UNICEF
Innocenti Research Centre

Innocenti Working Paper

INSTITUTIONS, INEQUALITY AND GROWTH: A REVIEW OF THEORY AND EVIDENCE ON THE INSTITUTIONAL DETERMINANTS OF GROWTH AND INEQUALITY

Richard Bluhm and Adam Szirmai

IWP-2011-02

May 2011

unite for
children

unicef

Innocenti Working Papers

UNICEF INNOCENTI RESEARCH CENTRE

The UNICEF Innocenti Research Centre in Florence, Italy, was established in 1988 to strengthen the research capability of the United Nations Children's Fund and to support its advocacy for children worldwide. The Innocenti Research Centre (formally known as the International Child Development Centre) generates knowledge and analysis to support policy formulation and advocacy in favour of children; acts as a convener and catalyst for knowledge exchange and strategic reflections on children's concerns; and supports programme development and capacity-building.

Innocenti studies present new knowledge and perspectives on critical issues affecting children, informing current and future areas of UNICEF's work. The Centre's publications represent contributions to a global debate on child rights issues, and include a range of opinions. For that reason, the Centre may produce publications that do not necessarily reflect UNICEF policies or approaches on some topics.

The Centre collaborates with its host institution in Florence, the Istituto degli Innocenti, in selected areas of work. Core funding for the Centre is provided by the Government of Italy and UNICEF. Additional financial support for specific projects is provided by governments, international institutions and private sources, including by UNICEF National Committees, as well as by UNICEF offices in collaborative studies.

For further information and to download or order this and other publications, please visit the IRC website at http://www.unicef-irc.org

Correspondence should be addressed to:

UNICEF Innocenti Research Centre
Piazza SS. Annunziata, 12
50122 Florence, Italy
Tel: (+39) 055 20 330
Fax: (+39) 055 2033 220
Email: florence@unicef.org

Acronyms and abbreviations

2SLS	Two-stage Least Squares
GDP	Gross Domestic Product
cGDP	Gross Domestic Product per Capita (PPP)
Gini	Gini coefficient of inequality
GMM	Generalized Method of Moments
ICRG	International Country Risk Guide
ILS	Indirect Least Squares
IV	Instrumental Variables
LPM	Linear Probability Model
OLS	Ordinary Least Squares
PPP	Purchasing Power Parity
TFP	Total Factor Productivity (growth)
USA	United States of America
WDI	World Development Indicators
WGI	Worldwide Governance Indicators

Institutions, inequality and growth: a review of theory and evidence on the institutional determinants of growth and inequality

Richard Bluhm [a] and Adam Szirmai [b]

[a] Maastricht Graduate School of Governance, Maastricht University, e-mail: richard.bluhm@maastrichtuniversity.nl
[b] UNU-MERIT, and Maastricht Graduate School of Governance, Maastricht University, e-mail: szirmai@merit.unu.edu

Summary: This paper provides a detailed review of selected contributions to the study of the interrelationships between institutions, inequality and economic growth. We focus on the works of Engerman and Sokoloff, and Acemoglu, Johnson and Robinson for the study of long-run growth, as well as Rodrik for bridging the gap from long-run to short-term growth. In addition, we review a wide array of supplementary econometric evidence and criticisms. The emphasis of this review is on identifying differences and commonalities in the underlying theories of economic development, proposed causal mechanisms and econometric specifications. We contrast the findings by using a sources-of-growth framework which distinguishes between ultimate, intermediate and proximate causes of growth, as well as socioeconomic outcomes.

Keywords: growth, institutions, inequality, and development

Any remaining errors and omissions are entirely the responsibility of the authors. The statements in this paper are the views of the authors and do not necessarily reflect the policies or the views of UNICEF, Maastricht University, UNU-MERIT or affiliated institutions.

Table of contents

1.   INTRODUCTION

The difference in the development experiences between the most developed countries and the least developed countries of today is vast. While the inhabitants of Luxembourg enjoyed an income per capita of $77,783.5 in 2007, people in Liberia only had an income per capita of $385.6, which is close to subsistence levels.[1] Luxembourg's per capita income is 200 times larger than Liberia's. Even within the developing world, growth is very unequal. East Asia and parts of Latin America are growing at impressive rates, while many other countries – especially in Sub-Saharan Africa – struggle with sluggish and volatile growth. Likewise, the degree of income inequality varies sharply from the developed world to the developing world, as well as among developing countries and entire regions. Latin America is infamous for being the world's most unequal region, while the catch-up countries in East Asia are much more equal in terms of incomes but often unequal in terms of political access. In Africa, income inequality is smaller than in Latin America but poverty is widespread and political exclusion is a widespread phenomenon.

While growth theory has proved a powerful device for understanding the proximate sources of growth, empirical investigations have shown that much of growth still remains unexplained by factor accumulation.[2] Total factor productivity growth (TFP) in growth models and growth accounting is a catch all concept for everything but primary inputs. Parts of TFP can be attributed to technological change, but it often remains a black box item, modeled exogenously. In new growth theory and evolutionary theory, technological change is endogenized and the question shifts to what determines the rate of technological change. This opens the door to institutional analysis. In this spirit, we argue that development must rather be viewed in historical perspective linking historically shaped institutions, political and economic inequalities, and long-run development paths to development outcomes and development chances today.

In this paper, we provide an analytic review of the recent literature on the relationships between institutions, inequality and economic growth in the long-run, as well as the links between institutions, inequality and policy and growth in the short and medium term. We focus on the contributions to the study of long-run growth of Engerman and Sokoloff (1997, 2002), Acemoglu, Johnson and Robinson (2001, 2002) and their critics. We also discuss the contributions of Rodrik, Subramanian and Trebbi (2004) and Hausmann, Pritchett and Rodrik (2005) to the study of long-run and short-run growth. We analyze the underlying frameworks, models, empirical methods and data used to test the theoretical hypotheses.

There is now a consensus that institutions matter for growth, but disagreement about how, the extent to which this is the case, and which institutional arrangements affect growth more than others. Over time, early institutionalism (e.g. Veblen, 1899; Commons, 1936; Mitchell,

---

[1] This comparison is based on purchasing power parity adjusted real GDP per capita in 2007 from the Penn World Tables version 6.3 (in 2005 prices). Luxembourg is the richest non-oil-based economy and Liberia is the poorest country for which there is data available in the PWTs in 2007. The subsistence level of GDP per capita is often assumed to be around $400 in 1990 prices.

[2] There is plenty historical and contemporary evidence for this (see, for example, Solow, 1956, 1957; Denison, 1967; Abramovitz, 1989; Solow, 1991; Easterly and Levine, 2001; Hoff and Stiglitz, 2001).

1910), post-WWII institutionalism (e.g. Gruchy, 1947, 1978; Kapp, 1950; Hirschman, 1958; Myrdal, 1968) and later New Institutional Economics (e.g. North and Thomas, 1973; North, 1990; Williamson, 1985) have offered varying but often complementary approaches to linking differences in growth performance to institutions.

Neither the early institutionalists nor neoclassical economists placed specific emphasis on intra-country income inequality. In early neoclassical economics, inequality was often seen as promoting growth through higher savings by the rich and positive incentive effects (the equity versus efficiency debate).[3] Further, Kuznets (1955) hypothesized that inequality first rises and then falls during the development process. Initially inequality increases as the result of urbanization, increasing urban-rural income disparities and a higher degree of inequality in the urban economy than the rural economy. Kuznets's (1955) assumption was that causality runs more from development to the income distribution and not from inequality to growth, so that in the long run high inequality is only a transitory outcome at middle-income levels. This theory is known as the Kuznets curve. Similarly, the earlier institutionalist literature mainly focused on factors other than inequality, such as catch-up based on the advantages of technological backwardness and overcoming institutional deficiencies (Gerschenkron, 1962), escaping low-equilibrium traps though a big push (Rosenstein-Rodan, 1943), or forward and backward linkages and economies of scale (Hirschman, 1958). The exception was Myrdal (1968), who saw inequality as an obstacle for growth, as rigid class and status structures inhibit an efficient allocation of labor and talent.

