizing effects of the transmission variables and becomes insignificant.

The Within-Group (WG) estimations also strongly support the results of the baseline table. This technique resembles the estimation strategy conducted by Gerring et al. (2005), Rodrik and Wacziarg (2005) and Papaioannou and Siourounis (2008). However, one concern is that introducing a lagged dependent variable in a WG model most likely results in a Nickell (1981) bias. In addition, WG does not account for possible problems caused by endogeneity, which are typically to be expected in growth regressions.

6.5.4 Sensitivity analysis II: Regional and cultural waves of democratization

This section considers another branch of sensitivity analyses, conducting IV regressions in which SVMDI is instrumented with regional and cultural democratization.

This technique, used in some more recent studies of the topic (see, e.g., Acemoglu et al., 2014 and Madsen et al., 2015), is motivated by the empirical observation that democratization often occurs in waves. Section (6.4.4) demonstrates that the SVMDI measure implies a multinational trend in democratization in the world during the 1980s and the early 1990s, which Huntington (1991, 2012) refers to as "Democracy's Third Wave". In addition, the renunciation of authoritarian regimes during the Arab Spring provides more recent experience with regional entanglements in the process of democratization. Spreading from one country to another, waves of democratization may be a satisfactory determinant of exogenous variation in democracy (Persson and Tabellini, 2009).

The analysis in this section follows Acemoglu et al. (2014) in assuming that, conditional on covariates, democratization in neighboring countries should be uncorrelated with a country's national GDP.[115] This allows for the creation of external instruments of democracy which capture the effect of democratization waves.

The analysis uses two different approaches to generate external instruments. The first approach is based on that of Acemoglu et al. (2014), instrumenting country-year $\langle i, t\rangle$ with jack-knifed average SVMDI of region r (denoted by Z^r_{it}) in which i is located. In order to satisfy the exclusion restriction, the computation leaves out i in the calculation of Z^r_{it}. The crucial challenge in computing Z^r_{it} is the accurate definition of the decisive regions. Whereas a narrower concept is more likely to include the countries that directly influence national demand for democracy, it bears the risk of leaving out information necessary to accurately instrument national SVMDI scores. In addition, arbitrary classification of regions may cause a distortion in the results. For this reason, Table (6.3) uses two different definitions of region. The first (wide) definition refers to the country classification of the World Bank, the second (narrower) definition splits each continent into four disjoint regions, as illustrated in appendix A6-1. The second approach for obtaining external instruments weights the SVMDI of the countries by their cultural distance from i, denoted with $\tilde{Z}^r_{it}$. While this procedure builds on the method proposed by Madsen et al. (2015), the examination uses the cultural dimensions from Hofstede (2001) to capture cultural diversity rather than linguistic differences. The advantage of $\tilde{Z}^r_{it}$ is that the exclusion restriction may be more likely to be fulfilled, as culturally similar countries are not necessarily in the immediate geographic vicinity of one another. The creation of the instruments is described in detail in Section (A1) of the appendix of this chapter.

The estimation strategy used in Table (6.3) follows Acemoglu et al. (2014) and Madsen et al. (2015), using 2SLS with cluster-robust standard errors and including country-fixed and period-fixed effects.[116]

[115] Whereas one could imagine plausible reasons why this assumption may be violated—e.g. due to a decline in regional trade or capital flows—Acemoglu et al. (2014) provide evidence that controlling for such effects has little influence on the estimation results.

[116] Whereas the authors of both studies use real per capita GDP as the dependent variable in their IV regressions, the dependent variable in Table (6.3) is again the growth rate of real GDP per capita to ensure comparability with the baseline results. Note that exact replication with inclusion of SVMDI as democracy variable yields quite similar results. Note also that the results of a more direct comparison to the baseline table achieved by inclusion of the external instruments in the System GMM estimations strongly resemble the baseline findings.

Table 6.3 The effect of SVMDI on growth, IV estimations. Dependent variable is real per capita GDP growth.

	Regional Democracy (World Bank)		Regional Democracy (Narrower definition)		Cultural Democracy (Culturally-weighted)	
	(1)	(4)	(1)	(4)	(1)	(4)
	Panel A: 2SLS regression results					
SVMDI	0.293***	0.224**	0.213***	0.119***	0.257***	0.0518
	(0.0747)	(0.109)	(0.0493)	(0.0429)	(0.0787)	(0.0496)
Log(GDP$_{pc}$)	-0.0573***	-0.0684***	-0.0514***	-0.0683***	-0.0416***	-0.0539***
	(0.0124)	(0.0143)	(0.00939)	(0.0121)	(0.0146)	(0.00917)
INVS		0.0897*		0.0859**		0.0691*
		(0.0541)		(0.0433)		(0.0401)
SCHOOLY		0.00110		0.00219		0.00261
		(0.00341)		(0.00231)		(0.00203)
Log(LIFEEX)		0.142***		0.135***		0.113***
		(0.0454)		(0.0266)		(0.0430)
GOVC		-0.0391		-0.0282		-0.0466*
		(0.0412)		(0.0298)		(0.0260)
INFL		-0.000527		-0.000536		-0.000734
		(0.000617)		(0.000584)		(0.000803)
OPEN		-0.0172		-0.0107*		-0.00971
		(0.0109)		(0.00609)		(0.00651)
Log(FERT)		0.0000956		-0.0196		-0.0361***
		(0.0259)		(0.0147)		(0.0120)
	Panel B: First-stage regression results					
Democracy wave	0.41551***	0.30557**	0.47916***	0.37924***	0.2568***	0.05182
$(t-1)$	(0.1070)	(0.1221)	(0.9224)	(0.1051)	(0.0787)	(0.0495)
Observations	893	671	893	671	544	445
Countries	157	128	157	128	94	83
F p-val	0.00001	0.00000	0.00000	0.00000	0.00287	0.00000

Notes: Table reports 2SLS estimations, where SVMDI is instrumented with jack-knifed regional and cultural democracy. All estimations include country fixed effects, cluster robust standard errors in parentheses. Test statistics and number of included countries refer to the estimations conducted in Panel A. F p-val gives the p-value of the F-Statistic of the reported model. Labels of the columns refer to the corresponding specification reported in the baseline estimations in Table 5.2. $*p < .10$, $**\,p < .05$, $***p < .01$

Panel A of Table (6.3) reports the 2SLS results, with first-stage outcomes presented in Panel B. The results from this exercise strongly support the positive effect of democracy found in Table (5.2). However, when instrumenting SVMDI with regional democratization waves, the reduced models imply an increase in the marginal effect of SVMDI from 0.0264 in the baseline specification to 0.293 in Table (6.3). The results also seem to be relatively unaffected by the classification of regions r, as both the categorization of the World Bank and the narrower concept yield outcomes strongly comparable in significance and magnitude. This also holds if the SVMDI variable is instrumented by culturally-weighted waves of democracy. The marginal effect in the reduced model strongly resembles the effect detected in Column (1). As in the baseline estimations, the SVMDI ceases to be significant once the fertility rate is introduced in the model.[117]

Panel B highlights a strong effect of regional democratization waves in $t-1$ on national SVMDI scores, suggesting that Z^r_{it-1} is a valid instrument for SVMDI.[118] The first-stage regressions also highlight that $\tilde{Z}^r_{it-1}$ is less valid than Z^r_{it-1}. In the reduced model, cultural waves of democratization are significantly related to national democracy; however, the marginal effect is smaller compared with regional democratization waves. In the comprehensive model specification in the last column, there is no indication for a contribution of cultural democracy waves to the SVMDI in country i.

Comparing the outcomes of Table (6.3) to a similar analysis conducted by Acemoglu et al. (2014), the results suggests that utilization of SVMDI is superior to application of a rough dichotomous measure, as it yields much more significant results.[119] The reason for this is the substantial increase in the level of detail achieved by the Support Vector classification of the underlying data. Even when controlling for regional democratization waves, the strong heterogeneity in the subset of democratic (autocratic) countries—which necessarily occurs when conducting a binary classification—results in a loss of information that causes a distortion of the estimated results. Note also that the IV approach is likely to suffer from a Nickell bias unless the (bold) assumption holds that $\mathrm{E}[Z^r_{it-1}\varepsilon_{it}] = 0$ and ε_{it} is serially uncorrelated.

6.5.5 The effect of alternative democracy indices on growth

Whereas the previous results provide strong evidence for a positive effect of democracy on growth when applying the SVMDI measure, it is important to examine if these results are superior when compared with estimations which use alternative indices of democracy. Whenever the available indices lack observations for recent periods (e.g. Vanhanen, 2012) or have not yet been made available (e.g. Acemoglu et al., 2014), missing values are computed according to the algorithms reported in the original documentations. To assess the effect of these variables on growth, the analysis is based on two different estimation techniques, difference GMM and system GMM.

Difference GMM has been used in a number of recent studies (e.g. in Gerring et al., 2005 and Acemoglu et al., 2014). The general idea of this technique, shown in Equation (6.5), is to eliminate unobserved heterogeneity by first-differencing the specified model, i.e. first-differencing Equation (6.4). However, this transformation removes the information in the equation in levels, so that the estimation relies solely on the within-country information. In the context of the relationship between democracy and growth, this means that the estimated parameter essentially captures

[117] Similar to the baseline results reported in Table (5.2), SVMDI significantly contributes to income increases in each specification other than model (4).

[118] SVMDI are instrumented by only one lag of Z^r_{it}. In accordance with Acemoglu et al. (2014), the effects differ only slightly when using more lags of Z^r_{it} as instruments.

[119] The same increase in significance occurs if the utilized specifications are replicated more directly, using Log(GDP$_{pc}$) as dependent variable.

Table 6.4 The effect of different democracy indicators on growth. Dependent variable is real per capita GDP growth.

	SVMDI	POLITY	VANHANEN	ACEMOGLU	FREEDOM	BOIX	UDS
	Panel A: Difference GMM estimations						
DEMOCRACY	0.0320**	0.000680	0.000860	0.00829	0.00851	0.00755	0.00850
	(0.0151)	(0.000918)	(0.000529)	(0.0116)	.00698	(0.0104)	(0.00723)
Observations	616	616	616	616	616	616	616
Countries	122	122	122	122	122	122	122
Hansen p-val	0.214	0.170	0.0968	0.221	0.210	0.199	0.226
AR(1) p-val	0.118	0.120	0.116	0.122	0.115	0.120	0.115
AR(2) p-val	0.229	0.240	0.237	0.229	0.220	0.231	0.236
Instruments	99	99	99	99	99	99	99
	Panel B: System GMM estimations						
DEMOCRACY	0.0161**	0.000972**	0.000684***	0.0119**	0.00585	0.00640	0.00698**
	(0.00751)	(0.00049)	(0.00026)	(0.00543)	0.00356	(0.0055)	(0.00336)
Observations	737	737	737	737	737	737	737
Countries	122	122	122	122	122	122	122
Hansen p-val	0.946	0.924	0.904	0.945	0.959	0.930	0.949
Diff-Hansen	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AR(1) p-val	0.121	0.120	0.118	0.122	0.119	0.122	0.118
AR(2) p-val	0.345	0.348	0.352	0.342	0.337	0.344	0.346
Instruments	154	154	154	154	154	154	154

Notes: Table reports two-step system GMM estimations. All estimations use period fixed effects and Windmeijer-corrections, robust standard errors in parentheses. The instrument matrix is restricted to lag 3. Hansen p-val. gives the p-value of Hansen's J-test, AR(1) p-val. and AR(2) p-val. report the p-values of the AR(1) and AR(2) test. Diff-in-Hansen reports the C statistic of the difference in the p-values of the restricted and the unrestricted model. The unrestricted model ignores the Arellano and Bover (1995) conditions. $*p < .10$, $**p < .05$, $***p < .01$

the effect of democratization *within* countries, i.e. the process of transformation towards or away from democracy.

Panel A of Table (6.4) illustrates the results of the difference GMM estimations, replicating the specification of Column (3) in Table (6.2) using SVMDI and six commonly used democracy indicators. To exclude the possibility of a sample selection bias, the estimations rely on the set of observations that are available for all indicators. As in Section (6.5.3), the SVMDI detects a positive and significant effect of the democratization process within countries on their growth rate. However, neither of the alternative indicators suggests a similarly significant influence, a result which strongly resembles the effects identified in many recent studies.[120] Since (non-)democratic countries differ in numerous historical, cultural, political, and institutional aspects, first-differencing the model requires indicators that react quite sensitively to political events in order to capture the effect of transition towards democracy within countries. As illustrated in Section (6.3), hitherto existing democracy indicators are unable to react with sufficient sensitivity to political events and regime changes. For this reason, raw measures of democracy—particularly dichotomous indices—provide little

[120] Note that this result also occurs if other model specifications are used, e.g. Column (4) of Table (6.2) and Columns (2)–(4) of the baseline estimations of Table (6.1).

indication of an income-enhancing effect of democratization, whereas such an effect is clearly visible in Table (6.4).

Since most of the variation in traditional democracy indicators stems from the cross-section rather than the time-dimension, the utilization of additional orthogonality conditions proposed by Arellano and Bover (1995) and Blundell and Bond (1998) is beneficial, as these additional restrictions ensure that some of the information from the equation in levels is maintained. With respect to the estimation of the democracy-growth nexus, this implies that the estimated parameters also capture the between variation, i.e. the variation in the level of democracy between the countries in the sample. In addition, as difference GMM requires information from at least three consecutive periods in order for a country to be included in the estimation, the exploitation of the Arellano and Bover (1995) orthogonality conditions also yields an increase in the number of observations. This is crucial, as one might expect a loss of observations for developing countries in particular, which possess a higher within variation of democratization than advanced economies. Panel B of Table (6.4) reports the results of system GMM using the same model specifications as in Panel A. This yields a change in the picture. The SVMDI index maintains its positive and strongly significant effect on growth. Additionally, four of the six alternative indices now point to a similar influence of democracy on growth.

Overall, the results of Table (6.4) broadly indicate that democracy is positively related to growth. However, only the SVMDI indicates that the road to democracy is beneficial to growth. From an economic perspective, this implies that small steps towards democracy already lead to long-run increases in living standards, even if political rights in the countries do not catch up with those of established democracies. Meanwhile, reverse waves of democratization are always harmful to growth in the long-run. Once the econometric specification allows for the investigation of differences in the democracy level across countries, the positive effect of democracy can be observed as a clear empirical pattern, even if the model relies on raw measures of democracy.

6.6 The transmission channels of democracy

In line with Tavares and Wacziarg (2001), the results of the previous estimations suggest that political rights exert their influence on growth via a number of transmission channels. This section is concerned with a more in-depth analysis of these mechanisms.

Table (6.5) illustrates the effect of democracy on schooling, investment, redistribution, and fertility. Each of these variables plays an important role in the growth process; however, it is crucial to disentangle the effects of democracy from those of credit availability. Whereas democracy may increase schooling and investment via a more equal distribution of opportunities and fewer government interventions in the private sector, it simultaneously contributes to better credit availability. It has been emphasized in the growth literature that mitigation of credit market imperfections yields an increase in education and physical capital investments (see, e.g., Galor and Zeira, 1993, Galor and Moav, 2004). Similarly, Chapter (3) highlights that this growth effect of the financial sector is particularly pronounced in developing countries, while the effects dissipate during the development process. For this reason, the analysis in this section specifies two models for each of the transmission variables: the first variant simply uses the variables of the specifications in Table (5.2), while the second variant additionally introduces private credit to GDP (CREDIT) as a proxy for credit availability.[121] As expected, the correlation between SVMDI and CREDIT is high (50 percent). In order to account for a change in the transmission channels of CREDIT due to different levels of development, the specifications also incorporate the level of real per capita GDP.

The empirical framework follows Acemoglu et al. (2014), conducting Within-Group (Panel A) and 2SLS (Panel B) estimations. The latter once more uses regional waves of democratization as external instruments for domestic democracy. Due to the high probability of a potential Nickell (1981) bias in the "small" T panel, the specification does not include lagged dependent variables. SVMDI enters in the regressions with a lag of one period to ensure that causality runs from democracy to the transmission variables, rather than the reverse.

The first transmission channel in Table (6.5) is concerned with education. The results imply that wealthier economies exhibit a higher average level of school attainment. In addition, better health as measured by life expectancy enhances education. The trade-off between the quantity and the education of children is clearly visible, as the outcomes suggest a significantly negative impact of fertility on education. Controlling for these effects, the influence of democratization is positive in the Within-Group estimations and becomes significant in Column (2) when CREDIT is introduced. Likewise, SVMDI is significant in both specifications of the 2SLS estimations. The results imply that better credit availability softens the budget constraints of the household, thereby contributing to a higher level of education of individuals. However, even when controlling for this effect, the impact of democracy still acts as an additional source of educational improvements.

The second transmission channel illustrates the effect on investment, which is positive in both the Within-Group and the 2SLS estimations. Apparently, democratic structures and political rights facilitate both national and foreign investments and capital inflows. These findings are in line with the well-known results of Perotti (1996), who finds that political stability—which is considerably greater in democracies (see, e.g., Feng, 1997)—has a huge impact on investment and growth. Thus, the growth mechanisms discussed in Chapters (2) and (3) seem to be particularly pronounced in countries with advanced political rights. Resembling the analysis conducted in Chapter (3), CREDIT again has no significant effect on investment if the whole sample of developed and advanced economies is employed. Chapter (3) reveals that this effect is due to a fundamental change in the effect of finance during the development process. Moreover, the result may also suggest that the positive contribution of the SVMDI stems largely from foreign investments, which are not necessarily financed by loans acquired in the target country.

[121] The data source is World Bank (2014b).

Table 6.5 Estimations for the transmission channels of democracy.

	Schooling		Investment		Redistribution		Fertility	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
	Panel A: Within-Group regression results							
Log(GDP$_{pc}$)	1.036***	0.650***	0.00803	0.0000725	0.000969	-0.00556	0.00845	-0.0501
	(0.189)	(0.183)	(0.0223)	(0.0234)	(0.00606)	(0.00553)	(0.0347)	(0.0333)
SVMDI($t-1$)	0.202	0.293*	0.0294**	0.0295*	0.00178	0.000810	-0.114**	-0.0522
	(0.185)	(0.161)	(0.0132)	(0.0151)	(0.00379)	(0.00429)	(0.0453)	(0.0364)
INVS	-0.166	-0.327			-0.0781***	-0.0783***	-0.310**	-0.370***
	(0.702)	(0.701)			(0.0216)	(0.0220)	(0.148)	(0.138)
SCHOOLY			0.000906	-0.00284	0.00571***	0.00492***	-0.115***	-0.120***
			(0.00552)	(0.00594)	(0.00152)	(0.00140)	(0.0165)	(0.0166)
Log(LIFEEX)	2.585***	2.123**	0.230***	0.229***	0.0300**	0.0263**	-0.210	-0.192
	(0.961)	(0.848)	(0.0729)	(0.0768)	(0.0140)	(0.0131)	(0.160)	(0.151)
GOVC	-0.729	-0.632	-0.107	-0.126	0.00198	0.00242	0.0972	0.0950
	(0.720)	(0.689)	(0.0920)	(0.0862)	(0.00904)	(0.00880)	(0.109)	(0.0985)
INFL	-0.0139**	-0.0108	-0.000165	-0.00119*	0.000405	0.000387	0.00034	0.00072
	(0.00668)	(0.00727)	(0.00134)	(0.000601)	(0.000275)	(0.000307)	(0.0007)	(0.0006)
Log(FERT)	-2.710***	-3.022***	-0.0633**	-0.0810***	0.0162**	0.0125*		
	(0.306)	(0.296)	(0.0294)	(0.0296)	(0.00679)	(0.00654)		
OPEN	0.107	0.122	-0.00218	-0.00359	0.00524	0.00469	0.00246	0.0112
	(0.123)	(0.112)	(0.0105)	(0.00979)	(0.00348)	(0.00334)	(0.0221)	(0.0189)
CREDIT		0.688***		0.00757		0.0126**		0.122***
		(0.220)		(0.0185)		(0.00581)		(0.0434)
REDIST			-0.683***					
			(0.189)					
	Panel B: 2SLS regression results							
SVMDI($t-1$)	2.052***	2.192***	0.0755*	0.0867	0.0161	0.0162	-0.469***	-0.420**
	(0.706)	(0.713)	(0.0424)	(0.0539)	(0.0107)	(0.0125)	(0.146)	(0.174)
Observations	670	648	560	648	560	544	670	648
Countries	128	126	121	126	121	119	128	126
R-squared	0.607	0.645	0.240	0.200	0.130	0.150	0.532	0.568
F-statistic	43.10	46.57	7.454	8.023	3.545	3.307	28.59	29.24
Model p-val	1.23e-30	6.06e-34	4.94e-08	1.08e-08	0.00102	0.00124	2.52e-23	2.99e-25

Notes: Table reports Within-Group and 2SLS estimations. Model specification of the 2SLS estimations is identical to the Within-Group variant. Cluster robust standard errors in parentheses. Test statistics refer to the Within-Group models. F-statistic reports the test statistic of joint significance of the model, Model p-val gives the p-value of the F-test. $*p < .10, **p < .05, ***p < .01$

To examine a possible negative effect of an increase in political rights in countries with a medium or high level of SVMDI, Column (1) also incorporates the level of effective redistribution, measured identically as in Chapter (5) by the difference in the Gini coefficient of household incomes before and after taxes and transfers.[122] The results show a strongly significant impact of redistribution on investments, where a greater amount of redistribution is negatively related to investment activity. This, in principle, supports the hypothesis that a higher level of democratization may be an impediment to growth. However, this mechanism only comes into play if democracy enhances redistribution.

This effect is investigated in the third branch of transmission analyses. The findings imply that redistribution is lower in countries with a higher average level of education. Meanwhile, countries with higher life expectancy, higher government consumption, and higher fertility rates typically tend to redistribute more. Controlling for these effects, there is no additional contribution of SVMDI to redistribution, neither in the Within-Group regressions nor in the 2SLS estimations. This implies that the strong bivariate correlation between SVMDI and REDIST (63 percent) is not due to an inherent causality running from democracy to redistribution, but is the result of numerous variables that are affected by democracy. The ambiguous effect of democracy on redistribution strongly resembles the recent findings of Acemoglu et al. (2013). However, Feld and Schnellenbach (2014) emphasize that the manner in which income is redistributed differs between countries, depending on the respective constitutional framework.

The last transmission channel deals with the effect of democracy on fertility. The first column highlights that democratization yields a significant decline in fertility rates. The process of democratization is often accompanied by a substantial increase in social security systems and a reduction of uncertainty due to greater political stability, both of which reduce families' incentives to have children as a substitute for social protection. However, it is crucial to disentangle the different effects of democracy and credit availability, as illustrated in Column (2). When holding constant CREDIT, the effect of democracy shrinks, but remains negatively and—in case of the 2SLS estimations—significantly associated with fertility. Better credit availability increases the fertility rate, as access to capital markets alleviates the otherwise binding trade-off between the quantity and the education of children.

In sum, the findings suggest that democracy exerts its influence on growth via better education, higher investment shares, and lower fertility rates. In contrast, there is little evidence for a redistribution-enhancing effect of democratization.[123]

[122] The data source is the SWIID version 5 compiled by Solt (2009, 2016). For a detailed description of the inequality data in the SWIID, see Chapter (5).

[123] There is also no robust effect of democracy on health, even though both variables reveal a high bivariate correlation (53 percent). However, it turns out that there is a significant impact of initial wealth on life expectancy. Whereas one would suspect that democratic countries provide better public health services, the estimations imply that incomes are much more decisive for health than regime types are. Yet, life expectancy may be a poor proxy in this context, as changes in this variable may only occur a considerable amount of time after democratization has taken place.

6.7 Conclusions

Employing reliable measurements regarding democracy is essential for achieving a sound understanding of democratization and its effects on political and economic outcomes. The overwhelming majority of existing indicators, however, are fraught with methodical problems. Scholars using such rough measurements will find, not infrequently, that an inappropriate democracy indicator is the Achilles heel of empirical analyses, particularly when working with panel data.

By maximizing comparability for the broadest possible sample of countries, the SVMDI algorithm facilitates empirical investigations of democracy. A direct result of this methodological progress is a substantial increase in the level of detail in comparison with established approaches. In addition, the algorithm places the crucial question of how to aggregate the underlying attributes—undoubtedly the main weak point of alternative indicators—into the context of a nonlinear optimization problem, thereby obtaining much more consistent and plausible results. The unprecedented potential of machine learning enables researchers to make highly accurate classifications, and may also yield very promising results for problems in the field of economics beyond its utilization for measuring democracy.

Using the SVMDI in empirical growth regressions reveals a robust positive influence of democracy on long-run economic growth. The results imply that the ambiguity in recent studies stems from two main sources. First, in light of the diversity of political institutions across countries, the inability of traditional democracy indicators to react sufficiently to political events and regime changes only allows for a rough classification of democracy. Second, when using empirical models that rely on the within-country variation, the problem of inadequate and insensitive measurement of democracy becomes particularly severe.

A more in-depth analysis of the democracy-growth nexus provides little indication of a nonlinear relationship between the variables. The analysis of the transmission channels through which democracy exerts its influence on growth illustrates why: whereas democratic countries typically have better educated populations, higher investment shares and lower fertility rates, there is little evidence for a redistribution-enhancing effect of democratization.

Taken together, the results emphasize that democratic structures facilitate economic growth in the long-run, and their implementation may be a beneficial strategy for less-developed countries. However, countries differ in numerous cultural, historical, political, and institutional dimensions. Isolating the growth effect of different aspects of democratic institutions may thus be an advantageous field of future research. Likewise, it would be beneficial to acquire a deeper empirical understanding of the transmission channels of democracy, particularly with regard to health, inequality, and redistribution.

6.A Appendix of Chapter (6)

Appendix A1: Description of the external instruments used in the IV regression

Let $\mathcal{R} = \{1, ..., R\}$ denote a set of regions, where each country i belongs to exactly one region r. In addition, let N_{rt} be the number of countries in region r at period t and d_{it} denote the level of democracy in country-year $\{i, t\}$. Then the regional democratization wave—i.e. instrumental variable Z_{it}^r—is calculated via

$$Z_{it}^r = \frac{1}{N_{rt} - 1} \sum_{\{j \neq i | r' = r, r' \in \mathcal{R}\}} d_{jt}.$$

To build the culturally weighted instrumental variable of democracy, four of the cultural dimensions—*Power Distance* (PD), *Individualism* (IN), *Masculinity* (MC), and *Uncertainty Avoidance* (UA)—provided by Hofstede (2001) are used to calculate instrument variables via a four-stage approach. First, it is required to calculate the Euclidean distance

$$\delta_{ij} = \sqrt{(PD_i - PD_j)^2 + (IN_i - IN_j)^2 + (MC_i - MC_j)^2 + (UA_i - UA_j)^2} \quad (6.6)$$

for each set of countries $\{i, j\}$. Subsequently, δ_{ij} is normalized to the interval from 0 to 1 by application of the standard formula

$$\bar{\delta}_{i,j} = \frac{\max_{i,j}\{\delta_{i,j}\} - \delta_{i,j}}{\max_{i,j}\{\delta_{i,j}\} - \min_{i,j}\{\delta_{i,j}\}}, \quad (6.7)$$

which is used, for instance, for generation of the Human Development Index (see United Nations, 2013). The third stage calculates the cultural weights $\lambda_{i,j}$ via

$$\lambda_{i,j} = \frac{\bar{\delta}_{i,j}}{\sum_{k \neq i} \bar{\delta}_{i,k}} \quad (6.8)$$

to ensure that the weights sum up to 1 for each country i. Finally, the external instrument $\tilde{Z}_{it}^r$—which gives the cultural weighted democracy score for a certain country-year $\{i, t\}$—is computed as follows

$$\tilde{Z}_{it}^r = \sum_{k \neq i} \lambda_{i,k} \mathrm{SVMDI}_{k,t}. \quad (6.9)$$

Table A6-1 Classification of regions in the IV regression in Table (6.3).

I. ASIA	
Central Asia	Afghanistan, Armenia, Azerbaijan, Bhutan, Georgia, India, Iran, Kazakhstan, Kyrgyzstan, Maldives, Mongolia, Nepal, Pakistan, Sri Lanka, Tajikistan, Turkmenistan, Uzbekistan
East-Southeast Asia	Bangladesh, Cambodia, China, Japan, Laos, Myanmar, North Korea, South Korea, Taiwan, Thailand, Vietnam
Arabic Region	Bahrain, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates, Yemen
Oceania	Australia, Brunei Darussalam, Fiji, Indonesia, Malaysia, New Zealand, Papua New Guinea, Philippines, Samoa, Singapore Solomon Islands, Tonga, Vanuatu
II. EUROPE	
Central-Northern Europe	Austria, Belgium, Denmark, Finland, Germany, Iceland, Ireland, Luxembourg, Netherlands, Norway, Sweden, Switzerland, United Kingdom
South-Southwest Europe	Cyprus, France, Greece, Italy, Malta, Portugal, Spain
East Europe	Belarus, Czech Republic, Estonia, Latvia, Lithuania, Moldova, Poland, Russia, Slovakia, Ukraine
Balkan States	Albania, Croatia, Bulgaria, Hungary, Kosovo, Macedonia, Montenegro, Romania, Serbia, Slovenia
III. AFRICA	
North Africa	Algeria, Egypt, Libya, Morocco, Tunisia
Central-East Africa	Cameroon, Central African Republic, Chad, Djibouti, Eritrea, Ethiopia, Kenya, Somalia, South Sudan, Sudan
West Africa	Benin, Burkina Faso, Cape Verde, Cote d'Ivoire, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone, Togo
Southern Africa	Angola, Burundi, Comoros, Democratic Republic of the Congo, Republic of the Congo, Equatorial Guinea, Gabon, Lesotho, Madagascar, Malawi, Mauritius, Mozambique, Namibia, Rwanda, São Tomé and Príncipe, Seychelles, South Africa, Swaziland, Tanzania, Uganda, Zambia, Zimbabwe
IV. AMERICA	
North America	Bahamas, Canada, United States
Central America	Belize, Costa Rica, El Salvador, Grenada, Guatemala, Honduras, Mexico, Nicaragua, Panama
South America	Argentina, Brazil, Chile, Colombia, Ecuador, Guyana, Paraguay, Peru, Suriname, Uruguay, Venezuela
Caribbean	Antigua and Barbuda, Barbados, Cuba, Dominica, Dominican Republic, Haiti, Jamaica, St. Kitts and Nevis, St. Lucia, St. Vincent, Trinidad and Tobago

Chapter 7
Concluding remarks

> *"To understand why the human race has become so much wealthier and why our wealth is shared so inequitably among the inhabitants of the world, we need to understand what drives economic growth."*
> (Philippe Aghion and Peter Howitt)

The astounding variation in living standards across countries is one of the most striking phenomena of the current era. These massive differences are the result of economic growth during the past centuries. The goal of this book was to contribute to a better understanding of the empirical determinants that drive economic development. What can be learned from the analysis? This chapter briefly reviews the main conclusions that can be drawn from the previous investigations.

7.1 The key findings of this book and their implications

The point of departure was the analysis of the effect of entrepreneurship on economic development. Chapter (2) highlighted that a higher average propensity of a nation's citizens to carry out entrepreneurial activity leads to a corresponding increase in long-run growth. The reason for this positive contribution lies in the advancement of knowledge and technological progress. As Chapter (2) has documented, attitudes towards entrepreneurship are substantially coined by cultural socialization, which is why there are few measures which economic policy may implement to stimulate growth *directly*. However, a direct implication of the empirical findings is that establishment of a fruitful environment for entrepreneurship—shaped by the rule of law, individual and contractual freedom, and a low degree of bureaucracy—facilitates economic development in the long run. For this reason, creation of an advantageous political and economic framework enhances the ability of individuals to carry out entrepreneurial activity, thereby stimulating economic development *indirectly*.

This is particularly important for poor countries, where institutions are on average much less developed than in the advanced economies. High degrees of bureaucracy and corruption are stumbling blocks for potential entrepreneurs in a number of nations in the developing world. The World Bank (2014b) reports the number of days required to legally start a business, emphasizing a relatively low level of bureaucracy in the developed countries, most notably in France (4 days), Norway (4), the United Kingdom (5), and the United States (6). While these figures also point to some room for improvement with respect to bureaucratic regulations in some of the rich nations—such as Germany, where eleven days are needed to meet the legal requirements—, the challenges pale in comparison to that of the developing world: with 73 days necessary to open a business in Laos, 66 days in Namibia, 84 days in Eritrea, 50 days in Bolivia, and 48 days in Indonesia, administrative barriers in a number of developing countries in Asia, Africa, and Latin America are much higher. These hurdles also include corruption: The CPIA index that measures transparency, accountability, and corruption in the public sector points to particularly unfavorable administrative environments in Zimbabwe, Sudan, Yemen, Cambodia, Eritrea, Bangladesh, and Laos (see World Bank, 2014b).

In the second step, the analysis was enlarged by the introduction of the financial sector. The analysis showed that in cases where the initial wealth of individuals is insufficient to carry out their desired investment projects, the financial system may work as a lubricant for the growth engine identified in Chapter (2). Additionally, the provision of funds may also facilitate investments in human capital, providing means to soften the budget constraint of families. However, the analysis also demonstrated quite clearly that finance and growth are linked via a nonlinear relationship. According to the results of Chapter (3), the financial sector primarily contributes to growth in developing countries, but becomes less relevant once incomes reach a more sophisticated level. The reason is that the transmission channels through which the financial sector affects growth change in the development process.

In poor economies, loans from the financial to the private sector facilitate investments in education, physical capital, and innovation activity. Policy measures that encourage financial development in this case may contribute to an increase in the growth rate. Closely related, Belke and Wernet (2015) show that policy actions that benefit investment in physical capital help to reduce poverty rates. However, barriers to carrying out investment in many of the developing countries in Asia and Africa do not only depend on financing, but also on cultural forces, preferences, and detrimental economic policies (see Banerjee and Duflo, 2012). To circumvent some of these problems, an intense discussion has recently arisen on how to best provide financial funds in the developing world. While the concept of micro credits has gained an increasing amount of attention, some studies disenchant their potential to enhance small business investment, health, education, and women's empowerment (see, for instance, Banerjee et al., 2015).

The development process of economies brings with it an improvement in the quality of public schooling systems, greater equality of opportunities, and lower fertility rates. These fundamental changes over the transition process yield a reduction in the importance of financial development for educational achievements. In advanced economies, the financial sector primarily fosters productivity gains, which is why the overall growth effect of financial development in this case depends on the rate of technological progress. Consequently, the analysis shows that the negative effect of finance is particularly strong in times when factor productivity grows at low rates, providing little growth potential via support of innovation activity.

In fact, therein lies the key for the explanation of the "vanishing effect of finance", as the past 15 years saw a tremendous decline in technological progress, particularly in the advanced economies. This development was initiated around the turn of the millennium and led to disproportionately low growth rates compared to the historical development of incomes during the past 70 years. Chapter (4) shows that a lack of radically new ideas contributed to this development, arguing for a supply-side explanation for what has recently become known as "secular stagnation" (see Summers, 2015). The chapter illustrates that the growth mechanism driven by innovative entrepreneurs and financial institutions falters in times when new ideas are scarce. For this reasons, policy measures that provide incentives and support for (basic) research may help reinitiate the growth mechanism documented in Chapters (2)–(3). A second emphasis of Chapter (4) is that countries with below-average progress in educational achievements are particularly exposed to secular stagnation. This may explain why the German growth rate was particularly affected by the decline of productivity gains during the early 2000s. In accordance with this view, a widely recognized study conducted by the OECD highlights that equity in the German education sector substantially lags behind many of the OECD member states (OECD, 2012). This presents a number of challenges for economic policy, most notably the need to prevent "cream skimming" of schools, the reduction of segregation, and the integration of migrants and refugees. The data provided by OECD (2012) suggests that particularly the education output of the latter group falls alarmingly short of the mean value.

Whereas large parts of Chapters (2)–(4) of this book indirectly implied that wealth is distributed equally among individuals, Chapter (5) introduces inequality of incomes into the empirical analysis. In the context of the growth mechanisms discussed in Chapters (2)–(4), disparity of incomes implies that some households have the ability to conduct their investment projects, while insufficient wealth of others prevents some individuals from exploiting their full intellectual and entrepreneurial potential. The result is a decline in education, innovation activity, and investment in physical capital, which translates to a lower rate of economic growth. In general, functioning capital markets can help to mitigate this problem. However, particularly in the case of developing economies, capital markets are often imperfect. To prevent negative effects on growth, direct redistribution by the government may seem to be an adequate policy measure to equalize investment opportunities among individuals. The empirical analysis in Chapter (5) shows that income inequality impairs growth via several channels, most notably by raising fertility rates, decreasing education and—to a lesser extent—lowering investment in physical capital. However, these effects—important as they are for poorer economies—vanish during the development process. When estimating the aggregate growth effect of redistribution—its direct negative effect and its indirect positive effect resulting from lower net inequality—the findings imply that the two effects are offsetting. While this result would advocate for more governmental redistribution, a more in-depth examination that distinguishes between different development levels reveals that redistribution is beneficial in poor economies, but rather harmful in rich economies. It bears underscoring, however, that when pursuing the goal of increasing living standards in the developing world, the concrete form of redistributive policy actions is highly decisive and often counter-intuitive (see Banerjee and Duflo, 2012). The effects on redistribution and growth that are triggered by different policy actions that seek to enhance incomes are still poorly explored and understood.

Finally, the last chapter of this book considered the political environment of countries. In light of the growth mechanisms discussed in Chapter (2)–(3) that sets in via investment decisions of individuals, we may expect that political institutions have a substantial influence in the growth process. This particularly includes individual and economic freedom, freedom of contract, property rights, and legal liability. There is much reason to suspect that these requirements are more likely to be met in democracies rather than in authoritarian regimes. Chapter (6) introduces a machine learning algorithm to classify the degree of democratization. The most fundamental improvement of this technique is that the aggregation rule used to combine information from different variables does not depend on assumptions. Rather, the problem of classification is transferred to a nonlinear optimization problem, where the unknown functional link between the output and the input variables is "learned" by the algorithm. Based on this new measure, the chapter highlighted that there is a distinct relationship between political institutions and growth. Countries with a higher level of democratization are on average able to increase their standards of living more rapidly compared with countries with authoritarian forms of government. The analysis of the transmission channels showed that democracy exerts its influence via better education, higher investment shares, and lower fertility rates. In contrast, there is little evidence for a redistribution-enhancing effect, which may explain why the analysis does not detect nonlinear effects of democracy in comprehensive model specifications.

The methodological contribution of Chapter (6) provides promising potential to solve further classification problems in the social science. In fact, scholars in economics, sociology, political science, and psychology routinely utilize composite measures for a wide range of different purposes. The vast and highly promising literature concerning artificial intelligence, machine learning, and mathematical optimization provides advantageous possibilities to transfer the SVM-based approach to these purposes. Additionally, there are some possibilities to augment the SVM approach in future research projects. Further research may, for instance, concentrate on statistical tests to identify the significance of attributes used in the Support Vector regression and classification tools, as well as the utilization of ε-tube widths that are permitted to adapt automatically to the data. The close relationship between the Support Vector procedure and the Maximum Likelihood framework discussed in Section (A.2.6) enables construction of a statistical test whose functioning is similar to that of a Likelihood-ratio test. In addition, Oracle inequalities may be particularly helpful for calibration and empirical variable selection (see, for instance, George and Foster, 2000).

7.2 A brief outlook at the future of economic development

While this book provides some empirical insights into the causes of economic growth, the challenge of understanding development processes requires further exploration of uncharted territory. Viewing growth through a development economist's lens, Banerjee and Duflo (2005) show that there is a large disparity in marginal returns of certain investment opportunities in both physical and human capital. This argues for a departure from aggregate production functions, where marginal returns would have to be more or less equalized. The reasons for these differences in returns are market failures in the markets for labor, credit, and insurance, slow diffusion of knowledge concerning new opportunities, and a multitude of behavioral problems (for a detailed discussion, see Banerjee and Duflo, 2012). These observations may yield a new strand of growth theories which move beyond endogenous growth models. In fact, there have been some recent attempts to derive novel classes of growth theories, of which the Unified Growth Theory (UGT) outlined in Section (1.2) is certainly the most popular representative. Aiming to contribute to a better understanding of growth patterns during the entirety of human history, the UGT also seeks to explain the large period of Malthusian Stagnation spanning from the year 1 to (at least) 1500. While the theory has been developed in a number of promising articles (see Galor and Weil, 1999, 2000, Galor and Moav, 2002, 2004, and Galor, 2011), there is still little empirical evidence concerning the UGT, and its main ideas are far from being broadly accepted in the economic profession (see, for instance, the criticism of Nielsen, 2013).

While understanding economic growth is difficult enough, predicting its future development is a task which is impossible to accomplish. In the words of the great physicist Niels Bohr, *"prediction is difficult, especially about the future"*. However, what is possible is to consider the factors that are decisive for the future development of prosperity. Weil (2005) provides an illustrative summary of the crucial drivers upon which the growth potential of the next several decades will be dependent. Obviously, for rich countries, the most important source of future growth will be technological progress. By its very nature, technological advancement is difficult to forecast. In fact, the last decades saw a number of predictions that turned out to be woefully misguided.

The growth potentials of technological change also depend on the question of whether global economic integration is likely to continue along its current path. Although gains from trade and technological diffusion raise national living standards on average, there are undeniably some individuals within every country who lose out as a result of globalization. Another factor is the future of the world demographic. While the 20th century was shaped by spectacular demographic changes in the form of a quadrupling of the world population and steep increases in longevity, the great unknowns of the future demographic path are worldwide epidemics and the trend in fertility rates.

A related issue is the question of the future development of the world income distribution. Even though incomes in the world are distributed highly unequally (Milanovic, 2013 estimates that the global Gini coefficient is around 70 percent), large parts of the world saw a substantial reduction in poverty during the past decades. In fact, since the early 1980s, more people escaped poverty than ever before in human history. While at that time 44.3 percent of the world population lived with less than 1.90 USD per day, this headcount ratio dropped to 12.7 percent in 2011, the latest period for which comparable data is accessible (see World Bank, 2014b). Particularly in Eastern Asia, poverty decreased tremendously from 80.6 percent in 1981 to 7.2

percent in 2011. However, there are still several regions which have made very little progress. Particularly the Sub-Saharan African countries—where, on average, 42.7 percent of the population live with less than 1.90 USD per day—still struggle with alarmingly poor living conditions. If the growth wave that affected China, India, and other countries from Eastern Asia washes over Africa and Latin America as well, then the next decades may see a similar extent of poverty reduction. In order for this effect to occur, massive changes in the economic policy of some of the countries are necessary so that they may escape their poverty gaps. The hope is that this book has contributed to the understanding of some of these mechanisms.

Appendix A
Mathematical Appendix

This appendix briefly illustrates the mathematical methods used in the chapters of this book, with a particular focus on the basic ideas underlying the application of machine learning, pattern recognition algorithms, and Support Vector Machines for estimation of real-valued and indicator functions. Machine learning is sometimes considered a subfield of computer science, having evolved from the computational learning theory of artificial intelligence. However, in light of machine learning's strong ties to mathematical optimization, this appendix covers both the mathematical concepts which make up its foundation and the algorithms themselves.

A.1 Machine learning

The field of machine learning is concerned with the study and construction of algorithms that are able to learn from an underlying dataset and make predictions based on this data. In short, these algorithms operate on the basis of a model drawn from example inputs to make data-driven predictions or decisions (see, for instance, Bishop, 2006). The huge advantage gained via application of this method is that computers are given the ability to learn without being explicitly programmed (Samuels, 1959). Substantial progress was made in this field with the introduction of the "Probability Approximately Correct Learning" (PAC) by Valiant (1984), which enables establishment of a real theory on the learnable. A further milestone in the field was reached as a result of the work of Vladimir Naumovich Vapnik. Whereas Vapnik had an early influence on the field via introduction of the Generalized Portrait algorithm (see Vapnik and Lerner, 1963 and Vapnik and Chervonenkis, 1964), probably his most influential contribution was placing the problem of "learning" into the context of an optimization problem, thereby allowing utilization of the large body of already available knowledge on mathematical optimization. Three essential optimization procedures proved to be particularly helpful for the field of machine learning and Support Vector classification. These are Fermat's Theorem for local extrema, the Lagrange Multiplier Rule for conditional optimization under constraints of equality type, and the Theorem of Kuhn-Tucker (sometimes Karush-Kuhn-Tucker) which enables solutions to convex optimization problems under constraints of inequality type. This section briefly introduces the basic ideas of machine learning, while the subsequent sections provide a more elaborate discussion of Support Vectors and their application for function estimation (Section A.2) and classification (Section A.3).