Since the 1970s, inequality became more widely seen as a possible barrier to growth, as newer theories suggested that a more equitable distribution could contribute to positive growth dynamics (Chenery, Ahluwalia, Duloy, Bell and Jolly, 1974). But especially modern theory since the 1990s formalized how, given credit-market imperfections, inequality is a constraint to human capital investments (Galor and Moav, 2004), entrepreneurship (Banerjee and Newman, 1993) and intergenerational mobility (Galor and Zeira, 1993). In addition, political economy theory suggested that inequality directly harms growth through raising demand for taxation over time (Alesina and Rodrik, 1994; Persson and Tabellini, 1994). These theories are accompanied by a vast and conflicting empirical literature on the direct effects of income inequality on development.[4]

Economic inequality did not feature centrally in the institutional literature until Engerman and Sokoloff (1997), who found that economic inequality in the age of colonization adversely affects suffrage, schooling, banking and other institutions to this very day. Their approach links factor endowments, inequality and institutions to long-run development outcomes. Acemoglu, Johnson and Robinson (2001) emphasize that political inequality and historical junctures matter. They relate the quality of colonial institutions set up by European settlers to the mortality rates of these settlers to explain subsequent development trends. Both theories

---

[3]This argument can be traced to Keynes (as discussed in Thirlwall, 2005) and was also made in the earlier works of Sen (1960). There is a tradition of this line of reasoning. For example, Lewis (1954) provided a model of economic development with unlimited supplies of labor. In the model, exploitation of cheap labor and inequality are directly linked to growth through the assumption that the rich capitalist classes are responsible for savings, investment and accumulation.

[4]See, for example, Perotti (1996), Deininger and Squire (1998), Barro (2000), and Forbes (2000).

offer explanations for the puzzling "reversal of fortune" – the phenomenon that formerly relatively rich colonies have now become poorer and relatively poor colonies have become richer (Acemoglu, Johnson and Robinson, 2002).

These studies, however, have provoked a critical debate. In the case of Acemoglu et al. (2001) in particular, Albouy (2008) strongly criticized the underlying empirical methodology. Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004) argued that human capital matters more for development than institutional explanations. Further, the research of Rodrik et al. (2004) weighs in on the debate on the primacy of different "ultimate causes" of growth, which is often understood as an interplay between geography, institutions and integration into international trade. Rodrik also links studies of long-run and short-run growth by exploring growth collapses (Rodrik, 1999), growth accelerations (Hausmann, Pritchett and Rodrik, 2005) and by developing methods for growth diagnostics (Hausmann, Rodrik and Velasco, 2005, 2008; Rodrik, 2010).

In the following review of this more recent literature, we use the sources-of-growth framework by Szirmai (2008, 2012) in order to identify the most pertinent mechanisms, organize the literature, and analyze the interrelationships between inequality, institutions, and economic growth. In this framework, we distinguish between proximate, intermediate, and ultimate causes of growth, as well as socio-economic outcomes.

Our review is guided by three broad working hypotheses. *First*, the institutional arrangements that shape human behavior are among the crucial determinants of the long-run economic growth performance of nations since 1500. *Second*, political and economic inequality affects growth in two important ways: (a) indirectly by influencing both the nature of institutions which in turn have an influence on long-run growth, and, (b) directly by influencing the accumulation of human capital, one of the important proximate sources of growth. *Third*, past institutional arrangements affect the degree of contemporary economic and political inequality.

The paper is organized as follows. Section 2 briefly presents our framework. Section 3 reviews the authors/topics in the terms of their own theory and models, and presents some of the most important criticisms. It consists of three subsections, for the works of (3.1) Engerman and Sokoloff, (3.2) Acemoglu, Johnson and Robinson, and (3.3) Rodrik. Section 4, the synthesis, discusses the results from section 3 building on the framework presented in section 2. Section 5 concludes.

2.   THE FRAMEWORK

Figure 1 illustrates our conceptualization of ultimate, proximate and intermediate causes of growth including socio-economic outcomes. The distinction between proximate and ultimate sources of growth has been developed by several authors such as Maddison (1988), Abramovitz (1989) and more recently by Rodrik (2003). The addition of intermediate causes and socio-economic outcomes is based on Szirmai (2008).

Why is this framework useful? First, it allows us to highlight the different levels of growth analysis. On the one hand, studies referring to long-run growth, or levels of GDP per capita, usually refer primarily to ultimate causality though sometimes trying to include variables capturing proximate and intermediate causality. On the other hand, studies of growth in the short and medium term mainly refer to intermediate and proximate causes and may attempt to link these with contemporary socio-economic outcomes.

Figure 1: Proximate, intermediate, and ultimate sources of growth and development

Ultimate sources of growth
· Institutions (political & economic)
· Geographic conditions
· Demographic characteristics
· Class and power-relationships
· Historical shocks
· Culture and attitudes
· Long-run science & technology cycles
· Distance to technological frontier

Intermediate sources of growth
· Economic, technology & social policy
· Demand trends (economic cycles)

Proximate sources of growth
economic actors
$$Y = f[(K, L, R)^e] + A + P$$

Socio-economic outcomes
· Standards of living and consumption
· Poverty and inequality
· Health and education
· Environmental sustainability

Source: adapted from Szirmai (2008).

Second, using the framework we can visualize the concept of endogeneity, which is mostly a function of time and interdependencies with other variables. Analyses of ultimate causality are most challenged by endogeneity, as nearly every factor but geography is endogenous in the long-run. However, analyses of proximate causality also have to deal with endogeneity issues. For instance, as Rodrik (2003) notes, capital accumulation and efficiency in the use of resources are themselves endogenous. Causality may well run backwards from growth to accumulation and productivity. Third, it helps to clarify that development is a nonlinear process, subject to simultaneity and circular causation (Myrdal, 1968) at almost every level, as is evident from the many feedback relationships in Figure 1. This is closely related to the problem of endogeneity. Fourth, it allows us to distinguish between two important aspects of development: growth of productive capacity (GDP, cGDP) and socio-economic outcomes. In the following, we briefly review the components of this framework.[5]

---

[5] See Szirmai (2008) and also the forthcoming second edition of "The dynamics of socio-economic development" by Szirmai (2012) for a much more detailed discussion of the framework.

The *proximate sources of growth* (Denison, 1967; Maddison, 1987, 1988) are directly measurable sources of output growth, or, in other words, the inputs into the production function (for both classical and endogenous growth theory). We understand the equation in Figure 1 as the result of decisions of a variety of heterogeneous economic actors responding to constraints and incentives provided by policies and the institutional framework. In Figure 1, Y refers to output, K, L and R refer to the primary factors of production capital, labor and natural resources, the exponent e refers to the efficiency with which the primary factors are used to turn intermediate inputs into final goods and services.[6] The term A denotes net income from capital investments and labor from abroad (net factor income) and P refers to colonial plunder and expropriation (negative) or voluntary transfers and development aid (positive).

Once we have quantified the proximate sources of growth, we can subsequently explore their links with the wider economic and social sources of growth and development. *Intermediate sources of growth* refer to two types of factors: first, trends in domestic and international demand and, second, economic, social and technology policies. Policies include a wide range of interventions such as trade policy, macroeconomic policy, industrial policy or subsidies to stimulate innovation and industry. They also include all kinds of social policies in the area of social protection/insurance, education and welfare, which affect the distribution of the fruits of growth. Including demand as an intermediate factor in this framework shifts the emphasis away from conceiving of growth in the medium term and short term as only supply-side driven.

Underlying both the proximate and intermediate sources, there are even "deeper" factors, which we call the *ultimate sources of growth*. These include economic, political and social institutions, institutional change, historical shocks, geographic conditions, long-run trends in scientific and technological knowledge, demographic conditions and trends, culture, basic social attitudes and capabilities, changes in class structures and relationships between social groups, and long-run developments in the international economic and political order. Many of these themes are analyzed in this review. For example, the critical junctures approach of Acemoglu, Johnson and Robinson (2001, 2002) and much of their subsequent research emphasizes the historical shocks of colonization in conjunction with demographic and geographic factors, as well as the dynamics of the relationships between elites and the mass of the population.