A.1.1 The general problem: an informal introduction

Machine learning can be used in a number of different ways to compile models used for statistical inference, but the most important implementation in this book concerns the problem of pattern recognition. This technique further establishes the foundation upon which large parts of the statistical learning theory rest. To illustrate its general idea, Figure (A.1) shows a simple two-dimensional pattern recognition problem where the goal is to design a machine that is able to separate the black dots from the white dots.

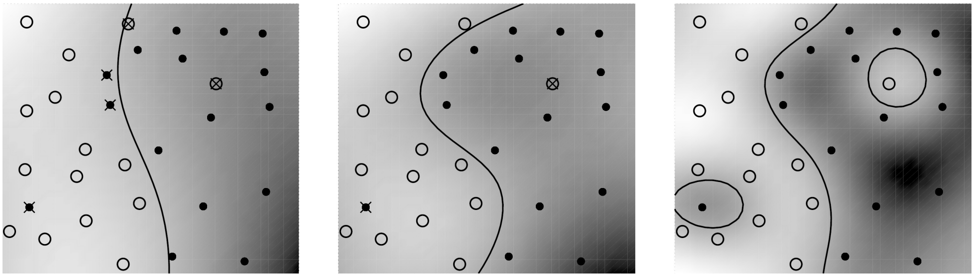


Figure A.1 2-D toy example of binary classification, which is solved with the help of three different models of varying complexity. The example is taken from Schölkopf and Smola (2001).

The picture on the right side highlights a classification function that correctly categorizes each of the training points $i = 1, \ldots, n$. It is, however, not certain beforehand that the same would also be true for a potential training point $n + 1$ that enlarges the given set. The classification would be particularly rough if the new training point emerges close to one of the illustrated "outliers". To avoid overfitting, it is possible to construct a simpler version of the classification, depicted in the picture on the left-hand side of the figure. This classification, however, exhibits the disadvantage of misclassification of a number of dots. This pertains not only to some of the outliers but also to some rather "easy" points that are located near to the separating boundary. The picture in the middle seems to provide an agreeable compromise, conducting the separation with an intermediate level of complexity.

This intuitive example illustrates the challenge machine learning is confronted with. However, the crucial problem is placing these arguments, as obvious as they are intuitively, in a consistent mathematical context.

A.1.2 The general problem: a more formal introduction

In the next step, we want to put the general problem of machine learning in a more formal context. Let $\mathcal{X}$ be a space of examples, $\mathcal{Y}$ a space of labels, and $\mathcal{H}$ a space of hypotheses containing functions mapping $\mathcal{X}$ to $\mathcal{Y}$. Define further the loss function $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ and the probability measure P over the space $\mathcal{X} \times \mathcal{Y}$.

Definition 1. *Risk of hypotheses conditional to L and P.* For any hypothesis $h \in \mathcal{H}$ we define the risk of h conditional on L and P as

$$\rho_P(h) = \underset{P}{E}\left(L(h(x), y)\right).$$

The general problem of machine learning can be formulated as the hypothesis $h \in \mathcal{H}$ that solves the optimization problem

$$\min_{h \in \mathcal{H}} \rho_P(h)$$

based on a labeled sample $S = (x_i, y_i)_{i=1}^n \in \mathcal{X} \times \mathcal{Y}$. The complexity of this problem is reduced when the distribution P is given only over the space $\mathcal{X}$ and there exists a labeling function $l : \mathcal{X} \to \mathcal{Y}$. In this case, the optimization problem becomes

$$\min_{h \in \mathcal{H}} \underset{P}{E}\left(L(h(x), l(x))\right).$$

Definition 2. *Agnostically PAC-learnable problem.* A problem is agnostically PAC-learnable if there exists an algorithm $\mathcal{A}$ for any given (ϵ, δ) such that after seeing a sample of the size $n = n(\epsilon, \delta)$, the algorithm returns a hypothesis that satisfies

$$\rho_P(h) - \min_{h \in \mathcal{H}} \rho_P(h) < \epsilon$$

with probability of at least $1 - \delta$ over the sampling process, where $n(\epsilon, \delta)$ is polynomial in $\left(\frac{1}{\epsilon}, \frac{1}{\delta}\right)$.

Proposition 1. *Convergence of $\rho_P(h) - \min_{h \in \mathcal{H}} \rho_P(h)$.* If a hypothesis class $\mathcal{H}$ has finite Vapnik-Chervonenkis (VC), then it holds that

$$\rho_P(h) - \min_{h \in \mathcal{H}} \rho_P(h) \to 0$$

in probability.

Remark 1. VC-dimensions were originally introduced in Vapnik and Chervonenkis (1971). The VC-dimension of a statistical classification algorithm is a measure of the complexity (capacity) of this algorithm. It is defined as the cardinality of the largest set of points that the algorithm can shatter.

Under Proposition 1, the optimization problem simplifies to

$$\min_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \left(L(h(x), y)\right). \tag{A.1}$$

The problem expressed in Equation A.1 is better known as empirical risk minimization (ERM). However, this simplification is often not sufficient to describe the underlying problem. For this reason, it is sometimes referred to as the "raw optimization problem" of machine learning.

A.1.3 Learning guarantees

Definition 3. *Rademacher complexity*. Let $V = \{v : Z \to \mathbb{R}\}$ be a set of functions and let $S = (z_i)_{i=1}^n$ be a training sample, the empirical Rademacher complexity of V defined as

$$\mathfrak{R}_S(V) = \underset{\sigma}{E}\left(\frac{1}{n}\sup_{v\in V}\sum_{i=1}^{n} g(z_i)\sigma_i\right),$$

where σ_i is a uniform random variable drawn from the Rademacher distribution with $\sigma_i \in \{-1; 1\}$. The Rademacher complexity is defined to be the expectation of the empirical Rademacher complexity, i.e.

$$\mathfrak{R}_n(V) = \underset{S}{E}\left(\mathfrak{R}_S(V)\right).$$

The Rademacher complexity measures richness (complexity) of V and gives the expected value of the maximal correlation with random noise. It can be shown that there exists a constant o such that any class of $\{0, 1\}$-indicator has Rademacher complexity $o\sqrt{\frac{VC}{n}}$, where VC is the Vapnik-Chervonenkis dimension.

Proposition 2. Risk and Rademacher complexity. Let $\tilde{L}$ be the $0-1$ loss function and $S = (x_i, y_i)_{i=1}^n$ a given sample, then it holds for any $h \in \mathcal{H}$ with probability of at least $1-\delta$ that (see Koltchinskii, 2001)

$$\rho_P(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{L}(h(x_i), y_i)\right) + \mathfrak{R}_n(\mathcal{H}) + \sqrt{\frac{\log(2/\delta)}{n}}.$$

If L_2 is the square loss and $L_2\left(h(x_i), y_i\right] \leq M\forall(x, y)$, then it follows that

$$\rho_P(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{L}(h(x_i), y_i)\right) + 4M\mathfrak{R}_n(\mathcal{H}) + M\sqrt{\frac{\log(2/\delta)}{n}}.$$

The bounds of Proposition 2 features the ubiquitous problem in machine learning of bias-variance trade-off. It illustrates that for large $\mathcal{H}$, the empirical error can be made small at the cost of an increase in the Rademacher complexity. In contrast, whereas small $\mathcal{H}$ would also be shaped by small values of $\mathfrak{R}_n(\mathcal{H})$, the empirical error in this case is substantially higher. This illustrates the machine learning method to solve the problems discussed in this section by finding the best hypothesis set for the underlying data sample.

A.2 Support Vector Machines for function estimation

A.2.1 Motivation

As noted previously, a natural way to solve the classification problem is empirical risk minimization, i.e.

$$\min_{h \in \mathcal{H}} \sum_{i=1}^{n} L\left(h(x_i), y_i\right).$$

The empirical risk can be thought of as the average training error, e.g. the average error made in the problem illustrated in Figure (A.1). This is where the general idea of Support Vector Machines (SVM) sets in. The SV algorithm is a nonlinear generalization of the Generalized Portrait algorithm, which has been developed by Vapnik and Lerner (1963) and Vapnik and Chervonenkis (1964) in the 1960s and further elaborated by a number of influential contributions during the 1990s. Among the most influential of these articles are Boser et al. (1992), Guyon et al. (1993), Cortes and Vapnik (1995), and Schölkopf (1995). The ideas introduced in these contributions will be discussed in detail in this chapter of the appendix. In its modern application, SVM is part of the standard toolbox of machine learning methods, and its implementation in real-world applications has provided huge benefits in various branches of sciences. For instance, it has produced very promising results for categorization of cancer cells (Guyon et al., 2002), classification of hyperspectral data (Gualtieri, 2009) and identification of imagine biomarkers of neurological and psychiatric diseases (Orrú et al., 2012). In addition Joachims (2002) and Cortes and Vapnik (1995) succeeded in categorization of texts and hand-written alphabetic characters by utilization of SVM. In general, Support Vectors can be used for estimation of classification (indicator) functions or real-valued functions. In what follows, the term “function estimation” refers to the estimation of the latter, while what is called the problem of classification is estimation of indicator functions.

This section is concerned with function estimations based on Support Vectors. This part introduces a number of important theories, theorems, and concepts. The subsequent section introduces Support Vectors for classification, where the description will often recourse to material covered in this section. Both estimation of indicator and real-valued functions are extensively used in this book to construct a new measurement of democracy, which is why this appendix describes both methods in great detail.

A.2.2 The basic idea behind SVM

The following more detailed illustration builds in large parts on Smola and Schölkopf (2004). Let $\{x_1, y_1\}, \ldots, \{x_\phi, y_\phi\} \subset \mathcal{X} \times \mathbb{R}$ where $\mathcal{X}$ is the space of the input pattern, for instance $\mathcal{X} = \mathbb{R}^d$. Vapnik (1995) introduced the approach of ε-SV regression, which aims to (1) find a function $f(x)$ with at most ε deviation from the targets y_i for each observations of the sample and (2) ensure that this function is as flat as possible.

In the linear case, f assumes the form

$$f(x) = \langle w, x \rangle + b \text{ with } w \in \mathcal{X}, b \in \mathbb{R},$$

where henceforth $\langle \cdot, \cdot \rangle$ denotes the dot product in $\mathcal{X}$. Flatness of the desired function means seeking small w, which can be specified via several methods. A widely-used approach to ensure flatness is to minimize the norm, i.e. $||w||^2 = \langle w, w \rangle$. This can be written as a convex optimization problem of the form

$$\min \frac{1}{2}||w||^2 \tag{A.2}$$

$$\text{s.t.} \begin{cases} y_i - \langle w, x_i \rangle - b & \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i & \leq \varepsilon, \end{cases}$$

which tacitly assumes feasibility of the optimization problem i.e. that the desired function exists. In many applications, however, this may not be the case. To solve this problem, Cortes and Vapnik (1995) use a version of the "soft margin" loss function introduced by Bennett and Mangasarian (1992) by introduction of slack variables ξ_i, ξ_i^*. These variables make it possible to deal with the otherwise infeasible constraint of Equation (A.2). In this case, the optimization problem adjusts to (Vapnik, 1995)

$$\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{\phi}(\xi_i + \xi_i^*) \tag{A.3}$$

$$\text{s.t.} \begin{cases} y_i - \langle w, x_i \rangle - b & \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i & \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* & \leq 0, \end{cases}$$

where the constant $C > 0$ determines the trade-off between flatness of f and the amount up to which deviations larger than the critical value ε are tolerated. This method corresponds to the application of the ε-insensitive loss function $|\xi|_\varepsilon$, which can be formulated via

$$|\xi|_\varepsilon := \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{otherwise .} \end{cases}$$

In the majority of cases, the optimization problem is easier to solve in its dual formulation, which is why Lagrange multipliers are applied here as a standard dualization procedure.

A.2.3 The dual problem and quadratic programs

Henceforth, the objective function of the optimization problem is called the *primal* objective function. The key idea of dualization is realized by construction of a Lagrange function of the primal objective function and the corresponding constraints with the help of introduction of a dual set of variables. It can be shown (see, e.g. Mangasarian, 1969) that this Lagrange function has a saddle point with respect to the primal and the dual variables at the solution.

The Lagrangian in this case is given by

$$\begin{aligned} L \quad := & \frac{1}{2}||w||^2 + C\sum_{i=1}^{\phi}(\xi_i + \xi_i^*) - \sum_{i=1}^{\phi}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ & \sum_{i=1}^{\phi}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i\rangle) + b) \\ & \sum_{i=1}^{\phi}\alpha_i^*(\varepsilon + \xi_i + y_i - \langle w, x_i\rangle) - b), \end{aligned}$$

where $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ denote Lagrange multipliers. For this reason, the dual variables must satisfy positivity constraints, such that $\tilde{\alpha}_i, \tilde{\eta}_i \geq 0$ for $\tilde{\alpha}_i$ referring to α_i and α_i^*.

As implied by the saddle point condition, the partial derivation of the Lagrangian with respect to the primal variables must be equal to zero, i.e.

$$\begin{aligned} \partial_b L = & \qquad \sum_{i=1}^{\phi}(\alpha_i^* - \alpha_i) = & 0 \\ \partial_w L = & \quad w - \sum_{i=1}^{\phi}(\alpha - \alpha_i^*)x_i = & 0 \\ \partial_{\tilde{\xi}_i} L = & \qquad C - \tilde{\alpha}_i - \tilde{\eta}_i = & 0 \end{aligned}$$

Substitution of these conditions into the Lagrangian gives the dual optimization problem

$$\max \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{\phi}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle \\ -\varepsilon\sum_{i=1}^{\phi}(\alpha_i - \alpha_i^*) + \sum_{i=1}^{\phi} y_i(\alpha_i - \alpha_i^*) \end{cases}$$

$$\text{s.t.} \sum_{i,j=1}^{\phi}(\alpha_i - \alpha_i^*) = 0, \alpha_i, \text{where } \alpha_i^* \in [0, C].$$

Reformulation of $\partial_{\tilde{\xi}_i} L$ to $\tilde{\eta}_i = C - \tilde{\alpha}_i$ can be used to eliminate η_i, η_i^*. Likewise, $\partial_w L$ can be rewritten to obtain the "Support Vector Expansion"

$$w = \sum_{i=1}^{\phi}(\alpha_i - \alpha_i^*), \text{thus } f(x) = \sum_{i=1}^{\phi}(\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b.$$

This provides the advantage that w can be described completely as a linear combination of x_i. Thus, the complexity of the SV representation of a function is independent of the dimensionality of $\mathcal{X}$, but depends on the number of SVs. It is also worth noting that the complete algorithm can be described in terms of the dot products between the sample data. Evaluation of $f(x)$ does not require explicit computation of w, which will be shown to be a beneficial property for the formulation of a nonlinear extension.

A.2.4 Computation of b

Computation of b can be achieved by making use of the conditions of Karush (1939) and Kuhn and Tucker (1951), often referred to as the Karush-Kuhn-Tucker (KKT) conditions. Application of KKT yields

$$\begin{aligned}
&\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) &= 0\\
&\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) &= 0\\
&(C - \alpha_i)\xi_i &= 0\\
&(C - \alpha_i^*)\xi_i^* &= 0
\end{aligned}$$

The above expressions allow two conclusions to be drawn. First, only samples with $C = \tilde{\alpha}_i$ lie outside the ε-insensitive tube. Second, it follows that $\alpha_i \alpha_i^* = 0$. This implies that there exists no set of dual variables where α_i and α_i^* are simultaneously zero. From this it follows that

$$\begin{aligned}
\varepsilon - y_i + \langle w, x_i \rangle + b \geq 0 &\text{ if } \alpha_i < C\\
\xi_i = 0 &\text{ if } \alpha_i < C\\
\varepsilon - y_i + \langle w, x_i \rangle + b \leq 0 &\text{ if } \alpha_i > 0,
\end{aligned}$$

which yields

$$\begin{aligned}
&\max\{-\varepsilon + y_i - \langle w, x_i \rangle | \alpha_i < C \text{ or } \alpha_i^* > 0\} \leq b \leq\\
&\min\{-\varepsilon + y_i - \langle w, x_i \rangle | \alpha_i > 0 \text{ or } \alpha_i^* < C\}.
\end{aligned}$$

For $\tilde{\alpha}_i^* \in (0, C)$ the inequalities become identities. Further methods to obtain b are discussed in Smola and Schölkopf (2004). From the previous constraints, it follows that the Lagrange multipliers may only be nonzero in the case that $|f(x_i) - y_i| \geq \varepsilon$. This means that the α_i, α_i^* vanish for all samples in the ε-tube. If it holds that $|f(x_i) - y_i| < \varepsilon$, then it follows that

$$\begin{aligned}
(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) \neq 0\\
(\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \neq 0.
\end{aligned}$$

In this case, satisfaction of the KKT requires that $\alpha_i = \alpha_i^* = 0$. This illustrates that the above corresponds to a sparse expansion of w in terms of x_i, meaning that it is not required to draw on information regarding all x_i in order to describe w.

Definition 4. *Support Vectors*. If $\alpha_i, \alpha_i^* \neq 0$, the examples are called "Support Vectors".

A.2.5 Kernels

In most cases, the linear SV model is not sufficient and it is required to make the SV algorithm nonlinear. This can be achieved by preprocessing the training patterns x_i into a space with higher dimension, called the *feature space* $\mathcal{F}$. Such methods have been described by Aizerman et al. (1964b) and Nilsson (1965). The idea is to shift x_i into the feature space by a map $\Phi : \mathcal{X} \to \mathcal{F}$ and then to apply the standard SV regression algorithm.

However, while this approach seems reasonable in some cases, it can easily become computationally infeasible. This holds true for polynomial features of higher order and higher dimensionality. For instance, the number of different monomial features of degree p is $\binom{\dim \mathcal{X}+p-1}{p}$. As illustrated, for instance, by Schölkopf (1995) and Vapnik (1995), typical values for OCR tasks are $p = 7$, $d = 28 \times 28$, which correspond to approximately 3.7×10^{16} features.

A computationally more efficient way is to use kernels instead. As noted previously, the Support Vector algorithm only depends on the dot products between x_i. For this reason, it is not required to that Φ is known explicitly, but suffices rather to know $k(x, x') := \langle \Phi(x), \Phi(x') \rangle$ instead. This allows reformulation of the SV optimization problem

$$\max \begin{cases} -\frac{1}{2} \sum_{i,j=1}^{\phi} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(x_i, x_j) \\ -\varepsilon \sum_{i=1}^{\phi} (\alpha_i - \alpha_i^*) + \sum_{i=1}^{\phi} y_i (\alpha_i - \alpha_i^*) \end{cases}$$

$$\text{s.t.} \sum_{i,j=1}^{\phi} (\alpha_i - \alpha_i^*) = 0, \alpha_i, \text{where } \alpha_i^* \in [0, C]$$

with the obvious expansion corresponding to the previous case

$$w = \sum_{i=1}^{\phi} (\alpha_i - \alpha_i^*) \Phi(x_i), \text{and } f(x) = \sum_{i=1}^{\phi} (\alpha_i - \alpha_i^*) k(x_i, x) + b.$$

Unlike in the linear case, this nonlinear variant no longer gives w explicitly. In addition, this optimization problem seeks to find the flattest function in the feature space rather than the input space. The theorem of Mercer (1909) characterizes functions $k(x, x')$ which correspond to a dot product in the feature space $\mathcal{F}$.

Theorem 1. *Mercer (1909)*. Suppose $k \in L_\infty(\mathcal{X})$ such that the integral operator $T_k : L_2(\mathcal{X}) \to L_2(\mathcal{X})$,

$$T_k f(\cdot) := \int_{\mathcal{X}} k(\cdot, x) f(x) \mathrm{d}\mu(x)$$

is positive (here μ denotes a measure on $\mathcal{X}$ with $\mu(\mathcal{X})$ finite and supp$(\mu) = \mathcal{X}$). Let $\psi_j \in L_2(\mathcal{X})$ be the eigenfunction of T_k associated with the eigenvalue $\lambda_j \neq 0$ and normalized such that $||\psi_j||_{L_2} = 1$ and let ψ_j denote its complex conjugate. Then

1. $(\lambda_j(T))_j \in \phi_1$.
2. $\psi_j \in L_\infty(\mathcal{X})$and $\sup_j ||\psi_j||_{L_\infty} < \infty$.
3. $k(x, x') = \sum_{j\in\mathbb{N}} \lambda_j \psi_j(x)\psi_j(x')$ holds for almost all (x, x'), where the series converges absolutely and uniformly for almost all (x, x').

Remark 2. Less formally speaking, Mercer's theorem implies that if it holds that

$$\int_{\mathcal{X}\times\mathcal{X}} k(x, x')f(x)f(x')\mathrm{d}x\mathrm{d}x' \geq 0 \text{ for all } f \in L_2(\mathcal{X}),$$

it is possible to write $k(x, x')$ as a dot product in some feature space.

Subsequently, functions k that satisfy Mercer's conditions are referred to as admissible SV kernels.

Lemma 2. *Positive linear combinations of kernels*. Let k_1, k_2 be admissible SV kernels and $c_1, c_2 \geq 0$, then

$$k(x, x') := c_1k_1(x, x') + c_2k_2(x, x')$$

is an admissible kernel. This follows directly from Remark 2 by virtue of the linearity of integrals.

In addition, Berg et al. (1984) illustrate that the set of admissible kernels forms a convex cone, closed in the topology of pointwise convergence.

Lemma 3. *Integrals of kernels*. Let $s(x, x')$ be a symmetric function on $\mathcal{X} \times \mathcal{X}$ such that

$$k(x, x') := \int_{\mathcal{X}} s(x, z)s(x', z)\mathrm{d}z$$

exists. Then k is an admissible SV kernel.

Remark 3. This follows directly from Lemma 3 and Remark 2 by simply changing the order of integration.

In the next step, four theorems are stated which enable formulation of necessary and sufficient conditions for translation invariant kernels.

Theorem 4. *Products of kernels*. Let k_1 and k_2 be admissible SV kernels, then

$$k(x, x') := k_1(x, x')k_2(x, x')$$

is an admissible kernel.

This theorem can be derived by application of the Mercer theorem and Remark 2. Note that by Remark 2, each term in the double sum

$$\sum_{i,j} \lambda_i^1 \lambda_j^2 \psi_i^1(x)\psi_i^1(x')\psi_j^2(x)\psi_j^2(x')$$

gives rise to a positive coefficient.