Socio-economic outcomes are what ultimately matter in development. However, we argue that the most fundamental engine of development, especially in historical perspective, is sustained increases in productive capacity and output growth over long stretches of time. This statement holds although many contemporary outcomes can be positively or negatively altered even in the absence of growth and is not intended to downplay the feedback channels

---

[6] The concept of efficiency as used here refers to everything that increases output per unit of primary input. It includes economies of scale, efficient allocation of the factors of production within sectors (appropriate choice of technology), efficient allocation between less productive and more productive economic sectors (structural change), reallocation of resources towards more dynamic sectors (structural change), efficient allocation between countries (specialization and comparative advantage), utilization of capacity and, last but not least, disembodied technological change. Disembodied technological change refers to changes in the state of our knowledge which cannot be measured through changes in the quality of capital and labor.

that connect outcomes to proximate and ultimate causes. The degree to which productive capacity is transformed into desired social outcomes depends on the nature of social and economic policy (intermediate causality), institutions, and initial levels of social inequality (ultimate causality). This interaction between outcomes (inequality) and ultimate causes (geography and institutions) is, for example, the basis of the theory of Engerman and Sokoloff (1997), whose contribution we review in the following section.

3. THE REVIEW

3.1 The Engerman and Sokoloff hypothesis

In a series of papers focusing on the divergent developmental experiences of the New World,[7] Engerman and Sokoloff developed a controversial theory which has received considerable attention in the modern literature. They focus on the very long-run growth outcomes and regard economic inequality as the core factor shaping the institutions that account for the take-off of North America throughout the course of the nineteenth century and the subsequent relative decline of South America.

Table 1: GDP per capita in selected New World economies

	GDP per capita relative to the U.S.			
	1700	*1800*	*1900*	*1997*
Argentina	-	102%	52%	35%
Barbados	150%	-	-	51%
Brazil	-	50%	10%	22%
Chile	-	46%	38%	42%
Cuba	167%	112%	-	-
Mexico	89%	50%	35%	28%
Peru	-	41%	20%	15%
Canada	-	-	67%	76%
United States (1985$)	$550	$807	$3,859	$20,230

Source: Sokoloff and Engerman (2000).

They begin with a puzzle. During the 17th century there was parity in incomes between many colonies in South and North America. Some Southern colonies such as Cuba and Barbados even had higher per capita incomes than the USA (Table 1). Engerman and Sokoloff (1997, 2002) argue that the North was initially not economically attractive to early colonizers. Only later was this trend reversed, owing to differences in the institutional structures created during colonization. Until around 1700 the Southern colonies were very successful in raising GDP per capita and specialized heavily according to their comparative advantage in primary products (such as sugar). At the same time, they created institutions based on high inequality and limited access to economic opportunities. In contrast, the greater homogeneity of the population in the North was reflected in the genesis of political and social institutions which allowed for broad-based access to economic opportunities and which encouraged human

---

[7] See Engerman and Sokoloff (1997, 2002, 2005), Sokoloff and Engerman (2000), and the additional sources cited in the text.

capital accumulation.[8] Engerman and Sokoloff (1997) consider growth the cumulative impact of incremental economic advances of many individuals participating in the economy. In the long run, with increasing economic diversification, access to economic opportunities for a broad range of the population became the driving force behind continuous innovation and growth. This contrasts with a development path dominated by small and restrictive elites in the South. Hence, institutional developments offer a credible explanation for both the timing and the scale of the reversal in prosperity between North and South in the course of the 19th century.

While all former colonies had a high marginal product of labor in common, the factor endowments of the Southern regions made them more suitable for growing and extracting sugarcane, minerals and other high-value commodities during the early colonial period. A region's climate and quality of the soil determined the most profitable commodity, and the size and density of the existing native population determined the initially available workforce. Sugarcane exhibited large economies of scale and was most efficiently processed in large plantations exploiting native and imported slave populations. As slave trade was free and priced in international markets, Engerman and Sokoloff (1997) argue that the sheer amount of slaves imported plus the fact that relatively more of them went to the South taken together allows us to conceive of slaves simply as a highly mobile production factor, flowing to regions of high demand and profit.

In the South, it was precisely these factor endowments and the extreme inequality resulting from a small European elite governing a largely poor and enslaved population, which proved detrimental to the emerging institutional structures. In the North, the relative large and homogeneous European population relying on small-scale farming with little or no slavery created institutions favorable to later economic development. Hence, they reject theories linking development outcomes to the identity and culture of the colonizer[9] and instead argue that "the colonies that later came to make up the United States and Canada were quite unusual in the New World, because their factor endowments (including climate, soils, and the density of the native populations) predisposed them towards paths of development with relatively equal distributions of wealth and human capital and greater population homogeneity as compared with the great majority of their hemispheric neighbors" (Engerman and Sokoloff, 2002, p. 56).

In the Americas and the Caribbean, the different factor endowments resulted in three broad clusters of countries. The first cluster (Barbados, Cuba, the West Indies, Saint Dominguez and Brazil) was characterized by large sugar plantations (e.g. *Latifundia*), a high percentage of immigrant slaves, and, as a result, a small European elite. The countries in the second cluster (the Spanish colonies of Mexico and Peru) were well-endowed with minerals and had

---

[8] The price of homogeneity among the population of European descent was the complete marginalization and even elimination of the indigenous Indian population.

[9] For instance, Hartz (1964) explains the divergent development of South and North America by referring to the cultural differences between colonists from the Iberian peninsula and northwestern Europe, which were magnified in the new world. The colonists from the Iberian Peninsula transferred regressive institutions such as *Latifundia* to South America, while the Northern European colonists transferred institutions which were more conducive to economic development.

a tradition in mining. Mines also exhibit economies of scale but their workforce consisted mainly of a large native population in coerced labor rather than imported slaves and contract laborers. Combined with immigration restrictions enforced by the Spanish, these colonies also ended up with relatively small European elites.[10] The third cluster are the colonies that later became the United States and Canada (in particular those north of Chesapeake Bay). They were neither endowed with a warm climate and soil suitable for the production of sugar, nor was there a large native population.[11] These are similar clusters – albeit with a new meaning – as in Fieldhouse's (1982) distinction between plantation colonies (e.g. Brazil), mixed colonies of settlement (e.g. Mexico), and pure colonies of settlement (e.g. USA).

3.1.1   Institutions of (in)equality

Central to their hypothesis is that precisely these factor endowments and patterns of colonial rule resulted in institutions which first adversely affected and then maintained an unequal distribution of wealth, human capital, and access to economic opportunities. Relying mainly on qualitative evidence, they give examples of how these inequalities manifested themselves in six particular institutional spheres: *suffrage, schooling, land policy, taxation, patents, and banking.*

The pace of the *extension of the franchise* is their most crucial and direct evidence of economic inequality creating political inequality. While all colonies restricted the right to vote to the white male population, the North quickly abandoned wealth and literacy requirements. Engerman and Sokoloff (2005) attribute this to the greater homogeneity among the (white male) population in the North. In short, comparatively equal people demanded comparatively equal rights and would eventually get them. By comparison, the disparities within the Southern population allowed elites to close-off access to the ballot and maintain selection criteria based on correlates of status. In 1880, 18.3 percent of the U.S. population voted in secret regardless of their wealth or literacy, while in Ecuador, Mexico, Peru, Uruguay and Venezuela only a very small percentage of the population voted. There, access to the ballot was restricted by wealth and literacy.

Likewise, *access to schooling* also displays a strong divergence across the hemisphere (Engerman and Sokoloff, 2002; Mariscal and Sokoloff, 2000). At latest by the mid-1800s, the Northern colonies all had public, general tax financed, and universally accessible primary schools. The wealthy colonies in South America, however, failed to develop broad schooling institutions and even the most progressive colonies trailed the North by almost 75 years in

---

[10] In this category, it was mainly the practice of the Spanish to distribute land rights to small elites (the system of *encomiendas* and *haciendas*) and the limitations placed on immigration to their colonies that created a structure very similar to the first group. The two land regimes are often subsumed under the term *Latifundia*. Even when economies of scale in production were absent, the agricultural and industrial structure remained concentrated. However, the size of the European population was somewhat larger than in the pure plantation economies such as Brazil.

[11] However, the south of the United States was suitable for tobacco, cotton and other valuable commodities with economies of scale. Consistent with their theory, slavery is widely known to have been prevalent in the South. Nevertheless, according to Engerman and Sokoloff, the south is a special case as it inherited a large part of the institutions from the north through national legislation.

terms of literacy. Limited access to schooling directly reinforced limited access to economic opportunities and via literacy requirements it limited access to the ballot box.

The differing *land policy regimes* across the Americas also point to differences in the institutionalization of economic inequality. The more homogeneous white settler populations in the U.S. and Canada benefitted from an institutionalized policy of granting land to small holders, however, often at the expense of the native Indian population. In 1900, nearly 75% of U.S. household heads owned land, and in 1901 in Canada, almost 90% of all household heads were land owners. In Latin America, landholdings were highly concentrated and large landownership predominated.

Further, Sokoloff and Zolt (2007) suggest that Latin America's reliance on consumption taxes and comparatively regressive *tax systems* are the result of elites resisting an increased tax burden on wealth, income or property – all of which are more progressive. This pattern extends further to patents and banking. Contrary to Latin America and the Caribbean, the U.S. *patent system* evolved particularly early, with low access barriers and strict enforcement of property rights (Khan and Sokoloff, 1998).

Differences in *banking institutions* can also be traced to the colonial period. Farmers and planters were already providing loans to each other in the early 18th century. In colonies with large estates, this exchange was limited to narrow elites, while in the Northern colonies a higher percentage of the population could provide collateral (Engerman and Sokoloff, 2002). After the U.S declaration of independence, the federal structure together with the broad franchise (lower barriers to participation), then engendered a diverse and competitive banking system in which bank chartering was a routine administrative affair. In Latin America, by contrast, chartering banks was tightly controlled and restricted to a narrow elite often associated with the national governments, resulting in few banks and limited access to credit.

3.1.2   The model

To summarize, the causal mechanism proposed by Engerman and Sokoloff (1997, 2002) is a combination of exogenous factors predetermining a development path, based on greater or lesser economic and political inequality, and endogenous institutional dynamics that maintain path-dependence over time. Exogenous factor endowments (climate, soil, mining resources and native population) determined the initial conditions of the colonies during European conquest from 1492 to 1700. At one extreme, a tropical climate, very fertile soil or mineral resources and a large non-European population were favorable to growing and extracting sugar, cotton, and other high-value commodities. The arriving Europeans used domestic and foreign slave labor to extract these resources. The size of the domestically available unskilled labor force and the imported unskilled slave labor influenced the relative share of Europeans, who simultaneously were the economic (and later political) elites. At the other extreme, a dispersed native population combined with soils and climates suitable to growing wheat or similar commodities created conditions in which plantation slavery was of little use, small farming was more efficient and, as a result, Europeans represented a large proportion of the population. On the basis of these initial allocations, endogenous institutional dynamics

evolve. Countries with homogeneous populations,[12] large European elites and low levels of inequality, develop institutions which provide broad access to political, educational and economic opportunities (e.g. broad franchise, accessible public schooling, easy access to capital and jobs). Broad access to opportunities in turn maintains lower degrees of inequality and promotes economic growth by providing more incentives to larger segments of the population. Initially homogeneous countries have higher human capital accumulation, broadly accessible savings and investment institutions, and better protection of property rights for both intellectual capital and land. As a result, the social and private returns to investment are more closely aligned (North and Thomas, 1973). In countries with heterogeneous populations, high degrees of inequality and small European elites, the elites created institutions of unequal access (e.g. limited franchise, limited schooling, and limited property rights for the non-elite population) to capture economic opportunities. These diverse colonial experiences matter even today, as institutional path dependence and the reinforcing features of higher or lower inequalities created time-persistent institutions.

Figure 2: The causal link of inequality, institutions and long-run growth
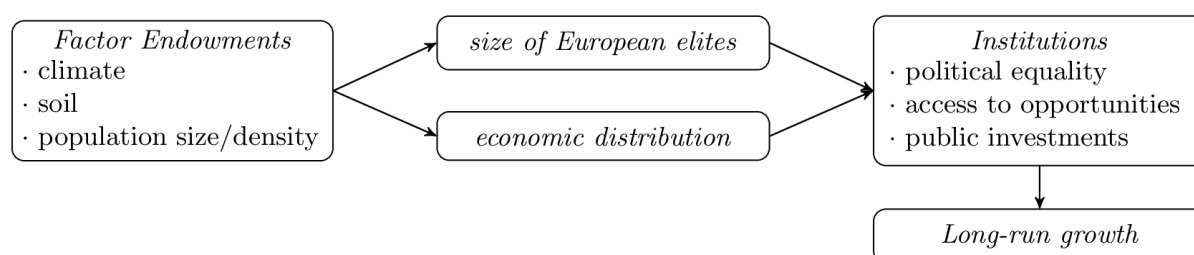
Source: authors' illustration.

Engerman and Sokoloff (1997, 2002) argue that the effects of economic and political inequality are intrinsically linked. In their theory, economic inequality often brings about political inequality. When the colonizers arrived in South America, it was the *de facto* economic inequality vis-à-vis the native population which then became institutionalized in the *de jure* structure of the political system and institutionalized access to economic and political opportunities. However, they make the qualification that this relationship only seems to hold in countries were unskilled labor was abundant (Engerman and Sokoloff, 2005). In conditions of abundant unskilled labor (natives or slaves), scarce capital and scarce skilled labor (Europeans), it is economic inequality rather than political inequality that results in the emergence of institutions that favor the owners of scarce factors of production.

The theory has both strengths and weaknesses. On the one hand, it is historically rooted and based on detailed country narratives without the loss of theoretical generality. On the other hand, it is questionable how well the hypothesized mechanisms apply outside the New World. The later colonies of occupation in Asia (Fieldhouse, 1982) followed a very different logic. Colonizers were intervening less in the existing institutional structures of indigenous societies. Further, Africa was colonized comparatively late in the 19th century but has been a

---

[12] In this context, homogeneity has two different meanings which need to be distinguished. On the one hand, homogeneity refers to more egalitarian societies, and on the other hand, it refers ethnic or racial homogeneity, where the European settlers have succeed in marginalizing or even eliminating the indigenous population. In many 'heterogeneous' countries in South America relatively more of the indigenous population has survived.

net slave exporter since the 15th century. Other confounding factors can be pointed out for Asia or even Europe. However, Engerman and Sokoloff (1997, 2002) do not claim the theory to be universal. This raises the question whether the causation running from economic inequality to political inequality is just one of many possible mechanisms. Untangling the concepts of economic inequality, political inequality and institutions is necessary to examine these relationships. This distinction is only beginning to emerge in the literature and poses additional difficulties in cross-country research, especially for indicators and theories in which access to institutions (political inequality) is a key feature of 'high quality institutions'.

3.1.3 Criticism and econometric evidence

The mechanism proposed by Engerman and Sokoloff (1997, 2002) proved controversial and provoked a series of empirical investigations seeking to confirm or discredit the relationship between economic inequality, institutional development and growth. We concentrate on the recent cross-country studies of Easterly (2007) and Nunn (2008a), combined with additional evidence the case study offered by Acemoglu, Bautista, Querubin and Robinson (2007) for Cundinamarca, Columbia. The key issue in all of these studies is to what extent economic inequality is really the causal factor that determines institutions and growth, or whether political inequality or the existence of slavery as such are alternative explanations of economically inefficient institutional structures and economic stagnation.

Easterly (2007) operationalizes the theory of Engerman and Sokoloff (1997, 2002) and uses the exogenous variation introduced by climate and soil to directly test the causal link between intra-country income inequality, the level of GDP per capita, the quality of institutions, and schooling levels. Building on Engerman and Sokoloff's argument that the cultivation of wheat had positive effects and growing sugar had negative effects on economic inequality, he derives a novel wheat-sugar suitability ratio as an instrumental variable (IV) for income inequality. Instrumental variables are commonly used to identify the direct effect of an endogenous variable on an outcome in one causal direction, without actually observing it or being able to estimate it directly. As the degree of inequality is in part an outcome of the growth process itself, it is subject to reverse causality. Moreover, it is measured with great imprecision. Using a valid instrumental variable introduces exogenous variation in income inequality, which can be used to overcome endogeneity, isolate the causal effect and shift the problem of measurement error away from the instrumented variable to the instrument. Easterly (2007) calls the variation introduced by this instrument "structural inequality", i.e. inequality which reflects historical events captured by the wheat-sugar suitability ratio including conquest, slavery and land distribution by the state or the colonial power. He tests the following two-stage least squares (2SLS) specifications:

$$y_i = \mu + \alpha I_i + \mathbf{X}'_i \gamma + \varepsilon_i \tag{1}$$

$$I_i = \zeta + \beta \ln W_i + \mathbf{X}'_i \delta + \upsilon_i \tag{2}$$

where y_i is the outcome variable of interest (log GDP per capita, institutions, or schooling), I_i is the Gini coefficient of income inequality, W_i is the log wheat-sugar suitability ratio, and $\mathbf{X}'_i$ is the transpose of a vector of covariates affecting all variables.