Theorem 5. *Smola et al. (1998c)*. A translation invariant kernel $k(x, x') = k(x - x')$ is an admissible SV kernel if and only if the Fourier transform

$$F[k](\omega) = (2\pi)^{-\frac{d}{2}} \int_{\mathcal{X}} e^{-i\langle w,x\rangle} k(x)\mathrm{d}x$$

is nonnegative.

The proof of this theorem is discussed in detail in Smola and Schölkopf (2004) and Smola et al. (1998c). It follows from interpolation theory and the theory of regularization networks. For kernels of the form $k(x, x') = k(\langle x, x'\rangle)$, there exist sufficient conditions for admissibility.

Theorem 6. *Burges (1999)*. Any kernel of dot-product type $k(x, x') = k(\langle x, x'\rangle)$ has to satisfy

$$k(\xi) \geq 0,\ \partial_\xi k(\xi) \geq 0 \text{ and } \partial_\xi k(\xi) + \xi \partial_\xi^2 k(\xi) \geq 0$$

for any $\xi \geq 0$ in order to be an admissible SV kernel.

These conditions are necessary, but not sufficient. They can be used in practical applications for checking whether a kernel is an admissible kernel and for construction of new kernels. The general case is given by the theorem of Schoenberg (1942).

Theorem 7. *Schoenberg (1942)*. A kernel of dot-product type $k(x, x') = k(\langle x, x'\rangle)$ defined in an infinite dimensional Hilbert space, with a power series expansion

$$k(t) = \sum_{n=0}^{\infty} a_n t^n$$

is admissible if and only if all $a_n \geq 0$.

A.2.6 The Risk Functional

The following chapters aim to relate the SV algorithm to existing methods of function estimation and the general case of machine learning. This illustration only considers the linear case. Extensions to the nonlinear case are easily possible via application of the kernel methods discussed in the previous section.

Consider the case that the sample data $\mathbf{X} := \{(x_1, y_1), \ldots, (x_\phi, y_\phi)\} \subset \mathcal{X} \times \mathbb{R}$ is drawn i.i.d. from the probability function $P(x, y)$. The goal here is to find a function f that minimizes the expected risk (see Vapnik, 1982)

$$\rho[f] = \int c(x, y, f(x)) \mathrm{d}P(x, y),$$

where $c(x, y, f(x))$ denotes a cost function that determines how estimation errors are penalized. In the case that the exact distribution $P(x, y)$ is known, utilization of $\mathbf{X}$ is sufficient for estimating a function that minimizes $\rho[f]$. The *empirical* risk functional is obtained by replacing the integration with the empirical estimate, i.e.

$$\rho_{\text{emp}}[f] := \frac{1}{\phi} \sum_{i=1}^{\phi} c(x_i, y_i, f(x_i)).$$

One possibility would be to find the empirical risk minimizer

$$f_0 := \text{argmin}_{f \in \mathcal{H}} \rho_{\text{emp}}[f]$$

for some function class $\mathcal{H}$. However, if the capacity of $\mathcal{H}$ is high—i.e. when dealing with few data in high-dimensional spaces—this may yield an overfitting problem. Adding a capacity control term helps to overcome this problem by obtaining the regularized risk functional (Vapnik, 1982, Morozov, 1984, Tikhonov and Arsenin, 1977). In the case of SV, the capacity control term is given by $||w||^2$, which yields the regularized risk function

$$\rho_{\text{reg}}[f] := \rho_{\text{emp}}[f] + \frac{\lambda}{2} ||w||^2. \tag{A.4}$$

The parameter $\lambda > 0$ is called the "regularization" constant. A number of algorithms such as regularization networks or neural networks with weight decay networks minimize expressions similar to the problem of Equation (A.4).

Recall the standard setting of the SV-case, which concerns the ε-insensitive loss

$$c(x, y, f(x)) = |y - f(x)|_\phi. \tag{A.5}$$

Minimizing Equation (A.4) with the loss function of Equation (A.5) is equivalent to minimization of Equation (A.3). In this case, however, it holds that $C = \frac{1}{(\lambda\phi)}$.

Loss functions of the form $|y - f(x)|_\phi^p$ with $p > 1$ may be less desirable as the superlinear increase impairs robustness of the estimator (see Huber, 1981). This argument is intensely discussed in the statistical appendix of this book, where the Generalized Least Squares (GLS) approach is introduced as an OLS-variant to obtain robust covariance estimates in case of heteroscedasticity. In the above case, the superlinear increase forces the derivative of the cost function to grow without bound, whereas for $p < 1$, c becomes nonconvex. In case of $c(x, y, f(x)) = (y - f(x))^2$ it is possible to use the least mean squares fit approach (Smola and Schölkopf, 2004), which yields a matrix inversion rather than a quadratic programming problem.

Under the assumption that the sample is generated by an underlying functional dependency plus noise such as $y_i = f_{\text{true}}(x_i) + \xi_i$ with density $p(\xi)$, it is possible to obtain an optimal cost function in a maximum likelihood sense as

$$c(x, y, f(x)) = -\text{log}p(y - f(x)).$$

Consider an estimate

$$\mathbf{X}_f := \{(x_i, f(x_1)), \ldots, (x_\phi, f(x_\phi))\}$$

for additive noise and i.i.d. data. The likelihood in this case is given by

$$p(\mathbf{X}_f|\mathbf{X}) = \prod_{i=1}^{\phi} p(f(x_i)|(x_i, y_i)) = \prod_{i=1}^{\phi} p(y_i - f(x_i)). \tag{A.6}$$

Note that maximizing $p(\mathbf{X}_f|\mathbf{X})$ is equivalent to minimization of

$$-\text{log}\ p(\mathbf{X}_f|\mathbf{X}).$$

Using this equivalence and the optimal cost function, it follows that

$$-\log p(\mathbf{X}_f|\mathbf{X}) = \sum_{i=1}^{\phi} c(x_i, y_i, f(x_i)). \tag{A.7}$$

A potential problem in this case is that the cost function in Equation (A.7) may be nonconvex. For this reason, it is necessary to find a convex approximation of the cost function, i.e. an efficient implementation of the corresponding optimization problem. For practical applications, it is required to find a cost function as close as possible to the specific function.

A.2.7 Solving the equations

To ensure existence and uniqueness of the optimization problem, assume that $f(x)$ in $c(x, y, f(x))$ is convex for fixed x and y. For reasons of simplicity, let c be symmetric with (at most) two discontinuities at $\pm\varepsilon, \varepsilon \geq 0$ in the first derivative. In addition, c is zero in the interval $[-\varepsilon, \varepsilon]$. This includes, for instance, the ε-insensitive loss function, as well as further commonly used specifications such as the Laplacian, the Gaussian and Huber's robust loss function. It directly follows that c will take the form

$$c(x, y, f(x)) = \tilde{c}(|y - f(x)|_\varepsilon).$$

In the case that the cost function is nonzero in the interval $[-\varepsilon, \varepsilon]$, it is required to use additional slack variables ξ. It is further possible to account for additional discontinuities by including extra Lagrange multipliers in the dual formulation of the optimization problem.

The optimization problem in this case is given as (Smola and Schölkopf, 1998)

$$\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{\phi}(\tilde{c}(\xi_i) + \tilde{c}(\xi_i^*))$$

$$\text{s.t.} \begin{cases} y_i - \langle w, x_i \rangle - b & \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i & \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq 0. \end{cases}$$

Again, the problem can be solved using the standard Lagrange multiplier method. The dual optimization problem becomes

$$\max \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{\phi}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j \rangle \\ +\sum_{i=1}^{\phi} y_i(\alpha_i - \alpha_i^*) - \varepsilon(\alpha_i - \alpha_i^*) \\ +C\sum_{i=1}^{\phi} T(\xi_i) + T(\xi_i^*) \end{cases}$$

$$\text{where} \begin{cases} w & = \sum_{i=1}^{\phi}(\alpha_i - \alpha_i^*)x_i \\ T(\xi) & := \tilde{c}(\xi) - \xi\partial_\varepsilon\tilde{c}(\xi) \end{cases}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{\phi}(\alpha_i - \alpha_i^*)x_i & = 0 \\ \alpha & \leq \partial_\varepsilon\tilde{c}(\xi) \\ \xi & = \inf\{\xi | C\partial_\varepsilon\tilde{c} \geq \alpha\} \\ \alpha, \xi & \geq 0, \end{cases}$$

where the indices are omitted where feasible.

A.2.8 Automatic tuning of the insensitive tube

It has been shown that the performance of a Support Vector machine crucially depends on the choice of the cost function (see, for instance, Smola et al., 1998b and Müller et al., 1997). While the ε-insensitive case brings with it a number of advantages, caution is recommended when utilizing other cost functions, as unless $\varepsilon \neq 0$, the advantage of sparse decomposition is lost. In recent efforts, a number of theoretical papers have addressed this issue and achieved considerable progress in retaining sparsity.

Even when applying the preferred ε-insensitive case, it is required to find a suitable parameter ε to obtain good performance of the SV machine. In the case that the noise model is known, the findings of Smola et al. (1998a) can be used, which state a linear dependency between the noise level and the optimal ε-parameter. A method that is more applicable in practical computations is constructed by Schölkopf et al. (2000). The idea is to adjust the general SV method to include ε as an additional parameter of the optimization problem, thereby allowing for automatic adjustment of ε. The objective function then becomes

$$\min R_v[f] := R_{emp}[f] + \frac{\lambda}{2}||w||^2 + v\varepsilon$$

for some $v > 0$. This yields

$$\min \frac{1}{2}||w||^2 + C\left(\sum_{i=1}^{\phi}(\tilde{c}(\xi_i) + \tilde{c}(\xi_i^*)) + \phi v\varepsilon\right)$$

$$\text{s.t.} \begin{cases} y_i - \langle w, x_i\rangle - b & \leq \varepsilon + \xi_i \\ \langle w, x_i\rangle + b - y_i & \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq 0. \end{cases}$$

In accordance with the optimization problem in the standard case, the dual form of this problem is

$$\max \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{\phi}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \\ +\sum_{i=1}^{\phi} y_i(\alpha_i - \alpha_i^*) \end{cases}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{\phi}(\alpha_i - \alpha_i^*) & = 0, \alpha_i, \alpha_i^* \in [0, C] \\ \sum_{i=1}^{\phi}(\alpha_i - \alpha_i^*) & \leq Cv\phi, \alpha_i, \alpha_i^* \in [0, C]. \end{cases}$$

Note that the target function is even simpler than in the standard case, even if it features an additional constraint. The above formulation bears the additional advantage of enabling pre-specification of the number of SVs.

Theorem 8. *Schölkopf et al. (2000).* Let ν be an upper bound on the fraction of errors and a lower bound on the fraction of SVs. Suppose the data is generated i.i.d. from a distribution $p(x, y) = p(x)p(y|x)$ with a continuous conditional distribution $p(y|x)$. Then it holds that ν asymptotically equals the fraction of SV and the fraction of errors.

The ν-SV regression improves upon ε-SV regression by allowing for automatic adoption of the tube width in accordance with the data. However, this does not apply for the shape of the tube. It is, however, possible to use parametric tube models with non-constant width, as proposed by Schölkopf et al. (2000). Combining ν-SV regression with the asymptotically optimal choice of ε for given noise models provides indication of the way to adjust ν in the case that the class of noise models is known.

Remark 4. *Optimal choice of* ν Let $\mathfrak{p}$ be a probability density with unit variance, and let $\mathfrak{P}$ be a family of noise models generated from $\mathfrak{p}$ by $\mathfrak{P} := \{p|p = \frac{1}{p}\mathfrak{p}\left(\frac{y}{\sigma}\right)\}$. Furthermore, assume that the data is generated i.i.d. from $p(x, y) = p(x)p(y - f(x))$ with continuous $p(y - f(x))$. Then by assumption of uniform convergence, the asymptotically optimal value of ν is

$$\nu = 1 - \int_{-\varepsilon}^{\varepsilon} \mathfrak{p}(t)\mathrm{d}t,$$

where

$$\varepsilon := \mathrm{argmin}_{\tau}(\mathfrak{p}(-\tau) + \mathfrak{p}(\tau))^{-}2 \left(1 - \int_{-\tau}^{\tau} \mathfrak{p}(t)\mathrm{d}t\right).$$

A.2.9 Regularization

Support Vector Machines possess a number of statistically interesting properties, and one of the particularly remarkable characteristics is the close relationship to Regularization Networks (RN). In fact, it can be shown that SV methods are essentially RNs with a particular choice of cost function and that the kernels of this operation are Green's functions of the corresponding regularization operators (see Smola et al., 1998c). The basic concept of Regularization Networks is—similar to the previously described approach—minimization of the regularized risk functional. In this case, optimization of some smoothness criterion lies at the heart of the approach rather than enforcing flatness in the feature space. It follows that

$$R_{\text{reg}}[f] := R_{\text{emp}}[f] + \frac{\lambda}{2}||\mathcal{P}f||^2, \tag{A.8}$$

where $\mathcal{P}$ is a regularization operator (see Tikhonov and Arsenin, 1977), i.e. $\mathcal{P}$ is a positive semidefinite operator mapping from the Hilbert space H of functions f under consideration of a dot product space $\mathfrak{D}$ such that the expression $\langle \mathcal{P}f \times \mathcal{P}g \rangle$ is well-defined for $g, f \in H$. A possible setting may be an operator $\mathcal{P}$ mapping from $L^2(\mathbb{R}^n)$ into some Reproducing Kernel Hilbert Space (RKHS), which is often used in the related literature.

Consider the expansion of f in terms of some symmetric function $k(\mathbf{x}_i, \mathbf{x}_j)$, i.e.

$$f(\mathbf{x}) = \sum_{i=1}^{\phi} \alpha_i k(x_i, x) + b.$$

By utilization of the ε-insensitive cost function, it is possible to illustrate a programming problem similar to the SV case. By application of

$$\mathfrak{D}_{ij} := \langle (\mathcal{P}k)(x_i, .) \times (\mathcal{P}k)(x_j, .) \rangle$$

it is possible to compute $\alpha = \mathfrak{D}^{-}1K(\beta - \beta^*)$, where β, β^* is the solution of

$$\min \begin{cases} \frac{1}{2}(\beta - \beta^*)^T K \mathfrak{D}^{-}1K(\beta - \beta^*) \\ -(\beta - \beta^*)^T y - \varepsilon \sum_{i=1}^{\phi}(\beta + \beta^*) \end{cases}$$

$$\text{s.t.} \sum_{i=1}^{\phi}(\beta - \beta^*) = 0 \text{ and } \beta_i, \beta_i^* \in [0, C].$$

However, this setting of the problem does not guarantee sparsity in terms of the coefficients. The reason is that a potentially sparse decomposition in terms of β_i and β_i^* is spoiled by $\mathfrak{D}^{-1}K$, which is not necessarily diagonal.

A natural question arising in this context is under which conditions the above method coincides with the SV regression case. This question is analogous to the investigation of the conditions under which RN methods might lead to sparse decompositions, i.e. only a few of the expansion coefficients α_i in f would differ from zero. A sufficient condition is

$$\mathfrak{D} = K \text{ and thus } K\mathfrak{D}^{-1}K = K.$$

By employing the concept of Green's functions (Girosi et al., 1993), it is possible to solve two problems. First, it makes it possible to find a kernel for a regularized operator $\mathcal{P}$ such that the SV machine not only enforces flatness but also corresponds to minimizing a regularized risk functional with $\mathcal{P}$ as regularizer. Second, given a kernel k, Green's functions allow for computation of a regularized operator $\mathcal{P}$ such that a SVM based on this kernel can be viewed as a RN.

Initially, Green's functions were introduced for the purpose of solving differential equations. In the context of SV regression, the advantage is that Green's functions $\mathfrak{G}_{x_i}(x)$ of $\mathcal{P} \times \mathcal{P}$ satisfy

$$(\mathcal{P} \times \mathcal{P}\mathfrak{G}_{x_i})(x) = \delta_{x_i}(x),$$

where $\delta_{x_i}(x)$ is the Delta-distribution with property $\langle f \times \delta_{x_i} \rangle = f(x_i)$.

Proposition 3. *Smola et al. (1998b)* Let $\mathfrak{p}$ be a regularization operator, and $\mathfrak{G}$ be the Green's function of $\mathcal{P} \times \mathcal{P}$. Then $\mathfrak{G}$ is a Mercer kernel such that $\mathfrak{D} = K$. SV machines using $\mathfrak{G}$ minimize risk functional (A.8) with regularization operator $\mathcal{P}$.

This condition can be used to compute Green's functions for a given regularization operator $\mathcal{P}$ and to infer the regularizer, given a kernel k.

Consider regularization operators $\mathcal{P}$ that may be written as multiplicators in Fourier space

$$\langle \mathcal{P}f \times \mathcal{P}g \rangle = \frac{1}{(2\pi)^{n/2}} \int_\Omega \frac{\overline{\tilde{f}(\omega)}\tilde{g}(\omega)}{\mathcal{P}(\omega)} \mathrm{d}\omega, \tag{A.9}$$

where $\tilde{f}(\omega)$ is the Fourier transform of $f(x)$. In addition, $P(\omega) = P(-\omega)$ is real valued, nonnegative and converging to 0 for $|\omega| \to \infty$ and $\Omega := \mathrm{supp}[\mathcal{P}(\omega)]$. Note that small values of $\mathcal{P}(\omega)$ for large ω are desirable since high frequency components of $\tilde{f}$ correspond to rapid changes in f. $\mathcal{P}(\omega)$ illustrates the filter properties of $\mathcal{P} \times \mathcal{P}$.

For regularization operators defined by Equation (A.9) in the Fourier space, exploitation of $\mathcal{P}(\omega) = \mathcal{P}(-\omega) = \overline{\mathcal{P}(\omega)}$ yields

$$\mathfrak{G}(x_i, x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp\{i\omega(x_i - x)\}\mathcal{P}(\omega)\mathrm{d}\omega,$$

which is a Green's function satisfying translational invariance. This is

$$\mathfrak{G}(x_i, x_j) = \mathfrak{G}(x_i - x_j) \text{ and } \tilde{\mathfrak{G}}(\omega) = \mathcal{P}(\omega)$$

and provides an efficient tool for analyzing SV kernels. The above can also be understood as a special case of Bochner's theorem (see Smola and Schölkopf, 2004), which states that the Fourier transform of positive measures constitutes a positive Hilbert Schmidt Kernel.

In the Support Vector approach of this book, the Gaussian RBF kernel is particularly important. For this reason, consider the example described in Girosi et al. (1993). For

$$||\mathcal{P}f||^2 = \int \mathrm{d}x \sum_m \frac{\sigma^{2m}}{m!2^m}(\hat{O}^m f(x))^2$$

with $\hat{O}^{2m} = \Delta^m$, $\hat{O}^{2m+1} = \nabla\Delta^m$, Δ denoting the Laplacian, and ∇ the Gradient operator, the resulting kernel is a Gaussian kernel. In addition, it is possible to achieve an equivalent representation of $\mathcal{P}$ in terms of its Fourier properties. This means

$$\mathcal{P}(\omega) = \exp\left\{-\frac{\sigma^2||\omega||^2}{2}\right\} \tag{A.10}$$

up to a multiplicative constant.

When programming a SVM with a Gaussian RBF kernel—such as is used in the applications of this book—this exercise corresponds to minimizing the specific cost function with a regularization operator of the above type. Note that Equation (A.10) enables computations of very smooth estimates. SV machines with Gaussian RBF perform particularly well, as Equation (A.10) implies that all derivatives of f are penalized. Without the RN formulation of this problem, the favorable performance of Gaussian RBF based SVM is difficult to describe, as it is by no means obvious that choosing a flat function in some high dimensional space will correspond to a simple function in low dimensional space (see Smola and Schölkopf, 2004). A similar case has been shown by Smola et al. (1998c) for Dirichlet kernels.

A.3 Support Vector Machines for classification

A.3.1 Motivation

In the next step, this appendix illustrates the SV algorithm used for classification purposes. Both SV regression and SV classification play important roles in the computation of democracy indices in this book. This section builds on the previous illustrations of SV for function estimation, particularly the described theorems and optimization methods. However, due to its important position in the SVMDI algorithm, it may be advantageous to describe the underlying concepts and the differences to SV used for regressions in greater detail. As in Chapter A.2, the inner product of vectors will play an important role in the computation of functions and hyperplanes. These products, however, assume more complex forms with respect to their notation. The chapter thus deviates from the conventional notation used in the previous section for reasons of lucidity, where the inner product of a and b is henceforth denoted by $(a \times b)$.

The general idea of pattern recognition

Developed in the late 1950s, the problem of pattern recognition can be described as follows (see Vapnik, 1998): A supervisor observes a number of situations and decides to which of k classes each one of them belongs. It is required to construct a machine that carries out the classification in a manner approximately equivalent to the classification of the supervisor. More formally, consider an environment characterized by a probability distribution function $\mathcal{F}(x)$ with situations x occurring randomly and independently. The supervisor classifies each x into one of k classes using the conditional probability distribution function $\mathcal{F}(\omega|x)$, where $\omega \in \{0, 1, \ldots, k-1\}$. Neither the decision rule, nor the properties of the environment $\mathcal{F}(\omega|x)$ are known. Yet as both functions exist, there is a joint distribution $\mathcal{F}(\omega, x) = \mathcal{F}(\omega|x)\mathcal{F}(x)$. Suppose that there is a set of functions $\phi(x, \alpha), \alpha \in \Lambda$ which assume values ω. Consider the loss function

$$\mathcal{L}(\omega, \phi) \begin{cases} 0 & \text{if } \omega = \phi \\ 1 & \text{if } \omega \neq \phi. \end{cases}$$

Pattern recognition is the problem of minimizing the functional

$$R(\alpha) = \int \mathcal{L}(\omega, \phi(x, \alpha)) \mathrm{d}\mathcal{F}(\omega, x) \tag{A.11}$$

on the set of functions $\phi(x, \alpha), \alpha \in \Lambda$ with unknown distribution and independent sample pairs $(\omega_1, x_1), \ldots, (\omega_l, x_l)$ given.

For the above loss function, the functional in Equation (A.11) determines the probability of a classification error for each given $\phi(x, \alpha)$. For this reason, the problem to be solved is minimization of the probability of classification errors when the data is given, but the probability distribution function $\mathcal{F}(\omega, x)$ is unknown.