The results of the regression analyses for GDP per capita are striking. A one standard deviation increase in the (instrumented) Gini coefficient leads to a 1.1 standard deviation reduction in income levels. Similarly, institutional quality[13] declines by one standard deviation and schooling[14] by 1.3 standard deviations. The specifications are robust to various changes and additions, such as controlling for natural resources, climate, and colonial/legal origin dummies. He concludes that structural inequality has a large and significant direct effect on GDP per capita and an indirect effect through its negative impact on institutions and schooling. This is in line with the Engerman and Sokoloff (1997) hypothesis. Nevertheless, at least three caveats are worth noting. First, Easterly uses a sample comprising the whole world, thereby extending a theory based solely on the New World colonies to virtually every country's development path regardless of its idiosyncratic circumstance and history.[15] Second, the robustness of cross-sectional cross-country instrumental variables regressions is debatable and depends strongly on the quality of the instrument used. Third, the causal mechanism is only tested indirectly and depends on the channels hypothesized by Engerman and Sokoloff (1997, 2002). Structural political inequality that coincides with economic inequality today could be driving this relationship, without necessarily having originated from economic inequality.

A second and more detailed study aiming to investigate the Engerman and Sokoloff (1997) hypothesis was conducted by Nunn (2008a). His analysis approaches the theory on multiple levels but focuses on slavery resulting from endowments favorable to larger scale farming as the primary mechanism in determining inequality and growth outcomes. He only analyzes colonies in the New World and does not claim universal applicability of the theory. Consequently, his approach has several advantages, such as the use of rich and detailed data for relevant countries, and a focus on the precise mechanism of slavery induced inequality. However, without instrumental variables inferences of causality are merely tentative and could be the result of either meaningful or spurious relationships depending on the unobserved factors.

In a first examination of a set of 29 New World colonies, Nunn (2008a) finds that the fraction of slaves in the total population in 1750 has a significant and large effect on GDP per capita in 2000. His baseline model begins with the assumed reduced form relationship between slavery, population density and income directly:

$$\ln y_i = \mu + \pi \frac{S_i}{P_i} + \varphi \frac{P_i}{A_i} + \mathbf{D}'_i \gamma + \varepsilon_i \qquad (3)$$

---

[13] Easterly (2007) measures the quality of institutions by using the average across all composite indicators of the World Bank's Worldwide Governance Indicators (WGI) developed by Kaufmann, Kraay and Zoido-Lobatón (1999). Easterly (2007) also uses their separate dimensions but finds that the results remain very similar across the board, which he attributes either to the inadequacy of the indicators in identifying separate dimensions, or to the effects of a dominant elite simply being similarly detrimental for all types of institutions.

[14] Schooling is measured as the average of secondary enrollment rates from 1998-2003 with data from the World Development Indicators (WDI).

[15] Interestingly, the exclusion of the Americas increases the negative effects of the instrumented inequality coefficient as opposed to weakening the relationship.

where $\ln y_i$ is the log of GDP per capita, S_i/P_i is the slave population as a fraction of the total population, P_i/A_i is the total population per unit of arable land, and $\mathbf{D}'_i$ is the transpose of a vector containing the country of origin of the colonizer.[16]

The effect of slavery on institutions is illustrated in the partial correlation plot below.[17] He provides an example to illustrate the scale. Jamaica had 90% of its population enslaved in 1750 and has a GDP per capita of \$3,640 (in 2000). If Jamaica's proportion of slaves would have been only 46% (close to the Bahamas) then GDP per capita would be more than 200% higher today (approximately \$11,580). These results seemingly confirm the basic premise of the Engerman and Sokoloff (1997) hypothesis.

Figure 3: Partial correlation plot of slavery in 1750 and income in 2000
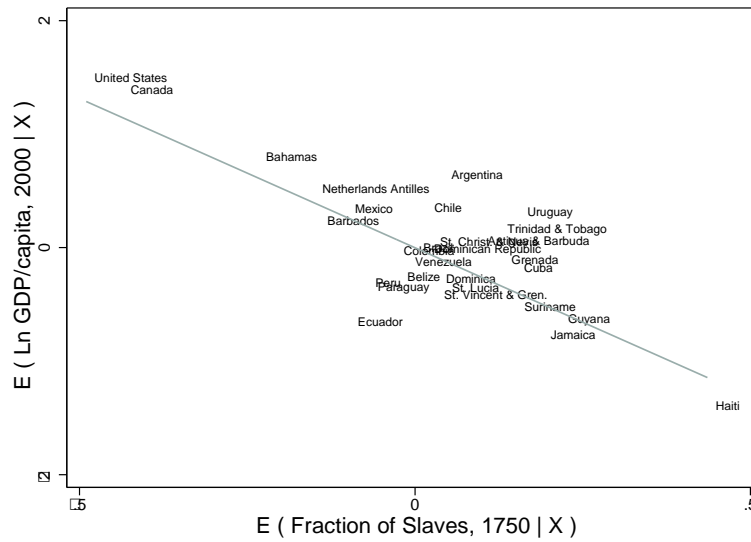
Source: computed using data from Nunn (2008a)

However, in a second examination of the British West Indies only, Nunn (2008a) casts doubt on the proposed mechanism of higher inequality in countries with widespread plantation slavery. Restricting the sample to the British colonies in the West Indies allows for the use of richer data on the size of the plantations and numbers of slaves, and indirectly controls for heterogeneity by concentrating on a more homogeneous group of colonies with similar characteristics. Nevertheless, this also reduces the sample size to a mere 12 countries. He modifies the specification in two ways to differentiate between plantation and non-plantation slavery (3) and to distinguish by size of slave holding (4):

$$\ln y_i = \mu + \pi_P \frac{S_i^P}{P_i} + \pi_{NP} \frac{S_i^{NP}}{P_i} + \varphi \frac{P_i}{A_i} + \varepsilon_i \tag{4}$$

$$\ln y_i = \mu + \pi_S \frac{S_i^S}{P_i} + \pi_M \frac{S_i^M}{P_i} + \pi_H \frac{S_i^H}{P_i} + \varphi \frac{P_i}{A_i} + \varepsilon_i \tag{5}$$

---

[16] Colonizer dummies are central the Engerman and Sokoloff (1997) hypothesis, as Spanish colonies had fewer slaves but nevertheless high inequality (see the description of country clusters).

[17] The outliers are easily identified on the graph. Nunn (2008a) removes the obvious candidates to test the robustness of his specification. Omitting the USA and Canada (countries with lower slave proportions) weakens the relationship, but removing Haiti (with an extremely high proportion of slaves) does not alter the strength.

13

where notation is as before. In (4), the indices denote plantation slavery (P) or non-plantation slavery (NP). In equation (5), (S) is defined as the ratio of slaves on small slave holdings to total population, (M) as the ratio for medium size slave holdings and (H) as the ratio for large slave holdings.

Differentiating between non-plantation slavery and large plantation slaves in 1830, Nunn (2008a) finds that the former, rather than the latter, has the most detrimental effect on development. The effect of non-plantation slavery is nearly twice as large as the effect of planation slavery. Similarly when differentiating by the size of the slave holdings, small holdings have a nearly four times higher effect than medium sized holdings, and two times higher effect than large holdings.[18] Contrary to the Engerman and Sokoloff (1997) hypothesis, Nunn (2008a) concludes that the institution of slavery per se, rather than the size of the slave holdings, predicts negative effects on economic development.