To characterize the pattern recognition problem, consider the two-class classification problem $\omega \in \{0, 1\}$ with loss function $\mathcal{L}(\omega, \phi)$. This is minimization of the risk based on empirical data. The difference between this setting and the more general case of minimization of risk functionals based on loss functions $Q(z, \alpha^*), \alpha \in \Lambda$ is that some restrictions are imposed. In the general case, let Z be a subset of vector space $\mathbb{R}^n$, $\{g(z)\}, z \in Z$ denote a set of admissible functions, and $R = R(g(z))$. The general problem is minimization of

$$R(g(z)) = \int \mathcal{L}(z, g(z)) \mathrm{d}\mathcal{F}(z),$$

where the set of functionals $g(z)$ will be given in a parametric form, i.e. $\{g(z, \alpha), \alpha \in \Lambda\}$. This yields the functional

$$R(\alpha) = \int Q(z, \alpha) \mathrm{d}\mathcal{F}(z), \alpha \in \Lambda,$$

where $Q(z, \alpha) = \mathcal{L}(z, g(z, \alpha))$. The general case of empirical risk minimization (ERM) is illustrated in Section (A.1.2).

Considering the problem of pattern recognition, the vector z consists of $n + 1$ coordinates: coordinate ω that assumes a finite number of values (in the case of the simplest problem considered above, the number of possible values is 2), and n coordinates $x^1, \dots, x^n$, forming the vector x. Likewise, the set of functionals $Q(z, \alpha), \alpha \in \Lambda$ is $Q(z, \alpha) = \mathcal{L}(\omega, \phi(x, \alpha)), \alpha \in \Lambda$. These specific features of the risk minimization problem characterize the pattern recognition problem. As it deals with the simplest loss function, the problem of pattern recognition forms the simplest learning problem (Vapnik, 1998). This loss function describes a set of indicator functions, i.e. functions that take only two possible values.

A.3.2 Rosenblatt's Perceptron

First, consider classical learning algorithms: perceptrons, radial basis functions, and neural networks. The basis is a machine proposed by Rosenblatt (1958), the Perceptron. Invented at the Cornell Aeronautical Laboratory, the Perceptron was designed as an algorithm for supervised learning of binary classifiers which allows for online-learning, i.e. it allows for procession of its training elements piece-by-piece. The algorithm was used in numerous practical implementations including software (e.g. in IBM 704") and hardware (e.g. in the "Mark 1 Perceptron"). The cornerstone is a set of indicator functions linear in their parameters

$$f(x, w) = \text{sign}\left\{\sum_{p=1}^{n} w^p \psi_p(x)\right\}, \tag{A.12}$$

where for reasons of simplicity the subsequent illustration considers indicator functions of the form sign$(u) \in \{-1, 1\}$ instead of the form $\theta(u) \in \{0, 1\}$. Rosenblatt (1958) suggested a procedure for computing an approximation function based on a sequence of examples $(y_1, x1), \dots, (y_l, x_l)$ with $y_i = 1$ if the vector x_i belongs to the first class, and $y_i = -1$ otherwise.

The recurrent procedure for choosing the function in Equation (A.12) with coefficients $w = (w^1, \dots, w^n)$ is as follows. First, the procedure uses $f(x, \mathbf{0})$ with $w(0) = (0, \dots, 0)$. In the subsequent steps t, the algorithm uses the elements (y_t, x_t) of the training sequence to change the vector of coefficients $w(t - 1)$ according to the rule

$$w(t) \begin{cases} w(t-1) & \text{if } y_t(w(t-1) \times \Psi_t) > 0, \\ w(t-1) + y_t \Psi_t & \text{if } y_t(w(t-1) \times \Psi_t) \leq 0. \end{cases}$$

In this case $\Psi_t = (\psi_1(x_t), \dots, \psi_n(x_t))$ denotes an n-dimensional vector and $(w(t-1) \times \Psi_t)$ is the inner product of two vectors.

Assume existence of a non-linear operator A that maps the vectors $x \in X$ into vector $u \in U$. The Perceptron constructs a separating hyperplane

$$f(u, w) = \text{sign}\{(u \times w)\}$$

passing through the origin in U space, which is called the *feature space*. The general idea is that the problem of computing the nonlinear decision rule in the space X can be reduced to constructing a separating hyperplane in the space U. Estimation of the unknown variables in this space follows

$$w(t) \begin{cases} w(t-1) & \text{if } y_t(w(t-1) \times u_t) > 0, \\ w(t-1) + y_t u_t & \text{if } y_t(w(t-1) \times u_t) \leq 0. \end{cases}$$

The theorems concerning the Perceptron Algorithm, which were initially developed in the early 1960s, originated the mathematical field of learning theory. Probably the most important of these theorems was proposed by Novikoff (1962), concerning the crucial issue of convergence of the algorithm.

Suppose that there is an infinite sequence of examples in feature space

$$\{\hat{Y}, \hat{U}\} = (y_1, u_1), \ldots, (y_l, u_l), \ldots$$

and a vector w_0 for which the inequality

$$\min_{(y,U) \in \{\hat{Y}, \hat{U}\}} \frac{y(w_0 \times u)}{|w_0|} \geq \rho_0 \tag{A.13}$$

holds true for some $\rho_0 > 0$. Then the Novikoff-Theorem can be stated as follows.

Theorem 9. *Novikoff (1962)* Assume existence of an infinite sequence of training examples $\{\hat{Y}, \hat{U}\}$ with elements satisfying $|u_i| < D$. Suppose further that there exists a hyperplane with coefficients w_0 that correctly separates the elements of this sequence and satisfies inequality A.13. Then the iterative procedure illustrated above yields the construction of a hyperplane that correctly separates all elements. For construction of this hyperplane, the Perceptron makes at least

$$M = \left[\frac{D^2}{\rho_0^2} \right]$$

corrections.

A comprehensive discussion of the Perceptron algorithm can be found in Vapnik (1998).

A.3.3 Method of potential functions and radial basis functions

Another approach was proposed in the mid 1960s, relying on potential functions and radial basis functions. This method was initially published by Aizerman et al.

(1964b,a). The idea introduced in these articles proposes estimation of the functional dependency from the data

$$(y_1, x_1), \ldots, (y_l, x_l)$$

using the set of functions

$$f(x,\alpha) = \text{sign}\left\{\sum_{i=1}^{l} \alpha_i \phi(|x - x_i|)\right\}\alpha, \tag{A.14}$$

where $\lim_{u\to\infty}\phi(|u|) = 0$. The function $\phi(|u|)$ is called a *potential function*, which receives its name via analogy with physical potentials. This function is monotonic and converges to zero with increasing $|u|$.[124]

With the help of the function illustrated in Equation (A.14), the vector x^* is classified as belonging to the first class if $f(x^*, \alpha) > 0$, and to the second class otherwise. Similar to the Perceptron discussed in the previous section, potential functions are online algorithms. After developing a theory of consistency in the 1970s, an offline variant of the method was intensely discussed in the 1980s. This offline variant uses functions that are called *radial basis functions* (RBFs). As in a number of techniques discussed in this appendix, RBF approximations use methods that minimize the empirical risk functional

$$R_{emp}(\alpha) = \sum_{i=1}^{L}\left(y_i - \sum_{j=1}^{l}\alpha_j \phi(|x_i - x_j|)\right)^2. \tag{A.15}$$

Under certain conditions, the matrix $\mathbf{K} = ||k_{ij}||$, consisting of elements $k_{ij} = \phi(|x_i - x_j|)$, is positive definite, resulting in Equation (A.15) having a unique solution.

The method of RBF was later adjusted using kernels that were not defined at every point but rather at some specific points of the training center, which are called *centers* $c_{ij}, j = 1, \ldots, N$. The empirical risk functional in this case becomes

$$R_{emp}(\alpha) = \sum_{i=1}^{L}\left(y_i - \sum_{j=1}^{l}\alpha_j \phi(|x_i - c_j|)\right)^2. \tag{A.16}$$

Several heuristics exist for specifying the number of centers N and their positions. Rather than illustrating these heuristics in detail, this appendix focuses its discussion on Support Vector Machines to construct an optimal separating hyperplane in the feature space.

A.3.4 Neural networks

The Perceptron, potential functions, and RBFs construct learning machines that non-linearly map the input vectors x into the feature space U and compute a linear function in this space. Support Vector Machines make this approach particularly

[124] For instance, consider $\phi(|u|) = \exp\{-\gamma|u|\}$.

attractive, which is why this appendix describes the SV method in greater detail in the next chapter. There is, however, an entirely different learning machine that is inspired by the neuro-physiological analogy. This approach—called *neural networks*—considers machines that are defined by a superposition of several neurons. This structure, having n inputs and one output, is defined by several neurons, each with own weight. Learning in these networks coincides with estimation of the coefficients of all neurons. Such an estimation requires calculation of the gradient loss function for neural networks, which can be achieved with the help of the *back-propagation method*. A description of this method can be found in LeCun (1986) and Rumelhart et al. (1986).

A.3.5 Formulating the classification problem in terms of Support Vectors

The previously described methods emphasize that the estimation of separating hyperplanes plays an important role in the construction of learning algorithms. By using Support Vector machines, it is possible to calculate optimal separating hyperplanes that turn out to have remarkable statistical properties. This particularly applies to Bellmans curse of dimensionality (Bellman, 1961), a significant impediment to machine learning methods which are required to learn structures of higher-dimensional spaces based on few data points. In fact, the dimensionality problem seems to be countermanded by SV applications, as the feature map makes problems appear easier and at the same time more reliable by transformation of the data into the feature space.

Introduce the set of training data

$$(y_1, x_1), \ldots, (y_\phi, x_\phi),\ x \in \mathbb{R}^n, y \in \{-1, 1\},$$

and define the subset S_1 with elements $y = 1$ and subset S_2 consisting of the elements $y = -1$. These subsets are separable by the hyperplane

$$(x \times \varphi) = c,$$

if there exists a constant c and a unit vector φ ($|\varphi| = 1$) such that

$$\begin{aligned} (x_i \times \varphi) &> c, \quad \text{if } x_i \in S_1 \\ (x_j \times \varphi) &< c, \quad \text{if } x_j \in S_2 \end{aligned}$$

holds true. Subsequently, these inequalities are referred to as the $(x \times \varphi)$-inequalities. Note that $(a \times b)$ denotes the inner product between vectors a and b. Consider two values for each unit vector φ as

$$\begin{aligned} c_1(\varphi) &= \min_{x_1 \in S_1} (x_i \times \varphi) \\ c_2(\varphi) &= \max_{x_j \in S_2} (x_j \times \varphi) \end{aligned}$$

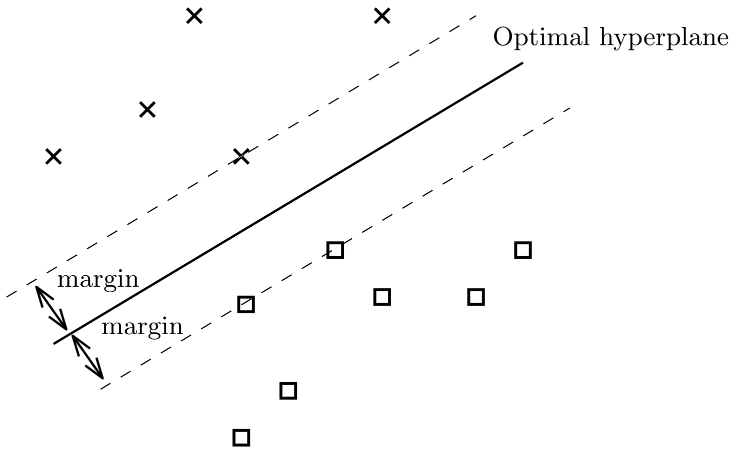


Figure A.2 The optimal hyperplane separates the data with the maximal margin (see Vapnik, 1998).

and a special unit vector φ_0 that satisfies the $(x \times \varphi)$-inequalities and maximizes the function

$$\rho(\varphi) = \frac{c_1(\varphi) - c_2(\varphi)}{2}, \ |\varphi| = 1.$$

In addition, consider the constant

$$c_0 = \frac{c_1(\varphi_0) + c_2(\varphi_0)}{2},$$

which together with ϱ_0 determines the hyperplane that separates the data points $(x_1, \ldots, x_a) \in S_1$ from $(x_1, \ldots, x_b) \in S_2$ and possesses the maximal margin $\rho(\varphi)$. This hyperplane is called *optimal hyperplane* or *maximal margin hyperplane*. This hyperplane is illustrated in Figure (A.2).

In principle, there exist numerous possible hyperplanes that separate the data. However, as Figure (A.2) shows, the optimal hyperplane in the context of machine learning and Support Vector classification is the hyperplane that has the maximal margin.

Two theorems are particularly decisive for the Support Vector estimation of indicator functions. The first theorem states uniqueness of the hyperplane with maximal margin.

Theorem 10. *Uniqueness of the optimal hyperplane.* The optimal hyperplane is unique.

Proof: It needs to be shown that the maximum point φ_0 of the function $\rho(\varphi)$ in the area $|\varphi| \leq 1$ exists and is reached at the boundary, i.e. at $|\varphi| = 1$. The existence of that maximum follows from the continuity of $\rho(\varphi)$ in $|\varphi| \leq 1$. Assume that this maximum would be achieved at some interior point φ^*. Then the vector

$$\varphi^* = \frac{\varphi_0}{|\varphi_0|}$$

would define a larger margin

$$\rho(\varphi^*) = \frac{\rho(\varphi_0)}{|\varphi_0|}.$$

As the function $\rho(\varphi_0)$ is convex, its maximum cannot be achieved on (two) boundary points. It is further impossible to achieve the maximum on the line that connects the points, which would be impossible given the preceding argument.

To find effective methods for constructing the optimal hyperplane, consider the equivalent statement of this problem: Find a pair consisting of a vector ψ_0 and a constant b_0 such that they satisfy

$$\begin{aligned} &(x_i \times \psi) + b_0 \geq 1, && \text{if } y_i = 1, \\ &(x_j \times \psi) + b_0 \leq -1, && \text{if } y_j = -1, \end{aligned} \tag{A.17}$$

and the vector ψ_0 has the smallest norm

$$|\psi|^2 = (\psi). \tag{A.18}$$

Theorem 11. *Margin of the optimal hyperplane.* The vector ψ_0 that minimizes the norm in Equation (A.18) by satisfying the constraints in (A.17) is related to the vector that forms the optimal hyperplane by equality $\varphi_0 = \frac{\psi_0}{|\psi_0|}$. The margin ρ_0 between the optimal hyperplane and the separated vectors is equal to

$$\rho(\varphi_0) = \sup_{\varphi_0} \frac{1}{2} \left(\min_{i \in S_1} (x_i \times \varphi_0) - \max_{j \in S_2} (x_j \times \varphi_0) \right) = \frac{1}{|\psi_0|}. \tag{A.19}$$

Proof: The proof can be found in Vapnik (1998).

It follows that the vector ψ_0 with the smallest norm that satisfies the constraints in (A.17) defines the optimal hyperplane (where in the special case of $b = 0$, the optimal hyperplane passes through the origin). The constraints in (A.17) can be expressed in the equivalent form to achieve simplification of the notation. The equivalent form is given by

$$y_i\{(x_i \times \psi_0) + b\} \geq 1, i = 1, \ldots, \phi. \tag{A.20}$$

Finding the optimal hyperplane is thus equivalent to minimization of the quadratic form expressed in Equation (A.18) with respect to the constraints in (A.20). This problem can be solved in the "original" *primal* space or the *dual* space, the space of Lagrange multipliers. As shown in the previous chapters of this appendix, utilization of the dual space brings with it a number of advantages. It is easy to show that the quadratic optimization problem can be solved by computation of the saddle point of the Lagrange function (see, for instance, Vapnik, 1998 and the descriptions in Chapter A.2.3)

$$L(\psi, b, \alpha) = \frac{1}{2}(\psi \times \psi) - \sum_{i=1}^{\phi} \alpha_i \{y_i((x_i \times \psi) + b)\}, \tag{A.21}$$

where $\alpha_i \geq 0$ are the Lagrange multipliers. The saddle point can be found by minimizing the Lagrange function over ψ and b, and maximizing $L(\cdot)$ over the nonnegative Lagrange multipliers. Application of the Fermat theorem implies that the minimum points of $L(\cdot)$ satisfy

$$\frac{\partial L(\psi, b, \alpha)}{\partial \psi} = \psi - \sum_{i=1}^{\phi} y_i \alpha_i x_i = 0, \frac{\partial L(\psi, b, \alpha)}{\partial b} = \sum_{i=1}^{\phi} y_i \alpha_i = 0.$$

These conditions imply that for ψ, the identities

$$\psi = \sum_{i=1}^{\phi} y_i \alpha_i x_i \tag{A.22}$$

and

$$\sum_{i=1}^{\phi} y_i \alpha_i = 0 \tag{A.23}$$

hold true. Substituting Equation (A.22) into the Lagrange function of Equation (A.21) and considering Equation (A.23), the function

$$W(\alpha) = \sum_{i=1}^{\phi} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\phi} y_i y_j \alpha_i \alpha_j (x_i \times x_j) \tag{A.24}$$

can be obtained. The notation $W(\alpha)$ is used to distinguish the function obtained by the last transformation from the "original" Lagrange function. Construction of the optimal hyperplane requires computation of α_i^0 that maximize $W(\alpha)$ in the nonnegative quadrant $\alpha_i \geq 0, i = 1, \ldots, \phi$ under Equation (A.23). Utilization of the attained coefficients yields the solution

$$\psi_0 = \sum_{i=1}^{\phi} y_i \alpha_i^0 x_i.$$

Computation of b_0 can be achieved by maximization of the margin $\rho(\phi)$. The optimal solutions ψ_0 and b_0 are required to satisfy the conditions of Kuhn and Tucker (1951), which in this case are given by

$$\alpha_i^0\{y_i((x_i \times \psi_0) + b_0) - 1\} = 0, i = 1, \ldots, \phi.$$

From these conditions, it can be concluded that nonzero values α_i^0 correspond only to the vectors x_i that satisfy the equality

$$y_i\{(x_i \times \psi_0) + b_0\} = 1.$$

Geometrically, these vectors are located closest to the optimal hyperplane, as illustrated in Figure (A.2). Analogous to the case of function estimation, these vectors are called *Support Vectors*. These vectors are crucial in constructing learning algorithms, as the vector ψ_0 that defines the optimal hyperplane is expanded with nonzero weight on Support Vectors, i.e.

$$\psi_0 = \sum_{i=1}^{\phi} y_i \alpha_i^0 x_i,$$

which yields the following form of the optimal hyperplane

$$f(x, \alpha_0) = \sum_{i=1}^{\phi} y_i \alpha_i^0 (x_s \times x) + b_0, \tag{A.25}$$

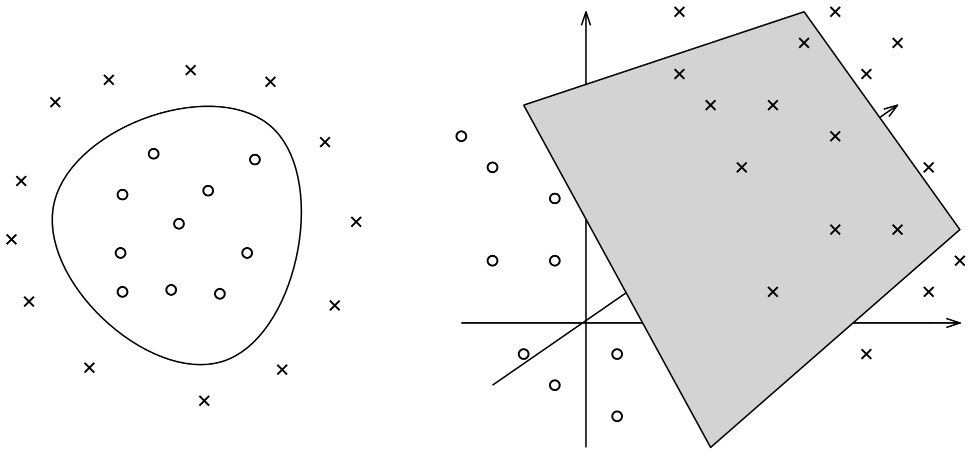
where the term $(x_s \times x)$ denotes the dot product of the two vectors. Both the hyperplane in Equation (A.25) and the resulting objective function

$$W(\alpha)) \sum_{i=1}^{\phi} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\phi} y_i y_j \alpha_i \alpha_j (x_i \times x_j) \tag{A.26}$$

do not explicitly depend on the dimensionality of the vector x, but only on the inner product of the two vectors. This property is highly advantageous, as it allows for construction of a separable hyperplane in spaces of higher dimensions, particularly in infinite-dimensional Hilbert spaces.

A.3.6 The construction of Support Vector Machines for estimation of indicator functions

The final step is to transfer the Support Vector procedure to the problem of estimation of indicator functions. The general idea of the Support Vector Machine (SVM) can be formulated analogously to the problem of function estimation as discussed in Chapter (A.2.5). The SVM maps the input vectors x into the high-dimensional *feature space* $\mathcal{F}$ via application of a nonlinear map, which is chosen a priori. In this space, the optimal hyperplane is constructed. As described in Chapter (A.2.9), the generalization ability of the constructed hyperplane is high, even if the feature space has a high dimensionality.[125] One advantage of the shift of the data into the higher-dimensional



Second dimension Third dimension

Figure A.3 Classification problems are often easier to solve in higher dimensions. Depicted are the second and third dimensional case.

space is that classification problems become simpler in higher dimensions. Figure (A.3) provides an intuitive illustration of the resulting simplification.

Whereas the optimal hyperplane can theoretically be found and under certain conditions generalizes well, the crucial question in this context is that of how to treat the higher dimensional space. As in the case of function estimation, it is not necessary to consider the *exact* form of $\mathcal{F}$, as it is sufficient for computation of the hyperplane to rely on the Support Vectors. Analogously to the case of function estimation, the *kernel trick* can be used, which is based on a special property of the Hilbert space. Let $k \in \mathbb{R}^n$ be a vector that is mapped into the Hilbert space with coordinates $\zeta_1(g), \ldots, \zeta_n(g), \ldots$. The Hilbert-Schmidt theory implies that $(\zeta_1 \times \zeta_2)$ has the following equivalent representation

$$(\zeta_1 \times \zeta_2) = \sum_{r=1}^{\infty} \kappa_r \zeta_r(g_1)\zeta(g_2) \Leftrightarrow k(g_1, g_2), \kappa_r \geq 0, \tag{A.27}$$

where $k(g_1, g_2)$ is a symmetric function that satisfies the theorems discussed in Chapter (A.2.5), particularly the theorem of Mercer (1909). This yields an admissible kernel identically to the case described previously. The remarkable feature of the structure of inner products in the Hilbert space is that for any kernel function $k(\cdot)$ that satisfies the theorem of Mercer (1909), there exists a feature space where the kernel function generates the inner product stated by the Mercer theorem.

Based on the inner product in the higher-dimensional space, a nonlinear decision function can be constructed via

[125] For the exact conditions under which this holds true, see Vapnik (1998).

$$\Psi(x, \alpha) = \text{sign}\left(\sum_{a=1}^{A} y_i \alpha_i^0 k(x, x_i) + b\right), \tag{A.28}$$

where A is the cardinal number of the set of Support Vectors with numbers $a = 1, \ldots, A$. This function is equivalent to linear decision functions in the feature space $\zeta_1(x), \ldots, \zeta_k(x), \ldots$

$$\Psi(x, \alpha) = \text{sign}\left(\sum_{a=1}^{A} y_i \alpha_i^0 \sum_{r=1}^{\infty} \zeta_r(x_i)\zeta_r(x) + b\right). \tag{A.29}$$

In this case, $k(\cdot)$ is the kernel that generates the inner product for this feature space. To construct the decision function of Equation (A.28), the previously discussed methods can be used, where the inner product (x, x_i) is replaced by the inner product defined by the kernel $k(x, x_i)$.

The Support Vector machine then follows three steps. First, it is necessary to compute the maximum of

$$W(\alpha) = \sum_{i=1}^{\phi} \alpha_i - \frac{1}{2}\sum_{i,j}^{\phi} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \tag{A.30}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{\phi} \alpha_i y_i & = 0 \\ \alpha_i & > 0 \end{cases}$$

to find the coefficients α_i in the separable case $y_i f(x_i, \alpha) = 1$. In the second step, it is sufficient to maximize the problem of Equation (A.30) under the constraints $\sum_{i=1}^{\phi} \alpha_i y_i = 0$ and $0 \geq \alpha_i \leq C$ to find the optimal soft margin solution or the nonseparable case.[126]

Finally, the optimal solution for a given margin $\rho = \frac{1}{Q}$

$$\Psi(x, \alpha) = \text{sign}\left(\frac{Q}{\sqrt{\sum_{i,j=1}^{\phi} \alpha_i^0 \alpha_j^0 y_i y_j k(x_i, x_j)}} \sum_{i=1}^{\phi} \alpha_i^0 y_i k(x_i, x) + b\right)$$

can be computed by maximization of the functional

$$W(\alpha) = \sum_{i=1}^{\phi} \alpha_i - Q\sqrt{\sum_{i,j}^{\phi} \alpha_i \alpha_j y_i y_j k(x_i, x_j)} \tag{A.31}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{\phi} \alpha_i y_i = 0 \\ q \leq \alpha_i \leq 1. \end{cases}$$

[126] The nonseparable case is not explicitly covered in this appendix. The method relies on the soft margin, which strongly resembles the referring approach described in Chapter (A.2). A detailed description of the topic can be found in Vapnik (1998).

The learning machines that construct these decision functions are called *Support Vector Machines*. Based on the techniques illustrated in this section, it is possible to derive an algorithm that qualifies objects with a particular value continuous on the $(0, 1)$ interval. In Chapter (6), this procedure is used to check robustness of the classification of country-years conducted in the second step of the SVMDI algorithm.

Bibliography

Acemoglu, D., Aghion, P., and Zilibotti, F. (2006). Distance to frontier, selection, and economic growth. *Journal of the European Economic Association*, 4(1):37–74.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics*, 117(4):1231–1294.

Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and democracy. *American Economic Review*, 98(3):808–842.

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

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2013). Democracy, redistribution and inequality. NBER Working Paper 19746, National Bureau of Economic Research.

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2014). Democracy does cause growth. NBER Working Paper 20004, National Bureau of Economic Research.

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2015). Chapter 21 - democracy, redistribution, and inequality. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution*, pages 1885–1966. Elsevier.

Acemoglu, D. and Robinson, J. A. (2012). *Why Nations Fail*. Crown, New York (US).

Acs, Z. J., Audretsch, D. B., Braunerhjelm, P., and Carlsson, B. (2004). The missing link: The knowledge filter and entrepreneurship in endogenous growth. Working Paper 805, Max Planck Institute for Research into Economic Systems Group Entrepreneurship, Growth and Public Policy.

Acs, Z. J., Audretsch, D. B., Braunerhjelm, P., and Carlsson, B. (2006). Growth and Entrepreneurship: An Empirial Assessment. CEPR Discussion Paper no.5409, Centre for Economic Policy Research.

Afonso, A. and Jalles, J. T. (2014). Fiscal composition and long-run growth. *Applied Economics*, 46(3):349–358.

African Election Database (2014). *A database of election results in Sub-Saharan Africa*. African Election Database.

Aghion, P., Angeletos, G.-M., Banerjee, A., and Manova, K. (2010). Volatility and growth: Credit constraints and the composition of investment. *Journal of Monetary Economics*, 57(3):246–265.

Aghion, P., Caroli, E., and Garcia-Penalosa, C. (1999). Inequality and economic growth: The perspective of the new growth theories. *Journal of Economic Literature*, 37(4):1615–1660.

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60(2):323–351.

Aghion, P. and Howitt, P. (1998). *Endogenous Growth Theory*. MIT Press, Cambridge (US) and London (UK).

Aghion, P. and Howitt, P. (2009). *The Economics of Growth*. MIT Press, Cambridge (US) and London (UK).

Aghion, P., Howitt, P., and Mayer-Foulkes, D. (2005). The effect of financial development on convergence: Theory and evidence. *The Quarterly Journal of Economics*, 120(1):173–222.

Aizenman, J., Jinjarak, Y., and Park, D. (2015). Financial development and output growth in Developing Asia and Latin America: A comparative sectoral analysis. NBER Working Paper No.20917, National Bureau of Economic Research.

Aizerman, M., Braverman, É., and Rozonoér, L. (1964a). The problem of pattern recognition learning and the method of potential functions. *Automation and Remote Control*, 25:1175–1193.

Aizerman, M., Braverman, E., and Rozonoer, L. I. (1964b). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837.

Alesina, A. and Giuliano, P. (2015). Culture and institutions. *Journal of Economic Literature*, 53(4):898–944.

Alesina, A., Özler, S., Roubini, N., and Swagel, P. (1996). Political instability and economic growth. *Journal of Economic Growth*, 1(2):189–211.

Alesina, A. and Perotti, R. (1996). Income distribution, political instability, and investment. *European Economic Review*, 40(6):1203–1228.

Alesina, A. and Rodrik, D. (1994). Distributive politics and economic growth. *The Quarterly Journal of Economics*, 109(2):465–490.

Allen, F. and Gale, D. (2000). *Comparing Financial Systems*. MIT Press.

Almunia, M., Bénétrix, A., Eichengreen, B., O'Rourke, K. H., and Rua, G. (2010). From great depression to great credit crisis: similarities, differences and lessons. *Economic Policy*, 25(62):219–265.

Alonso-Atienza, F., Rojo-Álvarez, J. L., Rosado-Muñoz, A., Vinagre, J. J., García-Alberola, A., and Camps-Valls, G. (2012). Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection. *Expert Systems with Applications*, 39(2):1956–1967.

Anderson, L. E. and Dodd, L. C. (2009). Nicaragua: Progress amid regress? *Journal of Democracy*, 20(3):153–167.

Anderson, T. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, 20:46–63.

Anderson, T. W. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1):47–82.

Ang, J. B. (2011). Financial development, liberalization and technological deepening. *European Economic Review*, 55(5):688–701.

Ang, J. B. and Madsen, J. B. (2012). Can second-generation endogenous growth models explain the productivity trends and knowledge production in the asian miracle economies? *Review of Economics and Statistics*, 93(4):1360–1373.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton (US).

Apolte, T. (2011). Democracy and prosperity in two decades of transition. *The Economics of Transition*, 19(4):693–722.

Arcand, J.-L., Berkes, E., and Panizza, U. (2012). Too Much Finance? IMF Working Paper No.12/161, International Monetary Fund.

Arcand, J.-L., Berkes, E., and Panizza, U. (2015). Too much finance? *Journal of Economic Growth*, 20(2):105–148.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297.

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

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In Nelson, R. R., editor, *The Rate and Direction of Inventive Activity: Economic and Social Factors*, NBER Chapters, pages 609–626. National Bureau of Economic Research, Cambridge (US).

Ashraf, Q. and Galor, O. (2011). Dynamics and stagnation in the malthusian epoch. *American Economic Review*, 101(5):2003–2041.

Atkeson, A. and Kehoe, P. J. (2000). Paths of development for early-and late-bloomers in a dynamic heckscher-ohlin model. Staff Report 256, Federal Reserve Bank of Minneapolis, Research Department.

Atkinson, A. B. and Brandolini, A. (2001). Promise and pitfalls in the use of secondary data-sets: Income inequality in oecd countries as a case study. *Journal of Economic Literature*, 39(3):771–799.

Atkinson, A. B. and Brandolini, A. (2009). On data: A case study of the evolution of income inequality across time and across countries. *Cambridge Journal of Economics*, 33(3):381–404.

Audretsch, D. B. (2007). Entrepreneurship capital and economic growth. *Oxford Review of Economic Policy*, 23(1):63–78.

Audretsch, D. B., Bönte, W., and Keilbach, M. (2008). Entrepreneurship capital and its impact on knowledge diffusion and economic performance. *Journal of Business Venturing*, 23(6):687–698.

Audretsch, D. B. and Keilbach, M. (2004a). Entrepreneurship capital and economic performance. *Regional Studies*, 38(8):949–959.

Audretsch, D. B. and Keilbach, M. (2004b). Entrepreneurship capital: Determinants and impact. Working Paper 3704, Max Planck Institute for Research into Economic Systems, Research Group on Entrepreneurship, Growth and Public Policy.

Audretsch, D. B. and Keilbach, M. (2008). Resolving the knowledge paradox: Knowledge-spillover entrepreneurship and economic growth. *Research Policy*, 37(10):1697–1705.

Bagehot, W. (1873). *Lombard Street: A Description of the Money Market*. King, London.

Bai, J. and Ng, S. (2010). Instrumental variable estimation in a data rich environment. *Econometric Theory*, 26(6):1577–1606.

Baier, S., Dwyer, G., and Tamura, R. (2006). How important are capital and total factor productivity for economic growth? *Economic Inquiry*, 44(1):23–49.

Bandiera, O. and Caprio, G. (2000). Does financial reform raise or reduce saving? *Review of Economics and Statistics*, 82(2):239–263.

Banerjee, A. and Duflo, E. (2005). Growth theory through the lens of development economics. In Aghion, P. and Durlauf, S., editors, *Handbook of Economic Growth*, pages 473–552. North Holland, Amsterdam (NL).

Banerjee, A. and Duflo, E. (2012). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. Public Affairs, New York (US).

Banerjee, A., Duflo, E., Glennerster, R., and Kinnan, C. G. (2015). The miracle of microfinance? Evidence from a randomized evaluation. *American Economic Journal: Applied Economics*, 7(1):22–53.

Barndt, W. T., Bond, P., Gerring, J., and Moreno, C. (2005). Democracy and economic growth: A historical perspective. *World Politics*, 57(3):323–364.

Barro, R. (2001). Human capital and growth. *American Economic Review: Papers and Proceedings*, 91(2):12–14.

Barro, R. (2009). Rare disasters, asset prices, and welfare costs. *American Economic Review*, 99(1):243–264.

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

Barro, R. J. (2000). Inequality and growth in a panel of countries. *Journal of Economic Growth*, 5(1):5–32.

Barro, R. J. (2003). Determinants of economic growth in a panel of countries. *Annals of Economics and Finance*, 4(2):231–274.

Barro, R. J. (2013a). Education and economic growth. *Annals of Economics and Finance*, 14(2):301–328.

Barro, R. J. (2013b). Health and economic growth. *Annals of Economics and Finance*, 14(2):329–366.

Barro, R. J. and Lee, J.-W. (2013). A New Data set of Educational Attainment in the World, 1950–2010. *Journal of Development Economics*, 104:184–198.

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

Barro, R. J. and Sala-i-Martín, X. (2004). *Economic Growth*. MIT Press, Cambridge (US) and London (UK), 2nd edition.

Basso, A., Bodenhorn, H., and Cuberes, D. (2014). Fertility and financial development: Evidence from u.s. counties in the 19th century. NBER working paper no.20917, National Bureau of Economic Research.

Bazzi, S. and Clemens, M. A. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics*, 5(2):152–186.

Beck, R., Georgiadis, G., and Straub, R. (2014a). The finance and growth nexus revisited. *Economics Letters*, 124(3):382–385.

Beck, T. (2008). The econometrics of finance and growth. Policy Research Working Paper 4608, The World Bank.

Beck, T., Büyükkarabacak, B., Rioja, F. K., and Valev, N. T. (2012). Who Gets the Credit? And Does It Matter? Household vs. Firm Lending Across Countries. *The B.E. Journal of Macroeconomics*, 12(1):1–46.

Beck, T., Degryse, H., and Kneer, C. (2014b). Is more finance better? Disentangling intermediation and size effects of financial systems. *Journal of Financial Stability*, 10:50–64.

Beck, T., Demirgüç-Kunt, A., and Levine, R. (2003). Law and finance: Why does legal origin matter? *Journal of Comparative Economics*, 31(4):653–675.

Beck, T. and Levine, R. (2004). Stock markets, banks, and growth; panel evidence. *Journal of Banking and Finance*, 28(3):423–442.

Beck, T., Levine, R., and Loayza, N. (2000). Finance and the sources of growth. *Journal of Financial Economics*, 58(1–2):261–300.

Becker, G. S. (1960). Underinvestment in college education. *American Economic Review*, 50(2):346–354.

Becker, G. S. and Barro, R. J. (1988). A reformulation of the economic theory of fertility. *The Quarterly Journal of Economics*, 103(1):1–25.

Becker, G. S. and Lewis, H. G. (1973). On the interaction between quantity and quality of children. *Journal of Political Economy*, 81(2):279–288.

Belke, A. and Wernet, A. (2015). Poverty reduction through growth and redistribution policies—a panel analysis for 59 developing countries. *Review of Development Economics*, 19(1):143–162.

Bellman, R. (1961). *Adaptive Control Processes*. Princeton University Press, Princeton (US).

Benabou, R. (1996). Inequality and growth. In Bernanke, B. S. and Rotemberg, J. J., editors, *NBER Macroeconomics Annual 1996*, volume 11 of *NBER Chapters*, pages 11–92. MIT Press, Cambridge (US).

Benhabib, J. and Spiegel, M. M. (2000). The role of financial development in growth and investment. *Journal of Economic Growth*, 5(4):341–60.

Benhabib, J. and Spiegel, M. M. (2005). Human capital and technology diffusion. *Handbook of Economic Growth*, 1:935–966.

Bennett, K. P. and Mangasarian, O. (1992). Robust linear programming discrimination of two linearly inseparable sets. *Optimization methods and software*, 1(1):23–34.

Berg, A., Ostry, J. D., and Tsangarides, C. G. (2014). *Redistribution, Inequality, and Growth*. IMF Staff Discussion Note No.14/02, International Monetary Fund.

Berg, C., Christensen, J., and Ressel, P. (1984). *Harmonic Analysis on Semigroups*. Springer, New York (US).

Bergh, A. (2005). On the counterfactual problem of welfare state research: How can we measure redistribution? *European Sociological Review*, 21(4):345–357.

Berthold, N. and Gründler, K. (2015). The growth crisis of germany: A blueprint of the developed economies? *International Economic Journal*, 29(2):195–229.

Bertola, G. (1993). Market structure and income distribution in endogenous growth models. *American Economic Review*, 83(2):1184–99.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer, New York (US), 2nd edition.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Blundell, R., Bond, S., and Windmeijer, F. (2000). Estimation in dynamic panel data models: Improving on the performance of the standard gmm estimator. IFS Working Paper W00/12, Institute for Fiscal Studies.

Boix, C., Miller, M., and Rosato, S. (2013). A complete data set of political regimes, 1800–2007. *Comparative Political Studies*, 46(12):1523–1554.

Bollen, K. A. (2001). Cross-national indicators of liberal democracy, 1950–1990. ICPSR 2532, ICPSR Inter - University Consortium for Political and Social Research.

Bolt, J. and van Zanden, J. L. (2014). The maddison project: collaborative research on historical national accounts. *The Economic History Review*, 67(3):627–651.

Bolton, P., Santos, T., and Scheinkman, J. A. (2011). Cream skimming in financial markets. NBER Working Paper 16804, National Bureau of Economic Research.

Bond, S., Hoeffler, A., and Temple, J. (2001). Gmm estimation of empirical growth models. CEPR Discussion Paper 3048, Centre for Economic Policy Research.

Borio, C., Gambacorta, L., and Hofmann, B. (2017). The influence of monetary policy on bank profitability. *International Finance*, 20(1):48–63.

Boser, S., Guyon, I., and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In Haussler, D., editor, *Proceedings of the Annual Conference on Computational Learning Theory*, pages 144–152. ACM Press, Pittsburg (US).

Bourguignon, F. (1981). Pareto superiority of unegalitarian equilibria in stiglitz' model of wealth distribution with convex saving function. *Econometrica*, 49(6):1469–1475.

Brambor, T., Clark, W. R., and Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analyses*, 14(1):63–82.

Bresnahan, T. F. and Trajtenberg, M. T. (1995). General purpose technologies "engines of growth"? NBER Working Paper 4148, National Bureau of Economic Research.

Brownstone, D. and Valletta, R. (2001). The Bootstrap and Multiple Imputations: Harnessing Increased Computing Power for Improved Statistical Tests. *Journal of Economic Perspectives*, 15(4):129–141.

Brunnermeier, M. K., Dong, G., and Palia, D. (2012). Bank's non-interest income and systemic risk. Working paper.

Bruno, G. (2005a). Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models. *Economics Letters*, 87(3):361–366.

Bruno, G. (2005b). Estimation and inference in dynamic unbalanced panel-data models with a small number of individuals. *The Stata Journal*, 5(4):473–500.

Buera, F., Kaboski, J., and Shin, Y. (2011). Finance and development: A tale of two sectors. *American Economic Review*, 101(5):1964–2002.

Bun, M. J. and Kivet, J. F. (2003). On the diminishing returns of higher-order terms in asymptotic expansion of bias. *Economics Letters*, 79(2):145–152.

Bun, M. J. and Windmeijer, F. (2010). The weak instrument problem of the system gmm estimator in dynamic panel data models. *Econometrics Journal*, 13(1):95–126.

Bureau of Labor Statistics (2013). *Major Sector Multifactor Productivity database*. Bureau of Labor Statistics, Washington D.C.(US).

Burges, C. J. (1999). Geometry and invariance in kernel based methods. In Schölkopf, B., Burges, C. J., and Smola, A. J., editors, *Advances in Kernel Methods – Support Vector Learning*, pages 89–116. MIT Press, Cambridge (US).

Burkhart, R. and Lewis-Beck, M. (1994). Comparative democracy: The economic development thesis. *The American Political Science Review*, 88(4):903–910.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2014a). rdrobust: Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*, 14(4):909–946.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2014b). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2017a). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, forthcoming.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2017b). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2):372–404.

Campante, F. R. and Chor, D. (2012). Why was the Arab World poised for revolution? Schooling, economic opportunities, and the Arab Spring. *The Journal of Economic Perspectives*, 26(2):167–187.

Carr, A. (2014). *Psephos: Adam Carr's Electoral Archive*. Adam Carr's Electoral Archive, Melbourne (AU).

Carree, M., Van Stel, A., Thurik, R., and Wennekers, S. (2007). The relationship between economic development and business ownership revisited. *Entrepreneurship & Regional Development*, 19(3):281–291.

Caselli, F. and Coleman, W. J. (2001). Cross-country technology diffusion: The case of computers. *American Economic Review*, 91(2):328–335.

Cass, D. (1965). Optimum growth in an aggregate model of capital accumulation. *The Review of Economic Studies*, 32(3):233–240.

Castelló-Climent, A. (2010). Inequality and growth in advanced economies: An empirical investigation. *Journal of Economic Inequality*, 8(3):293–321.

Castells-Quintana, D. and Royuela, V. (2017). Tracking positive and negative effects of inequality on long-run growth. *Empirical Economics*, 54(4):1349–1378.

Cavalli-Sforza, L., Menozzi, P., and Piazza, A. (1994). *The history and geography of human genes*. Princeton University Press, Princeton (US).

Cervellati, M. and Sunde, U. (2007). Human capital, mortality and fertility: A Unified Theory of the Economic and Demographic Transition. IZA Discussion Paper no.2905, Institute for the Study of Labor (IZA).

Cervellati, M. and Sunde, U. (2011). Life expectancy and economic growth: The role of the Demographic Transition. *Journal of Ecoomic Growth*, 16(2):99–133.

Cheibub, J. A., Gandhi, J., and Vreeland, J. R. (2010). Democracy and dictatorship revisited. *Public Choice*, 143(1–2):67–101.

Chiao, J. and Blizinsky, K. (2010). Culture-gene coevolution of individualilsm-collectivism and the seronin transporter gene. *Proceedings – Royal Society: Biological sciences*, 277(1681):529–537.

Chinn, M. D. and Ito, H. (2006). What matters for financial development? Capital controls, institutions, and interactions. *Journal of Development Economics*, 81(1):163–192.

Cingranelli, D., Richards, D., and Clay, K. C. (2014). *The CIRI Human Rights Dataset*. CIRI Human Rights Data Project.

Comin, D., Easterly, W., and Gong, E. (2010). Was the wealth of nations determined in 1000 bc? *American Economic Journal: Macroeconomics*, 2(3):65–97.

Compton, R. A., Giedman, D. C., and Hoover, G. A. (2011). Panel evidence on economic freedom and growth in the united states. *European Journal of Political Economy*, 27(3):423–435.

Conley, T., Hansen, C., and Rossi, P. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272.

Contoyannis, P. and Jones, A. M. (2004). Socio-economic status, health, and lifestyle. *Journal of Health Economics*, 23(2):965–995.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

Cowen, T. (2011). *The Great Stagnation: How America Ate All The Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better*. Dutton, London (UK).

Dahl, R. A. (1971). *Polyarchy: Participation and Opposition*. Yale University Press, New Haven (US).

De Gregorio, J. and Guidotti, P. E. (1995). Financial Development and Economic Growth. *World Development*, 23(3):433–448.

De la Croix, D. and Doepke, M. (2003). Inequality and growth: Why differential fertility matters. *American Economic Review*, 93(4):1091–1113.