Finally, Nunn (2008a) uses data at the state and county level in the U.S. to verify this relationship and examine whether causality runs from plantation slavery to economic inequality and subsequent GDP levels. He runs two separate simple OLS regressions:

$$I_i = \mu + \pi \frac{S_i}{P_i} + \varphi \frac{P_i}{A_i} + \varepsilon_i \tag{6}$$

$$\ln y_i = \mu + \delta I_i + \pi \frac{S_i}{P_i} + \varphi \frac{P_i}{A_i} + \varepsilon_i \tag{7}$$

where notation is as before and I_i is the Gini coefficient of land inequality in 1860. The first specification used data from 1860 only and the second specification changes the dependent variable to log GDP per capita in 2000.

Two main effects emerge. First, in the USA the effect of slavery on development is negative but only differs minimally between small and larger slave holdings. Second, using the Gini coefficient of land inequality, the fraction of slaves, and population density in 1860, he confirms that slavery caused economic inequality. However, when regressing per capita income in 2000 on land inequality, the fraction of slaves and the population density in 1860, slavery independently has a highly significant negative effect, while inequality is even positively related to income per capita. For Nunn (2008a), these findings contradict the Engerman and Sokoloff (1997) hypothesis, which states that inequality negatively influences income levels and implies that the coefficient of slavery should become insignificant once inequality is accounted for. However, Nunn gives insufficient recognition to the fact that the North-South divide in the U.S. is put forth as a special case in the original argument. According to Engerman and Sokoloff (2002), the south was generally unsuitable for sugar cultivation and hence the share of slave plantations and total use of slaves was never as great as in the Caribbean or Brazil. Further, many of the key institutions in the southern states were determined nationally after 1864 and through competition with other states in the union. As a result, the South of the USA became more competitive and open than its counterparts with a

---

[18] Small holdings refers to 10 or less slaves, medium size holdings refers to 11 to 200 slaves, and large size holdings refers to 201 or more slaves. The size of the holding is highly correlated with non-plantation or plantation slavery and larger slave holdings cluster with sugar, coffee and tobacco plantations.

stronger legacy of slavery in South America (Engerman and Sokoloff, 2002, p. 86). Therefore, causal inference based on the USA only might be problematic at best.

Further emphasizing the independent effects of any kind of slavery on institutions and development, Nunn (2008b) shows in a related paper on Africa how the export of slaves is negatively associated with current economic performance of the countries of origin. This is an interesting extension of the body of work by Engerman and Sokoloff (1997, 2002), as not only the use of slaves but also their production and export were detrimental for a country's institutions and development paths. The causal mechanisms in slave exporting countries are quite different. Slavery hindered the formation of larger ethnic identities, contributed to ethnic fractionalization,[19] and subsequently led to underdeveloped political institutions. However, slavery is only one of the factors influencing the development of African political institutions, which were also influenced by very different patterns of political centralization and nation formation.

Acemoglu et al. (2007) study the state of Cundinamarca, Columbia, to directly test whether economic or political inequality shaped the region's institutional structure and long-run growth outcomes. They construct four measures of inequality: a Gini coefficient of landownership in 1897 and 1890, an overall Gini coefficient for landownership including non-land owners, an index of political concentration (operationalized as the number of individuals having held mayoral office over the number of times a mayor has been appointed between 1875 and 1895), and an index of overlap between land inequality and political concentration. They estimate the impact of these variables on both long-term outcomes (primary/secondary enrollment, urbanization, and poverty in 1993) and medium-term socio-economic outcomes (literacy, urbanization, and access to non-educational public goods in 1937). Their baseline model is:

$$y_i = \mu + \delta I_i + \phi Q_i + \varphi O_i + \psi L_i + \mathbf{X}'_i \gamma + \varepsilon_i \tag{8}$$

where I_i is average land inequality in 1890 and 1897, Q_i is political concentration in the period from 1875 to 1895, O_i is the overlap measure, L_i is contemporary land inequality and $\mathbf{X}'_i$ is the transpose of a vector of covariates affecting all variables.

In most of their regressions on contemporary outcomes, a higher land Gini is *positively* associated with schooling and urbanization, but negatively associated with poverty. However, political concentration is negatively associated with schooling and urbanization,[20] but robustly correlated with higher poverty. For medium-term outcomes, the only robust link is a negative association of political inequality and literacy in 1937. In these specifications too, the land Gini often enters positively. The effect of the overlap measure is very small and insignificant in most specifications. They repeat the exercise without the overlap measure and with the overall land Gini (including non-land owners) in addition to the traditional land Gini. The overall land inequality has a negative sign but remains mostly insignificant. The land Gini is positively associated with the outcome variables, yet often insignificantly. Political

---

[19] Ethnic fractionalization is an obstacle to development in its own right (Easterly and Levine, 1997).
[20] The effects of land inequality and political constraints are only significant at or above the 10% level for the secondary schooling and basic needs (poverty) regressions.

inequality remains negatively associated with the outcomes and is sometimes significant. For the poverty and access to public services outcomes all signs reverse, as before. Further, they show that political leaders disproportionately amassed more wealth with every year in power and that the probability of becoming a land-owner is higher for politicians than the probability of becoming a politician for landholders.

The pattern clearly contradicts the theory put forth by Engerman and Sokoloff (1997), which emphasizes the primacy of economic over political inequality. Acemoglu et al. (2007) conceptualize the results based on weak versus strong institutionalization of the polity,[21] meaning the strength of institutional constraints placed on political actors. Cundinamarca had few constraints on political actors and it was easy for them to consolidate their hold on power. In some regions, a separate land-holding elite could thus prove a critical political counterbalance and check on political concentration, as indicated by the positive effects of high land inequality vis-à-vis the negative effects of political concentration. This explanation opens up a plurality of possible interactions between economic inequality, political inequality, and elite configurations.

In sum, out of the many relationships reviewed here some prove very robust and others raise issues requiring further research. Slavery is detrimental to institutions and growth both for slave importing and exporting countries, regardless of the size of slave holdings, or if in the form of plantation or non-plantation slavery. Whether this effect is an independent effect of slavery, or if it works through economic inequality, or political inequality, or any combination of these is uncertain. Future studies of long-run growth need to distinguish between political and economic inequality. Further, economic inequality can lead to political inequality and hinder development but the conditions under which this is the case must be strictly identified. Political inequality by itself can be a considerable barrier to schooling and development. Further, while case studies add to our understanding of the processes involved they also introduce considerable complexity into the reasoning, which challenges parsimonious theory.

3.2   The critical junctures hypothesis

The institutions and growth studies of Acemoglu, Robinson, Johnson, and collaborating authors concentrate on three broad themes: institutions and long-run growth in former European colonies, formal theories of dictatorship, democracy and development, and empirical analyses of democratization. We concentrate on two of their seminal papers (Acemoglu et al., 2001, 2002), as these offer an explicit theory of development for former colonies and establish a causal link running from institutions to growth. Their models of dynamic games between citizens and elites are used to corroborate and specify the underlying mechanisms, conceptualize the role of inequality, and broaden the scope to the origins of regime types in general. Further, we review their critical work on the modernization hypothesis as an alternative theory of development and highlight some of the criticisms of

---

[21] This concept draws on a previous paper by Acemoglu, Robinson and Verdier (2004) and incorporates insights from Bates (1981).

their approach and theory put forth by Albouy (2008) and Glaeser, La Porta, Lopez-de Silanes and Shleifer (2004).