De la Torre, A., Feyen, E., and Ize, A. (2013). Financial development: Structure and dynamics. *World Bank Economic Review*, 27(3):514–541.

Deininger, K. and Squire, L. (1998). New ways of looking at old issues: inequality and growth. *Journal of Development Economics*, 57(2):259–287.

Demirgüç-Kunt, A., Feyen, E., and Levine, R. (2013). The Evolving Importance of Banks and Securities Markets. *World Bank Economic Review*, 27(3):476–490.

Demirgüç-Kunt, A. and Huizinga, H. (2010). Bank activity and funding strategies: The impact on risk and returns. *Journal of Financial Economics*, 98(3):626–650.

Diamond, J. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. Norton, New York and London.

Diamond, L. (2008). *The Spirit of Democracy. The Struggle to Build Free Societies Throughout the World*. Henry Holt, New York (US).

Diamond, L., Linz, J., and Lipset, S. M. (1990). *Politics in developing countries: Comparing experiences with democracy*. Lynne Rienner, Boulder (US).

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, 67(3):297–308.

Doepke, M. and Zilibotti, F. (2014). Culture, entrepreneurship, and growth. In *Handbook of Economic Growth*, volume 2, chapter 1, pages 1–48. Elsevier.

Domar, E. (1946). Capital expansion, rate of growth, and employment. *Econometrica*, 14(2):137–147.

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

Durlauf, S. N., Johnson, P. A., and Temple, Jonathan R. W. (2005). Chapter 8 growth econometrics. In Philippe Aghion and Steven N. Durlauf, editor, *Handbook of Economic Growth*, volume Volume 1, Part A, pages 555–677. Elsevier.

Easterly, W. (2007). Inequality does cause underdevelopment: Insights from a new instrument. *Journal of Development Economics*, 84(2):755–776.

Easterly, W., Islam, R., and Stiglitz, J. E. (2000). Shaken and Stirred: Explaining Growth Volatility. Working Paper prepared for the Annual World Bank Conference on development economics, The World Bank.

Easterly, W. and Levine, R. (2016). The european origins of economic development. *Journal of Economic Growth*, 21(3):225–257.

Eichengreen, B. (2015). Secular stagnation: The long view. *American Economic Review: Papers & Proceedings*, 105(5):66–70.

EIU (2009). *A New Ranking of the World's Most Innovative Countries*. Economist Intelligence Unit, London (UK).

Elias, J. S. (2015). Judges and democracy in argentina: An elite in search of legitimacy. In Turenne, S., editor, *Fair Reflection of Society in Judicial Systems-A Comparative Study*, pages 23–40. Springer, Cham (CH).

Ethier, W. J. (1982). National and international returns to scale in the modern theory of international trade. *American Economic Review*, 72(3):389–405.

Evans, D. S. and Leighton, L. S. (1989). The determinants of changes in the u.s. self-employment. *Small Business Economics*, 1(2):111–120.

Evans, D. S. and Leighton, L. S. (1990). Small business formation by unemployed and employed workers. *Small Business Economics*, 2(4):319–330.

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8(2):195–222.

Federico, P. J. (1964). Historical patent statistics 1791-1961. *Journal of the Patent Office Society*, 46(2):112–116.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2013). The next generation of the penn world table. NBER Working Paper 19255, National Bureau of Economic Research.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. *American Economic Review*, 105(10):3150–3182.

Feld, L. and Schnellenbach, J. (2014). Political institutions and income (re-)distribution: Evidence from developed economies. *Public Choice*, 159(3):435–455.

Feng, Y. (1997). Democracy, political stability, and economic growth. *British Journal of Political Science*, 27(3):391–418.

Fernández, R. (2011). Does culture matter? In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics*, volume 1, pages 481–510. North-Holland, Amsterdam (NL).

Ferreira, F. H. and Gignoux, J. (2011). The measurement of inequality of opportunity: Theory and application to latin america. *Journal of Income and Wealth*, 57(4):622–657.

Figueros, M. (1985). An assessment of overvoting in Jamaica. *Social and Economic Studies*, 34(3):71–106.

Finlay, K. and Magnusson, L. (2009). Implementing weak-instrument robust tests for a general class of instrumental-variable models. *The Stata Journal*, 9(3):398–421.

Finlay, K., Magnusson, L., and Schaffer, M. E. (2016). *WEAKIV: Stata module to perform weak-instrument-robust tests and confidence intervals for instrumental-variable (IV) estimation of linear, probit and tobit models*. Statistical Software Components, Boston College Department of Economics.

Forbes, K. J. (2000). A reassessment of the relationship between inequality and growth. *American Economic Review*, 90(4):869–887.

Frankel, M. (1962). The production function in allocation of growth: A synthesis. *American Economic Review*, 52(5):995–1022.

Freedom House (2014). *Freedom in the World Database*. Freedom House, Washington D.C.(US).

Freedom House (2017). *Freedom in the World database*. Freedom House, Washington D.C.(US).

Galor, O. (2009). Inequality and economic development: An overview. Working Paper 3, Brown University, Department of Economics.

Galor, O. (2011). *Unified Growth Theory*. Princeton University Press, New Jersey (US).

Galor, O. and Moav, O. (2002). Natural selection and the origin of economic growth. *Journal of Economics*, 117(4):1133–1191.

Galor, O. and Moav, O. (2004). From physical to human capital accumulation: Inequality and the process of development. *Review of Economic Studies*, 71(4):1001–1026.

Galor, O. and Weil, D. N. (1999). From malthusian stagnation to modern growth. *American Economic Review*, 89(2):150–154.

Galor, O. and Weil, D. N. (2000). Population, technology and growth: From the malthusian stagnation to the demographic transition and beyond. *American Economic Review*, 90(4):806–828.

Galor, O. and Zang, H. (1997). Fertility, income distribution, and economic growth: Theory and cross-country evidence. *Japan and the World Economy*, 9(2):197–229.

Galor, O. and Zeira, J. (1993). Income distribution and macroeconomics. *Review of Economic Studies*, 60(1):35–52.

GEM (2017). *Global Entrepreneurship Monitor Database 2017*. Global Entrepreneurship Monitor, London (UK).

George, E. I. and Foster, D. P. (2000). Calibration and empirical bayes variable selection. *Biometrika*, 87(4):731–747.

Gerring, J., Bond, P., Barndt, W. T., and Moreno, C. (2005). Democracy and economic growth. A historical perspective. *World Politics*, 57(3):323–364.

Gerschenkron, A. (1962). *Economic Backwardness in Historical Perspective: A Book of Essays*. Harvard University Press, Cambridge (US).

Giavazzi, F. and Tabellini, G. (2005). Economic and political liberalizations. *Journal of Monetary Economics*, 52(7):1297–1330.

Gibney, M., Cornett, L., Wood, R., and Haschke, P. (2013). *Political Terror Scale, 1976–2012*. Political Terror Scale, Asheville (US).

Girosi, F., Jones, M., and Poggio, T. (1993). Priors, stabilizers and basis functions: From regularization to radial, tensor and additive splines. A.i. memo no. 1430, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Glaeser, E., Scheinkman, J., and Shleifer, A. (2003). The injustice of inequality. *Journal of Monetary Economics*, 50(1):199–222.

Glaeser, E. L., Pekkala, S., and Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statitics*, 97(2):498–520.

Goldin, C. and Katz, Lawrence, F. (2007). The race between education and technology: The Evolution of U.S. educational wage differentials, 1890 to 2005. NBER Working Paper, National Bureau of Economic Research.

Goldsmith, R. W. (1969). *Financial Structure and Development*. Yale University Press, New Haven (US).

Gordon, R. J. (2012). Is US economic growth over? Faltering innovation confronts the six headwinds. NBER Working Paper 18315, National Bureau of Economic Research.

Gordon, R. J. (2015). Secular stagnation: A supply-side view. *American Economic Review: Papers & Proceedings*, 105(5):54–59.

Gorodnichenko, Y. and Roland, G. (2017). Culture, institutions and the wealth of nations. *Review of Economics and Statistics*, 99(3):402–416.

Gorodnichenko, Y. and Schnitzer, M. (2013). Financial constraints and innovation: Why poor countries dont't catch up. *Journal of the European Economic Association*, 11(5):1115–1152.

Greenberg, J. (2014). *After the Revolution: Youth, Democracy, and the Politics of Disappointment in Serbia*. Stanford University Press, Stanford (US).

Grossman, G. M. and Helpman, E. (1991). *Innovation and Growth in the Global Economy*. MIT Press, Cambridge (US).

Gründler, K. and Krieger, T. (2016). Democracy and growth: Evidence from a machine learning indicator. *European Journal of Political Economy*, 45(S):85–107.

Gründler, K. and Scheuermeyer, P. (2018). Growth Effects of Inequality and Redistribution: What are the Transmission Channels? *Journal of Macroeconomics*, 55(3):293–313.

Gualtieri, A. (2009). The support vector machine (svm) algorithm for supervised classification of hyperspectral remote sensing data. In Camps-Valls, G. and Bruzzone, L., editors, *Kernel Methods for Remote Sensing Data Analysis*, pages 49–83. John Wiley & Sons, Ltd, Chichester (UK).

Gugiu, M. R. and Centellas, M. (2013). The democracy cluster classification index. *Political Analysis*, 21(3):334–349.

Guiso, L., Sapienza, P., and Zingales, L. (2004). Does local financial development matter? *The Quarterly Journal of Economics*, 119(3):929–969.

Guiso, L., Sapienza, P., and Zingales, L. (2006). Does culture affect economic outcomes. *The Journal of Economic Perspectives*, 20(3):23–48.

Guyon, I., Boser, S., and Vapnik, V. (1993). Automatic capacity tuning of very large vc-dimension classifier. In Hanson, S. J., editor, *Advances in Neural Information Processing Systems 5*, pages 147–155. Morgan Kaufmann Publishers, Burlington (US).

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1–3):389–422.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, insights and methodological tools. NBER Working Paper 8498, National Bureau of Economic Research.

Hall, R. E. and Jones, Charles, I. (1999). Why do some countries produce so much more output per capita than others? *Quarterly Journal of Economics*, 114(1):83–116.

Halter, D., Oechslin, M., and Zweimüller, J. (2014). Inequality and growth: the neglected time dimension. *Journal of Economic Growth*, 19(1):81–104.

Hansen, A. H. (1939). Economic progress and declining population growth. *American Economic Review*, 29(1):1–15.

Hanushek, E. and Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4):267–321.

Harrod, Roy, F. (1939). An essay in dynamic theory. *The Economic Journal*, 49(193):14–33.

Haustein, H.-D. and Neuwirth, E. (1982). Long waves in world industrial production, energy consumption, innovations, inventions, and patents and their identification by spectral analysis. *Technological Forecasting and Social Change*, 22(1):53–89.

Helpman, E. and Trajtenberg, M. (1994). A time to sow and a time to reap: Growth based on general purpose technologies. NBER Working Paper 4854, National Bureau of Economic Research.

Henderson, D. J., Papageorgiou, C., and Parmeter, C. F. (2013). Who benefits from financial development? New methods, new evidence. *European Economic Review*, 63:47–67.

Heston, A., Summers, R., and Aten, B. (2012). *Penn World Tables, Version 7.1*. Center for International Comparisons of Production, Income, and Price at the University of Pennsylvania, Philadelphia (US).

Hofstede, G. (2001). *Culture's Consequences. Comparing Values, Behaviors, Institutions, and Organizations Across Nations.* Sage, Thousand Oaks (US) and London (UK).

Huang, Y. and Lin, S.-C. (2009). Non-linear finance-growth nexus. *Economics of Transition*, 17(3):439–466.

Huber, P. (1981). *Robust statistics*. John Wiley and Sons, New York (US).

Huntington, S. P. (1991). Democracy's third wave. *Journal of Democracy*, 2(2):12–34.

Huntington, S. P. (2012). *The Third Wave: Democratization in the Late 20th Century*. University of Oklahoma Press, Norman (US).

IDEA (2014). *Voter Turnout Database*. International Institute for Democracy and Electoral Assistance, Stockholm (SE).

IPU (2014). *Parline Database on National Parliaments*. Inter-Parliamentary Union, Le Grand-Saconnex (CH).

Jain, R. K., Smith, K. M., Culligan, P. J., and Taylor, J. E. (2014). Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy*, 123:168–178.

Jappelli, T. and Pagano, M. (1994). Saving, growth, and liquidity constraints. *Quarterly Journal of Economics*, 109(1):83–109.

Jenkins, S. P. (2015). World income inequality databases: an assessment of wiid and swiid. *Journal of Economic Inequality*, 13(4):629–671.

Joachims, T. (2002). *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms*. Kluwer Academic Publishers, Norwell (US).

Johnes, G. (1993). *The Economics of Education*. St. Martin's Press, New York (US).

Jones, C. (1997). On the evolution of the world income distribution. *Journal of Economic Perspectives*, 11(3):19–36.

Jordà, Ò., Schularick, M., and Taylor, A. M. (2016). The great mortgaging: housing finance, crises and business cycles. *Economic Policy*, 31(85):107–152.

Jovanovic, B. and Rousseau, P. L. (2005). General purpose technologies. In Aghion, P. and Durlauf, S., editors, *Handbook of Economic Growth*, pages 1182–1224. North Holland, Amsterdam (NL).

Judson, R. A. and Owen, A. L. (1999). Estimating dynamic panel data models: a guide for macroeconomists. *Economics Letters*, 65(1):9–15.

Kaldor, N. (1955). Alternative theories of distribution. *Review of Economic Studies*, 23(2):83–100.

Kapetanios, G. and Marcellino, M. (2010). Factor-gmm estimation with large sets of possibly weak instruments. *Computational Statistics and Data Analysis*, 54(11):2655–2675.

Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints. Master's thesis, University of Chicago, Dept. of Mathematics.

Kashima, E. and Kashima, Y. (1998). Culture and language: The case of cultural dimensions and personal pronoun use. *Journal of Cross-Cultural Psychology*, 29(3):461–487.

Kihlstrom, R. E. and Laffont, J.-J. (1979). A general equilibrium entrepreneurial theory of firm formation based on risk aversion. *Journal of Political Economy*, 87(4):719–748.

Kim, K.-J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1):307–319.

King, R. G. and Levine, R. (1993a). Finance and growth: Schumpeter might be right. *Quarterly Journal of Economics*, 108(3):717–737.

King, R. G. and Levine, R. (1993b). Finance, entrepreneurship and growth: Theory and evidence. *Journal of Monetary Economics*, 32(3):513–542.

Kiviet, J. F. (1995). On bias, inconsistency and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics*, 68(1):53–78.

Kleibergen, F. (2005). Testimg parameters in GMM without assuming that they are identified. *Econometrica*, 73(4):1103–1123.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126.

Kneer, C. (2013). Finance as a magnet for the best and the brightest: Implications for the real economy. DNP Working Papers 392, De Nederlandsche Bank.

Knight, F. H. (1921). *Risk, Uncertainty, and Profit*. Houghton Miflin, Boston (US).

Knowles, S. (2005). Inequality and economic growth: The empirical relationship reconsidered in the light of comparable data. *The Journal of Development Studies*, 41(1):135–159.

Koltchinskii, V. (2001). Rademacher penalties and structural risk minimization. *IEEE Transactions on Information Theory*, 47(5):1902–1914.

Koopmans, T. C. (1965). *On the Concept of Optimal Growth. The Econometric Approach to Development Planning*. North Holland, Amsterdam (NL).

Korotayev, A., Zinkina, J., and Bogevolnov, J. (2011). Kondratieff waves in global invention activity (1900–2008). *Technological Forecasting and Social Change*, 78(7):1280–1284.

Kortum, S. and Lerner, J. (2000). Assessing the contribution of venture capital to innovation. *RAND Journal of Economics*, 31(4):674–692.

Kraay, A. (2015). Weak instruments in growth regressions. implications for recent cross-country evidence on inequality and growth. Policy Research Working Paper 7494, The World Bank.

Kremer, M. and Chen, D. L. (2002). Income distribution dynamics with endogenous fertility. *Journal of Economic Growth*, 7(3):227–258.

Kroszner, R. S., Laeven, L., and Klingebiel, D. (2007). Banking crises, financial dependence, and growth. *Journal of Financial Economics*, 84(1):187–228.

Kuhn, H. and Tucker, A. (1951). Nonlinear programming. In Neyman, J., editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, page 481–492. University of Chicago Press, Berkeley (US).

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. W. (1997). Legal determinants of external finance. *Journal of Finance*, 52(3):1131–1150.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. W. (1998). Law and finance. *Journal of Political Economy*, 106(6):1113–1155.

Laeven, L., Levine, R., and Michalopoulos, S. (2015). Financial innovation and endogenous growth. *Journal of Financial Intermediation*, 24(1):1–24.

Lakner, C. and Milanovic, B. (2015). Global income distribution: From the fall of the berlin wall to the great recession. *World Bank Economic Review*, 30(2):203–232.

Landes, D. (1998). *The Wealth and Poverty of Nations*. Norton, New York.

Larkins, C. (1998). The judiciary and delegative democracy in Argentina. *Comparative Politics*, 30(4):423–442.

Law, S. H. and Singh, N. (2014). Does too much finance harm economic growth? *Journal of Banking & Finance*, 41:36–44.

LeCun, Y. (1986). Learning processes in an asymmetric threshold network. In Bienenstock, E., Fogelman-Soulié, F., and Weisbuch, G., editors, *Disordered systems and biological organization*, pages 233–240. Springer, Berlin (DE).

Levine, R. (1997). Financial development and economic growth: Views and agenda. *Journal of Economic Literature*, 35(2):688–726.

Levine, R. (2005). Finance and growth: Theory and evidence. In Durlauf, S. N. and Aghion, P., editors, *Handbook of Economic Growth*, volume 1A, pages 865–934. Elsevier, Amsterdam.

Levine, R., Loayza, N., and Beck, T. (2000). Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics*, 46(1):31–77.

Levitsky, S. and Murillo, M. V. (2008). Argentina: From Kirchner to Kirchner. *Journal of Democracy*, 19(2):16–30.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economic Statistics*, 30(1):67–80.

Li, H. and Zou, H.-F. (1998). Income inequality is not harmful for growth: Theory and evidence. *Review of Development Economics*, 2(3):318–334.

Licht, A. N. (2007). Culture rules: The foundation of the rule of law and other norms of governance. *Journal of Comparative Economics*, 34(4):659–688.

Lind, J. T. and Mehlum, H. (2010). With or without u? the appropriate test for a u-shaped relationship. *Oxford Bulletin of Economics and Statistics*, 72(1):109–118.

Lindert, P. (2004). *Social Spending and Economic Growth Since the Eighteenth Century*. Cambridge University Press, Cambridge (UK).

Lipset, S. M. (1959). Some social requisites of democracy. *American Political Science Review*, 53(1):69–105.

Lizzeri, A. and Persico, N. (2004). Why did the elites extend the suffrage? Democracy and the scope of government, with an application to britain's "age of reform". *The Quarterly Journal of Economics*, 119(2):705–763.

Loayza, N. V. and Ranciere, R. (2006). Financial development, financial fragility, and growth. *Journal of Money, Credit and Banking*, 38(4):1051–1076.

Lochner, L. and Monge-Naranjo, A. (2012). Credit constraints in education. *Annual Review of Economics*, 4(1):225–256.

Lucas, R. (1990). Why doesn't capital flow from rich to poor countries? *American Economic Review*, 80(2):92–96.

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

Lupu, N. and Pontusson, J. (2011). The structure of inequality and the politics of redistribution. *American Political Science Review*, 105(2):316–336.

Luttmer, E. F. P. and Singhal, M. (2011). Culture, context, and the taste for redistribution. *American Economic Journal: Economic Policy*, 3(1):157–179.

Maddison, A. (2013). *Maddison Project Database, Statistics on World Population, GDP and Per Capita GDP*. Maddison Project, Groningen (NL).

Madsen, J. B. (2008). Semi-endogenous versus schumpeterian growth models: testing the knowledge production function using international data. *Journal of Economic Growth*, 13(1):1–26.

Madsen, J. B. (2014). Human capital and the world technological frontier. *Review of Economics and Statistics*, 96(4):676–692.

Madsen, J. B. and Ang, J. B. (2016). Finance-led growth in the OECD since the nineteenth century: How does financial development transmit to growth? *Review of Economics and Statistics*, 98(3):552–572.

Madsen, J. B., Raschky, P. A., and Skali, A. (2015). Does democracy drive income in the world, 1500-2000? *European Economic Review*, 78:175–195.

Malahanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1):49–55.

Mangasarian, L. (1969). *Linear Programming*. McGraw-Hill, New York (US).

Mankiw, G. (2009). *Macroeconomics*. Worth Publishers, New York (US), 7th edition.

Mankiw, G. N., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2):407–437.

Marrero, G. A. and Rodriguez, J. G. (2013). Inequality of oppoprtunity and growth. *Journal of Development Economics*, 104:107–122.

Marrero, G. A., Rodriguez, J. G., and Van der Weide, R. (2016). Unequal opportunity, unequal growth. Policy Research Working Paper 7853, World Bank.

Marshall, M., Gurr, T. R., and Jaggers, K. (2014). *Polity IV Project. Political Regime Characteristics and Transitions, 1800–2013*. Center of Systematic Peace, Vienna (US).

Maseland, R. (2013). Parasitical cultures? The cultural origins of institutions and development. *Journal of Economic Growth*, 18(2):109–136.

McConnell, S. A. (2014). The 2011 Presidential and Legislative Elections in Nicaragua. *Electoral Studies*, 34:300–303.

McFaul, M. (2002). The fourth wave of democracy and dictatorship: Noncooperative transitions in the postcommunist world. *World Politics*, 54(2):212–244.

McKinnon, R. I. (1973). *Money and Capital in Economic Development*. Brookings Institution Press, Washington D.C. (US).

Meltzer, A. H. and Richard, S. F. (1981). A rational theory of the size of government. *Journal of Political Economy*, 89(5):914–927.

Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character*, 209(441-458):415–446.

Miguel, E. and Satyanath, S. (2010). Understanding transitory rainfall shocks, economic growth, and civil conflict. NBER Working Paper N. 16461, National Bureau of Economic Research.

Milanovic, B. (2000). The median-voter hypothesis, income inequality, and income redistribution: An empirical test with the required data. *European Journal of Political Economy*, 16(3):367–410.

Milanovic, B. (2013). Global income inequality in numbers: in history and now. *Global policy*, 4(2):198–208.

Mokyr, J. (2016). *A Culture of Growth: The Origins of the Modern Economy*. Princeton University Press, Princeton.

Morales, M. F. (2008). Financial intermediation in a model of growth though creative destruction. *Macroeconomic Dynamics*, 7:363–393.

Morand, O. (1999). Endogenous fertility, income distribution, and growth. *Journal of Economic Growth*, 4(3):331–349.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048.

Morozov, V. (1984). *Methods for Solving Incorrectly Posed Problems*. Springer, New York (US).

Muinelo-Gallo, L. and Roca-Sagalés, O. (2013). Joint determinants of fiscal policy, income inequality and economic growth. *Economic Modelling*, 30(0):814–824.

Müller, K.-R., Smola, A. J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., and Vapnik, V. (1997). Predicting time series with support vector machines. In Gerstner, W., Germond, A., Hasler, M., and Nicoud, J.-D., editors, *Artificial Neural Networks ICANN'97*, volume 1327, pages 999–1004. Springer, Berlin (DE).

Munck, G. L. and Verkuilen, J. (2002). Conceptualizing and meassuring democracy: Evaluating alternative indices. *Comparative Political Studies*, 35(1):5–34.

Murray, D. and Schaller, M. (2010). Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture. *Journal of Cross-Cultural Psychology*, 41(1):99–108.

Murtin, F. and Wacziarg, R. (2014). The democratic transition. *The Journal of Economic Growth*, 19(2):141–181.

Nelson, R. R. and Phelps, E. S. (1966). Investment in humans, technological diffusion, and economic growth. *American Economic Review*, 56(1/2):69–75.

Neves, P. and Silva, S. (2014). Inequality and growth: Uncovering the main conclusions from the empirics. *Journal of Development Studies*, 50(1):1–21.

Newey, W. K. and Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, 77(3):687–719.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Nielsen, R. W. (2013). Unified growth theory: A puzzling collection of myths based on hyperbolic illusions. Working paper, Environmental and Future Research Institute.

Nikulainen, T. and Kulvik, M. (2009). How general are general purpose technologies? evidence from nano-, bio-and ict-technologies in finland. ETLA Discussion Paper 1208, The Research Institute of the Finish Economy.

Nilsson, N. (1965). *Learning machines: foundations of trainable pattern classification systems*. McGraw-Hill, New York (US).

Nohlen, D. (2005). *Elections in the Americas: A Data Handbook: Volume 2 South America*, volume 2. Oxford University Press, Oxford (UK).

Nohlen, D. (2010). *Elections in Europe: A data handbook*. Nomos, Baden-Baden (DE).

Nohlen, D., Grotz, F., and Hartmann, C. (2001). *Elections in Asia and the Pacific: A Data Handbook*. Oxford University Press, Oxford (UK).

Nohlen, D., Thibaut, B., and Krennerich, M. (1999). *Elections in Africa: A Data Handbook: A Data Handbook*. Oxford University Press, Oxford (UK).

Novikoff, A. B. (1962). On convergence proofs on perceptrons. Symposium on the mathematical theory of automata no.12, Polythechnic Institute of Brooklyn.

O'Donell, G. A. (1994). Delegative democracy. *Journal of Democracy*, 5(1):55–69.

OECD (2012). *Equity and Quality in Education: Supporting Disadvantaged Students and Schools*. OECD publishing, Paris (FR).

OECD (2013a). Patents database. Technical report, Paris (FR).

OECD (2013b). Productivity statistics. Technical report, Paris (FR).

OECD (2016). *ICT Investment (indicator)*. Organisation for Economic Co-operation and Development, Paris (FR).

Okun, A. (1975). *Equality and Efficiency: The Big Tradeoff*. The Brookings Institution, Washington D.C. (US).

Olea, J. L. M. and Pflüger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369.

Olson, M. (1982). *The Rise and Decline of Nations*. Yale University Press, New Haven (US).

Orrú, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., and Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience and Biobehavioral Reviews*, 36(4):1140–1152.

Panizza, U. (2002). Income inequality and economic growth: Evidence from american data. *Journal of Economic Growth*, 7(1):25–41.

Panizza, U. (2013). Financial development and economic growth: Known knowns, known unknowns, and unknown unknowns. IHEID working paper no.14, The Graduate Institute of International and Development Studies.

Papaioannou, E. and Siourounis, G. (2008). Democratisation and growth. *Economic Journal*, 118(532):1520–1551.

Pappas, G., Roussos, N., and Falagas, M. (2009). Toxoplasmosis snapshots: Global status of toxoplasma gondii seroprevalence and implications for pregnancy and congenital toxoplasmosis. *International Journal for Parasitology*, 39(12):1385–1394.

Pemstein, D., Meserve, S. A., and Melton, J. (2010). Democratic compromise: A latent variable analysis of ten measures of regime type. *Political Analysis*, 18(4):426–449.

Perla, J. and Tonetti, C. (2014). Equilibrium imitation and economic growth. *Journal of Political Economy*, 122(1):52–76.

Perotti, R. (1994). Income distribution and investment. *European Economic Review*, 38(3–4):827–835.

Perotti, R. (1996). Growth, income distribution, and democracy: What the data say. *Journal of Economic Growth*, 1(2):149–187.

Persson, T. and Tabellini, G. (1994). Is inequality harmful for growth? *The American Economic Review*, 84(3):600–621.

Persson, T. and Tabellini, G. (2008). The growth effect of democracy: Is it heterogenous and how can it be estimated? In Helpman, E., editor, *Institutions and Economic Performance*. Harvard University Press, Cambridge (US).

Persson, T. and Tabellini, G. (2009). Democratic capital: The nexus of political and economic change. *American Economic Journal: Macroeconomics*, 1(2):88–126.

Petsas, I. (2003). The dynamic effects of general purpose technologies on schumpeterian growth. *Journal of Evolutionary Economics*, 13(5):577–605.

Philippon, T. (2010). Financiers versus engineers: Should the financial sector be taxed or subsidized? *American Economic Journal: Macroeconomics*, 2(3):158–82.

Philippon, T. and Reshef, A. (2013). An international look at the growth of modern finance. *Journal of Economic Perspectives*, 27(2):73–96.

Picot, G., Manser, M. E., and Lin, Z. (1998). The role of self-employment in job creation in Canada and the US. Oecd-cerf-ciln international conference on self-employment, Organisation for Economic Co-operation and Development, Center for Economic Research and Forecasting, and Canadian International Labour Network, Burlington (CA).

Poschke, M. (2014). The firm size distribution across countries and skill-biased change in entrepreneurial technology. IZA Discussion Paper 7991, Institute for the Study of Labor (IZA).

Putterman, L. and Weil, D. N. (2010). Post-1500 population flows and the long-run determinants of economic growth and inequality. *Quarterly Journal of Economics*, 125(4):1627–1682.

Ragin, C. C. (2000). *Fuzzy-set social science*. University of Chicago Press, Chicago (US).

Ragin, C. C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. University of Chicago Press, Chicago (US).

Rajan, R. and Zingales, L. (1998). Financial dependence and growth. *American Economic Review*, 88(3):559–586.

Rajan, R. G. (2005). Has financial development made the world riskier? NBER Working Paper 11728, National Bureau of Economic Research.

Rajan, R. G. and Zingales, L. (2003). The great reversal: The politics of finacial development in the twentieth century. *Journal of Financial Economics*, 69(1):5–50.

Rancière, R., Tornell, A., and Westermann, F. (2008). Systemic Crises and Growth. *Quarterly Journal of Economics*, 123(1):359–406.

Rauch, J. and Trindade, V. (2002). Ethnic Chinese networks in international trade. *Review of Economics and Statistics*, 84(1):116–130.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press, Cambridge (US).

Reinhart, C. M. and Kaminsky, G. L. (1999). The twin crises: The causes of banking and balance-of-payments problems. *American Economic Review*, 89(3):473–500.

Reynolds, P. D. (1999). Creative destruction: Source or sympton of economic growth. In Acs, Z. J., Carlsson, B., and Karlsson, C., editors, *Entrepreneurship, Small and Medium-Sized Enterprizes and the Macroeconomy*, chapter 4, pages 97–136. Cambridge University Press, Cambridge (UK).

Ricardo, D. (1817). *On the Principles of Political Economy and Taxation*. Juhn Murray, London (UK), 3rd edition.

Robinson, J. (1952). *The Rate of Interest and other Essays*. Macmillan, London.

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

Roemer, J. E. and Trannoy, A. (2016). Equality of opportunity: Theory and measurement. *Journal of Economic Literature*, 54(4):1288–1332.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5):1002–1037.

Romer, P. M. (1987). Growth based on increasing returns due to specialization. *The American Economic Review*, 77(1):56–62.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5):71–102.

Romer, P. M. (1993). Idea gaps and object gaps in economic development. *Journal of Monetary Economics*, 32(3):543–573.

Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system gmm in stata. *Stata Journal*, 9(1):86–136.

Roodman, D. (2009b). How to do xtabond2: An introduction to difference and system gmm in stata. *Stata Journal*, 9(1):86.

Roodman, D. (2009c). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1):135–158.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Reviews*, 65(6):386–408.

Rousseau, P. L. and Wachtel, P. (2002). Inflation thresholds and the finance-growth nexus. *Journal of International Money and Finance*, 21(6):777–793.

Rousseau, P. L. and Wachtel, P. (2011). What is happening to the impact of financial deepening on economic growth? *Economic Inquiry*, 49(1):276–288.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons, Inc.

Rumelhart, D., Hinton, G., and Williams, R. (1986). Learning internal representations by error propagation. In Rumelhart, D. and McClelland, J., editors, *Parallel Distributed Processing: Explorations in the Macrostructure of Cognition*, volume 1, pages 318–362. Bradford Books, Cambridge (US).

Saint-Paul, G. and Verdier, T. (1993). Education, democracy and growth. *Journal of Development Economics*, 42(2):399–407.

Sala-i Martín, X. (1997). I just ran two million regressions. *The American Economic Review*, 87(2):178–183.

Salgado-Banda, H. (2007). Entrepreneurship and economic growth: An empirical analysis. *Journal of Development Entrepreneurship*, 12(1):3–29.

Samila, S. and Sorenson, O. (2011). Venture capital, entrepreneurship, and economic growth. *Review of Economics and Statistics*, 93(1):338–349.

Samuels, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229.

Sanderson, E. and Windmeijer, F. (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics*, forthcoming.

Sapir, E. (1970). Language. In Mandelbaum, D., editor, *Culture, Language and Personality: Selected Essays*. University of Calilfornia Press, Berkeley (US).

Sardadvar, S. (2011). *Economic Growth in the Regions of Europe. Theory and Empirical Evidence from a Spatial Growth Model.* Springer, Berlin (DE).

Schoenberg, I. (1942). Positive definite functions of spheres. *Duke Mathematical Journal*, 9(1):96–108.

Schölkopf, B. (1995). Extracting support data for a given task. In Fayyad, U. and Uthurusamy, R., editors, *Proceedings, First International Conference on Knowledge Discovery & Data Mining*. AAAI Press, Palo Alto (US).

Schölkopf, B., Smola, A., Williamson, R., and Bartlett, P. (2000). New support vector algorithms. *Neural Computation*, 12(5):1207–1245.

Schölkopf, B. and Smola, A. J. (2001). *Learning with Kernels*. MIT-Press, Cambridge (US) and London (UK).

Schularick, M. and Taylor, A. (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises. *American Economic Review*, 102(2):1029–1061.

Schumpeter, J. (1947). The creative response in economic history. *Journal of Economic History*, 7(2):149–159.

Schumpeter, J. A. (1911). *Theorie der wirtschaftlichen Entwicklung*. Duncker & Humblot, Leipzig (DE).

Schumpeter, J. A. (1934). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*. Transaction Publilshers, New Brunswick (US) and London (UK).

Schumpeter, J. A. (1939). *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process*. McGraw-Hill, New York (US).

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. Harper and Row, New York (US).

Schwartz, S. H. (2006). A theory of cultural value orientations: Explication and applications. *Comparative Sociology*, 5(2–3):137–182.

Severinghaus, S. R. (2001). Mongolia in 2000 the pendulum swings again. *Asian Survey*, 41(1):61–70.

Shin, K.-S., Lee, T. S., and Kim, H.-j. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28(1):127–135.

Silverberg, G. and Verspagen, B. (2003). Breaking the waves: A poisson regression approach to schumpeterian clustering of basic innovations. *Cambridge Journal of Economics*, 27(5):671–693.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. William Strahan Stahan, London (UK).

Smola, A., Murata, N., Schölkopf, B., and Müller, K.-J. (1998a). Asymptotically optimal choice of ε-loss for support vector machines. In Niklasson, L., Bodén, M., and Ziemke, T., editors, *Proceedings of the International Conference on Artificial Neural Networks. Perspectives in Neural Computing*, pages 105–110. Springer, Berlin (DE).

Smola, A. J. and Schölkopf, B. (1998). On a kernel-based method for pattern recognition, regression, approximation and operator inversion. *Algorithmica*, 22(1–2):211–231.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222.

Smola, A. J., Schölkopf, B., and Müller, K.-R. (1998b). The connection between regularization operators and support vector kernels. *Neural Networks*, 11(4):637–649.

Smola, A. J., Schölkopf, B., and Müller, K.-R. (1998c). General cost functions for support vector regression. In Downs, T., Frean, M., and Gallagher, M., editors, *Proceedings of the 8th International Conference on Artificial Neural Networks*, pages 79–83. University of Maryland, Brisbane (US).

Smolny, W. (2000). Post-war growth, productivity convergence and reconstruction. *Oxford Bulletin of Economics and Statistics*, 62(5):589–606.

Soartes, R. R. (2005). Mortality reductions, educational attainment, and fertility choice. *American Economic Review*, 95(3):580–601.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1):65–94.

Solow, R. M. (1957). Technical change and the aggregate production function. *The Review of Economics and Statistics*, 39(3):312–320.

Solt, F. (2009). Standardizing the world income inequality database. *Social Science Quarterly*, 90(2):231–242.

Solt, F. (2015). On the assessment and use of cross-national income inequality datasets. *The Journal of Economic Inequality*, 13(4):683–691.

Solt, F. (2016). The standardized world income inequality database. *Social Science Quarterly*, 97:1267–1281.

Spolaore, E. and Wacziarg, R. (2009). The diffusion of development. *Quarterly Journal of Economics*, 124(2):469–529.

Spolaore, E. and Wacziarg, R. (2013). How deep are the roots of economic development? *Journal of Economic Literature*, 51(2):325–369.

Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Springer Verlag, New York (US).

Stiroh, K. J. (2004). Diversification in banking: Is noninterest income the answer? *Journal of Money, Credit, and Banking*, 36(5):853–882.

Stock, J. A. and Yogo, M. (2005). Testing for weak instruments in linear iv regression. In Andrews, D. W., editor, *Identification and Inference for Econometric Models*. Cambridge University Press, New York.

Summers, L. H. (2014). Reflections on the 'new secular stagnation hypothesis'. In Teulings, C. and Baldwin, R., editors, *Secular Stagnation: Facts, Causes, and Cures*, pages 27–40. CEPR Press, London (UK).

Summers, L. H. (2015). Demand side secular stagnation. *American Economic Review: Papers & Proceedings*, 105(5):60–65.

Swan, T. W. (1956). Economic growth and capital accumulation. *Economic Record*, 32(2):334–361.

Sydsæter, K. and Hammond, P. (2008). *Essential Mathematics for Economic Analysis*. Pearson Education, Boston (US).

Tabellini, G. (2008). Presidential address: Institutions and culture. *Journal of the European Economic Association*, 6(2-3):255–294.

Tabellini, G. (2010). Culture and institutions: Economic development in the regions of Europe. *Journal of the European Economic Association*, 8(4):677–716.

Tavares, J. and Wacziarg, R. (2001). How democracy affects growth. *European Economic Review*, 43(8):1341–1378.

Tay, F. E. and Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29(4):309–317.

TED (2017). *The Conference Board Total Economy Database (TED)*. The Conference Board, New York (US).

The Conference Board (2015). *The Conference Board Total Economy Database (TED)*. The Conference Board, New York (US).

Thewissen, S. (2014). Is it the income distribution or redistribution that affects growth? *Socio-Economic Review*, 12(3):545–571.

Thurik, R. and Wennekers, S. (1999). Linking entrepreneurship and economic growth. *Small Business Economics*, 13(1):27–55.

Tikhonov, A. N. and Arsenin, V. (1977). *Solution of ill-posed problems*. V.H. Winston and Sons, Washingtion D.C. (US).

Treier, S. and Jackman, S. (2008). Democracy as a latent variable. *American Journal of Political Science*, 52(1):201–217.

United Nations (2013). Human development index - technical notes. Technical report, United Nations, New York (US).

US Patent and Trademark Office (2013). *U.S. Patent Activity, 1790-Present*. US Patent and Trademark Office, Alexandria (US).

UTIP (2012). *Estimated Household Income Inequality Data Set*. Universitiy of Texas Inequality Project, Austin (US).

Uzawa, H. (1965). Optimum technical change in an aggregative model of economic growth. *International Economic Review*, 6(1):18–31.

Valiant, L. G. (1984). A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142.

Van den Bosch, K. and Cantillon, B. (2008). Policy impact. In Goodin, R. E., Moran, M., and Rein, M., editors, *The Oxford Handbook of Public Policy*, pages 296–318. Oxford University Press, Oxford (UK).

Van Duijn, J. J. (2013). *The Long Wave in Economic Life*. Allen & Unwin, London (UK).

Van Stel, A., Carree, M., and Thurik, R. (2005). The effect of entrepreneurial activity on national economic growth. *Small Business Economics*, 24(3):311–321.

Van Stel, A. and Storey, D. (2004). The link between firm births and job creation: Is there a upas tree effect? *Regional Studies*, 38(8):893–909.

Vandenbussche, J., Aghion, P., and Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *Journal of Economic Growth*, 11(2):97–127.

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

Vanhanen, T. (2012). *Polyarchy Dataset 2.0*. International Peace Research Institute, Oslo (NO).

Vapnik, V. (1982). *Estimation of Dependences Based on Empirical Data*. Springer, Berlin (DE).

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer, New York (US).

Vapnik, V. and Chervonenkis, A. (1964). A note on one class of the algorithms of pattern recognition. *Automation and Remote Control*, 25(6):821–837.

Vapnik, V. and Chervonenkis, A. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264–280.

Vapnik, V. and Lerner, A. (1963). Generalized portrait method for pattern recognition. *Automation and Remote Control*, 24(6):774–780.

Vapnik, V. N. (1998). *Statistical Learning Theory*. John Wiley & Sons, New York (US).

Voitchovsky, S. (2005). Does the profile of income inequality matter for economic growth? *Journal of Economic Growth*, 10(3):273–296.

Voitchovsky, S. (2009). Inequality and economic growth. In Nolan, B., Salverda, W., and Smeeding, T. M., editors, *The Oxford Handbook of Economic Inequality*, pages 549–574. Oxford University Press, Oxford (UK).

Walker, L. D. (2009). Delegative democratic attitudes and institutional support in central america. *Comparative Politics*, 42(1):83–101.

Wang, G. and Ma, J. (2012). A hybrid ensemble approach for enterprise credit risk assessment based on support vector machine. *Expert Systems with Applications*, 39(5):5325–5331.

Way, B. M. and Lieberman, M. D. (2010). Is there a genetic contribution to cultural differences? collectivism, individualism, and genetic markers of social sensitivity. *Social Cognitive and Affective Neuroscience*, 5:203–2011.

Weber, M. (1905). Die protestantische Ethik und der Geist des Kapitalismus. *Archiv für Sozialwissenschaft und Sozialpolitik*, 21:1–110.

Weil, D. (2005). *Economic Growth*. Addison-Wesley, Boston (US).

Whorf, B. (1956). *Language, Thought and Reality*. MIT Press, Cambridge (US).

Williams, P. J. (1990). Elections and democratization in Nicaragua: The 1990 Elections in perspective. *Journal of Interamerican Studies and World Affairs*, 32(4):13–34.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics*, 126(1):25–51.

Wong, P., Ho, Y., and Autio, E. (2005). Entrepreneurship, innovation and economic growth: Evidence from gem data. *Small Business Economics*, 24(3):335–350.

World Bank (2012a). *Education Statistics (EdStats)*. World Bank, Washington D.C. (US).

World Bank (2012b). *World Development Indicators Database (WDI)*. World Bank, Washington, D.C. (US).

World Bank (2012c). *World Development Indicators (WDI) database*. World Bank, Washington D.C. (US).

World Bank (2014a). *The Database of Political Institutions*. The World Bank, Washington D.C. (US).

World Bank (2014b). *World Development Indicators Database (WDI)*. The World Bank, Washington D.C. (US).

World Bank (2016). *World Development Indicators Database (WDI)*. The World Bank, Washington D.C. (US).

World Bank (2017). *World Development Indicators (WDI)*. The World Bank, Washington D.C.

World Intellectual Property Organization (2013). *WIPO statistics database*. World Intellectual Property Organization, Geneva (CH).

World Value Survey (2014). *World Value Survey Wave 6, official aggregate v.2015041*. World Value Survey Association, Madrid (ES).

Wüst, A. M. (2005). Jamaica. In Nohlen, D., editor, *Elections in the Americas: A Data Handbook*, volume 1, pages 423–438. Oxford University Press, Oxford (UK).

Youtie, J., Iacopetta, M., and Graham, S. (2008a). Assessing the nature of nanotechnology: Can we uncover an emerging general purpose technology? *The Journal of Technology Transfer*, 33(3):315–329.

Youtie, J., Shapira, P., and Porter, A. L. (2008b). Nanotechnology publications and citations by leading countries and blocs. *Journal of Nanoparticle Research*, 10(6):981–986.

Zagorski, P. W. (2003). Democratic breakdown in Paraguay and Venezuela: The shape of things to come for Latin America? *Armed Forces and Society*, 30(1):87–116.

Zhang, J. and Zhang, J. (2005). The effect of life expectancy on fertility, saving, schooling and economic growth: Theory and evidence. *The Scandinavian Journal of Economics*, 107(1):45–66.

Why are some nations rich and others poor? What are the sources of long-run economic development and growth? How can living standards be increased? In this book, Klaus Gründler empirically analyses these central economic questions and puts a particular emphasis on the role of technology, inequality, and political institutions. To substantiate his empirical studies, he introduces a new method to compute composite measures and indices that is based on mathematical algorithms from the field of machine learning.

Würzburg University Press

ISBN 978-3-95826-072-6

9 783958 260726