3.2.1 Colonial origins

In an influential paper, Acemoglu, Johnson and Robinson (2001) develop a theory and provide evidence of the reasons behind the diverging paths of comparative development in former colonies. They argue that the institutions set up by former colonial powers differ vastly, ranging from "extractive states", as in the Belgian Congo, to the "neo-Europes" of the United States, Canada, Australia and New Zealand. In their view, these different institutional configurations are due to the varying sizes of European settlements relative to the native population, which were in part determined by the influence of the actual and perceived local disease environment on Europeans. For example, the local population in Africa and the Caribbean was partly immune to yellow fever and malaria, while as much as 80% of European deaths in the tropics can be attributed to these two diseases (Curtin, 1989). As a result, colonies with low settler mortality were predisposed to become colonies of settlement with inclusive institutions and strong protection of property rights, whereas in colonies in which Europeans had lower survival chances extractive states emerged. According to Acemoglu et al. (2001), Europeans essentially brought their institutions to where they could, or created extractive institutions where they could not settle in large enough numbers. Applying the categories of Fieldhouse (1982), the theory essentially states that pure colonies of settlement inherited institutions of private property rights, while mixed colonies of settlement and plantation colonies resemble points on a continuum towards political and economic insecurity.[22] Acemoglu et al. (2001) argue that key features of the institutions set-up during colonization persist until today, even though the political systems of former colonies underwent many changes since. Akin to the earlier work of Engerman and Sokoloff (1997), their theory focuses on the initial conditions Europeans faced in the colonies and how these predetermined highly dissimilar development paths. This "critical junctures" approach emphasizes the role of historical factors in shaping institutions, the political system and development outcomes. However, contrary to Engerman and Sokoloff (1997, 2002), they do not stress factor endowments and inequality, but settler mortality, as the determining factor of the size of European settlements relative to the native population. Following Acemoglu et al. (2001), the model can be summarized as a system of equations:

$$\ln y_i = \mu + \alpha R_i + \mathbf{X'}_i \gamma + \varepsilon_i \tag{9}$$

$$R_i = \lambda_R + \beta_R C_i + \mathbf{X'}_i \gamma_R + v_{R_i} \tag{10}$$

$$C_i = \lambda_C + \beta_C S_i + \mathbf{X'}_i \gamma_C + v_{C_i} \tag{11}$$

$$S_i = \lambda_S + \beta_S \ln M_i + \mathbf{X'}_i \gamma_S + v_{S_i} \tag{12}$$

---

[22] Interestingly, Acemoglu et al. (2001) do not distinguish between periods of colonization. Implicitly, trading posts or colonies of occupation had similar effects on institutions as mixed or plantation colonies. However, in the latter two Europeans intervened heavily in the indigenous structures, while in the former the influence of Europeans was intentionally marginal. The theory treats extractive institutions in Africa, which was colonized late and only briefly, equivalent to those in Latin America, which was colonized early. This time effect is only captured indirectly, though high (potential) settler mortality in large parts of Africa effectively deterring Europeans from large-scale settlement.

where y_i is the log of GDP per capita for country i, R_i is a measure of "current institutions", C_i is a measure of "early institutions", S_i is a measure of people of European descent, M_i is the log mortality rate of the settlers, $\mathbf{X}'_i$ is the transpose of a vector of covariates affecting all variables and the γ's are the coefficient vectors. Only M_i, R_i, y_i and parts of $\mathbf{X}'_i$ are actually observed.

The advent of European colonialism can be regarded as a natural experiment of history. Acemoglu et al. (2001) propose an innovative instrument for exploiting this historical juncture to estimate the causal effect of institutions on national income. They argue that settler mortality is exogenously determined by geographic factors and should not be systematically correlated to any unobserved factors influencing development today in any other way than through institutions. Therefore, it can be used to isolate the variation in institutions due to differences in settler mortality and to infer the causal effect of institutions on income levels. Their preferred measure of current institutions (R_i) is an index of protection against the risk of expropriation (averaged from 1985-95), which assesses the strength of property rights[23]. Their settler mortality (M_i) data is mostly taken from Curtin (1989, 1998) and Gutierrez (1986). To operationalize the theory, they test a two-stage least squares model consisting of equation (9) and the following collapsed version of equations (12) to (10) as the first stage specification:

$$R_i = \zeta + \beta \ln M_i + \mathbf{X}'_i \delta + v_i \tag{13}$$

The results point to a very large and highly significant effect of property rights institutions on long-run economic performance. In their baseline estimate, the resulting coefficient is 0.94 with a standard error of 0.16. They provide an example to illustrate the scale. The difference between Chile and Nigeria is 2.24 points on the expropriation index and they are "typical" in the sense that they are close to the regression line. The expected difference in GDP between Nigeria and Chile is 7.24-fold while in reality the distance is 11.46-fold. Hence, the difference in institutions explains more than 60% of the difference in economic performance between these two countries. The strength of the relationship is illustrated in the regression plot of the instrumented (predicted) values of the expropriation index and the logarithm of GDP per capita in 1995 in Figure 4 below.

The results are robust to controlling for the identity of the colonizer, religion, climate, soil quality, natural resources, landlocked countries, diseases, and ethno-linguistic fragmentation. The identity of the colonizer has been argued to be a decisive determinant of current institutions (e.g. La Porta, Lopez-de-Silanes, Shleifer and Vishny, 1999). However, Acemoglu et al. (2001) find that this effect is only relevant for British origin and just about significant at the 5% level. They conclude that Britain primarily colonized places where settler mortality allowed larger settlements relative to the native population and verify that the coefficient on institutions remains about the same when investigating former British

---

[23] Knack and Keefer (1995) first published this comprehensive index encompassing many institutional features (rule of law, repudiation of contracts, corruption in government and the quality of bureaucracy) based on data from the International Country Risk Guide (ICRG). Previous studies used revolutionary coups or assassinations to proxy for the risk of expropriation/property rights, but produced questionable country rankings. The ICRG data and the Knack and Keefer (1995) index have since become standard use in the literature.

colonies only. Contrary to the hypothesis of – *inter alia* – McArthur and Sachs (2001), geography and climate seem to have no independent effects on GDP per capita once institutions are treated as endogenous.

Figure 4: Plot of predicted expropriation risk and GDP per capita
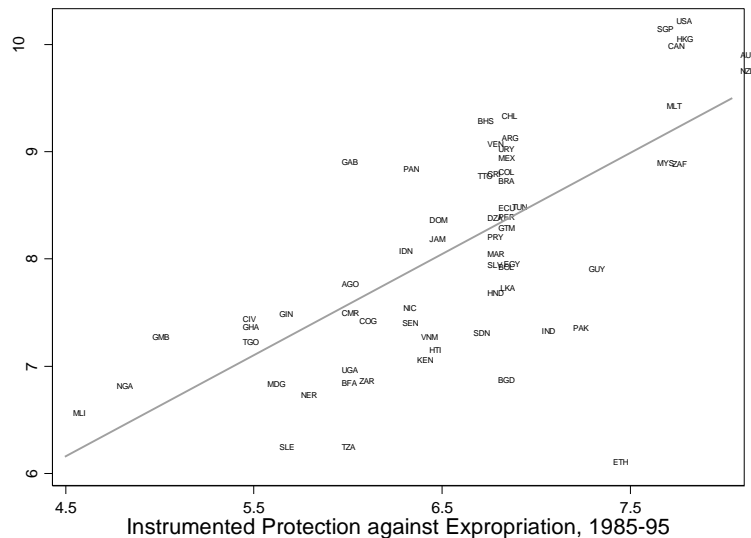
Source: computed using data from Acemoglu et al. (2001)

3.2.2 *Reversal of fortune*

In a second major contribution, Acemoglu, Johnson and Robinson (2002) systematically document and analyze a reversal in income per capita among former European colonies. They argue that, particularly during industrialization, institutions can be causally linked to this reversal and extend their theory to incorporate population density as a determinant of initial conditions. In addition, they cast further doubt on theories linking modern development outcomes to geographic factors.

While the data on per capita income in 1500 are fragmentary estimates at best, historical accounts suggest that many of the pre-colonial civilizations in South America were comparatively richer than those in North America but also than New Zealand and Australia.[24] Acemoglu et al. (2002) argue that urbanization and population density can be used as proxies to measure prosperity before the advent of colonization. In their view, urbanization is a direct measure of development, as it required an advanced network of transportation and agricultural surplus to be sustainable. To validate this assumption they show how highly urbanization and income are correlated when considering both cross-sectional and panel data since 1913. However, in theory population density is less strongly linked to GDP per capita. This weaker link is, for example, explained in Malthus' classic argument. On the one hand, Malthus associates growth of population with increasing standards of living, but, on the other hand, he also stresses the checks and balances of famine and hunger as food production fails

---

[24] The reversal essentially took place among colonies that later became known as the Western Offshoots and all others. For GDP per capita estimates see, for example, Engerman and Sokoloff (1997), Sokoloff and Engerman (2000), and also Coatsworth (1999) or Table 1 presented earlier.

to keep up with population growth.[25] In recent cross-sections, population density is not associated with prosperity, which Acemoglu et al. (2002) attribute to the changed nature of the relationship between income and number of children. Nevertheless, population density and urbanization in 1500 are highly correlated, which for Acemoglu et al. (2002) justifies the use of both in the analysis, in spite of the Malthusian principle.

Figure 5: Urbanization/log population density in 1500 and log GDP per capita in 1995
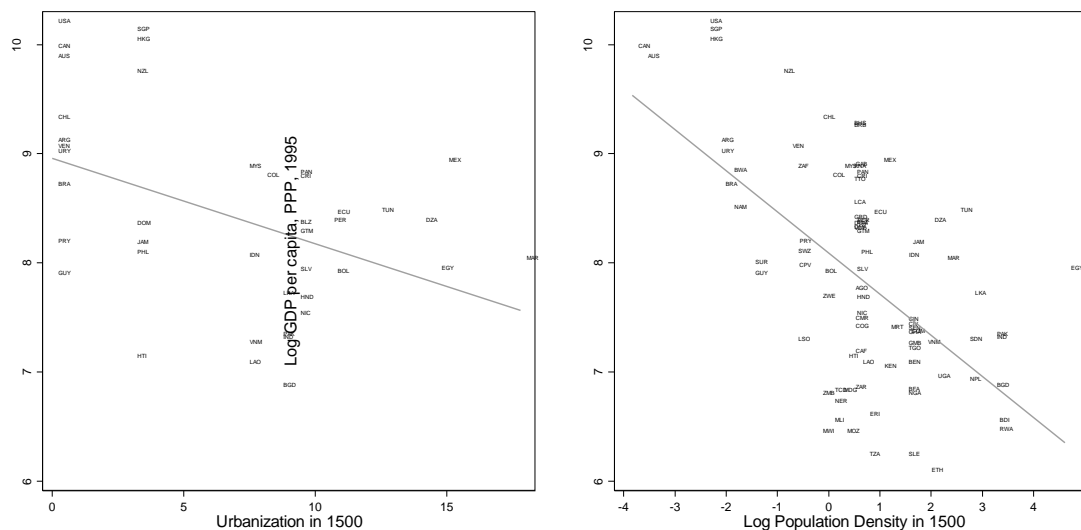
Source: computed using data from Acemoglu, Johnson and Robinson (2002).

They find that urbanization in 1500 and income per capita in 1995 are significantly negatively correlated (see Figure 5 above). The estimated coefficient on urbanization is -0.078. Consequently, a 10% decrease in urbanization results in an approximately two times higher GDP per capita in 1995. These results just about account for the current income difference between, for example, Uruguay and Guatemala. They repeat the analysis using log population density in 1500. The coefficient is -0.38 and is highly significant. A 10% increase in population density results in 4% lower per capita income in 1995. The results are robust to various controls, instrumenting urbanization with population density, changes in the sample, and alternative assumptions. In most extended specifications, the coefficients change only minimally. When both population density and urbanization are included, urbanization enters positively but insignificantly, while population density enters negatively and significantly. Interestingly, when examining countries that were never colonized the relationship between urbanization or population density and GDP per capita is *positive*, confirming the relevance of colonialism as a natural experiment or critical juncture. Acemoglu et al. (2002) place the timing of the reversal at the onset of industrialization, which they corroborate by showing that the great divergence in urbanization rates, industrial production and per capita income

---

[25] Specifically, higher living standards led to quicker ovulation in women, more successful pregnancies and more surviving children/adults. Before 1800 higher standards of living meant faster population growth, but these were preludes to so-called Malthusian catastrophes, such as the Great Famine (1315-1317) and the Black Death (1346-1351). Faster population growth also increases the scarcity of resources and land and reduces output per worker. Food supply could not keep up with population growth and, as a result, the standards of living declined again until population growth averaged zero. Much of the institutions literature is focused on how property rights, innovation and efficient production created the conditions for escaping these pre-industrial dynamics.

between the Western Offshoots and all other colonies did not occur until the turn of the nineteenth century. Before the 19th century many colonies had higher urbanization rates (per capita income) than the U.S., Canada, Australia and New Zealand.

How can this reversal be explained? Acemoglu et al. (2002) argue that neither the "simple" nor the "sophisticated" geography hypotheses, which have been put forth in different variants by many authors (e.g. Lewis, 1978; Myrdal, 1968; Sachs, 2001; Diamond, 1997), can account for this phenomenon. According to Acemoglu et al. (2002), the simple geography hypothesis suggests that time-invariant factors (such as natural resources, a coastline, and good conditions for agriculture or health) have lasting effects on development. In the view of Acemoglu et al. (2002), proponents of the sophisticated geography hypotheses, in turn, argue that time-variant geographic factors influence development. These are, for example, an interaction of the most suitably grown crop with plowing technology, or the interaction of geographically determined transport costs and industrialization.[26] Acemoglu et al. (2002) succinctly summarize the two hypotheses as follows:

$$Y_{i,t} = \alpha_0 + \alpha_1 \cdot G_i + \nu_t + \varepsilon_{i,t} \tag{14}$$

$$Y_{i,t} = \alpha_0 + \alpha_1 \cdot G_{i,t} + \alpha_2 \cdot T_t \cdot G_{i,t} + \nu_t + \varepsilon_{i,t} \tag{15}$$

where $Y_{i,t}$ is GDP per capita in country i and time t, G_i are time invariant geographic characteristics, T_t is the state of technology at time t, $G_{i,t}$ are time variant geographic characteristics, ν_t is a general time effect, and $e_{i,t}$ are country-time specific effects. The simple version (14) concentrates on α_1, while the sophisticated version (15) argues that α_2 has the most important effect.

Like Engerman and Sokoloff (1997), Acemoglu et al. (2002) suggest that, contrary to the geography hypothesis, it is in fact European colonialism which led to the reversal of incomes. They define two criteria as central for growth-enhancing institutions: well-defined private property rights (aligning private and social returns)[27] and inclusive institutions (enabling broad participation in productive opportunities). When secure property rights are only applicable to a wealth-owning or political elite they are not sufficient for lasting development. Their key argument is that "European colonialism not only disrupted existing social organizations, but led to the establishment of, or continuation of already existing, extractive institutions in previously prosperous areas and to the development of private property institutions in previously poor areas" (Acemoglu et al., 2002, p. 1263).[28]

---

[26] There are very elaborate arguments and models behind what Acemoglu et al. (2002) call the "geography hypothesis"; we follow their simplification here as we are mainly concerned with the robustness of the Acemoglu et al. (2002) model. For more detail see the original authors as referred to in the text above, but also McArthur and Sachs (2001). Many of these authors find distinct roles for geography and institutions. However, econometrically the debate has centered on establishing if geography has an independent effect on per capita income or if it is entirely captured by institutions. The possible (direct or indirect) influence of geographic determinants is widely acknowledged and is evident in the inclusion of a variety of geographic controls in virtually all of the model specifications in this line of research.

[27] This is the main point of North and Thomas (1973).

[28] It is not entirely clear how a continuation of already existing and the new establishment of extractive institutions can both be determinants of the reversal at the same time. Strictly following the logic of Acemoglu

Extending their earlier work (Acemoglu et al., 2001), they now identify two initial conditions as relevant determinants of the development paths of former colonies. On the one hand, the initial population density determined how much labor was available that could be enslaved or coerced to work in agriculture or mining. Densely populated areas were also more highly developed and often had a functioning tax system, which could be captured by the arriving Europeans. On the other hand, the feasibility of settlements (i.e. settler mortality rates) determined how large the proportion of European descent would be relative to the native population and in absolute numbers. In areas of low density and low settler mortality, European settlement in large numbers was easier. A larger relative quantity of Europeans also translated into a social stratification similar to their countries of origin, and the lower strata would demand rights comparable to those present in their country of origin. Interestingly, they attribute no distinct role to the differences in (weapons) technology between Europeans and native populations, which varied sharply from early colonization conquests to later campaigns (e.g. in Asia).

Acemoglu et al. (2002) test their theory utilizing a specification similar to that presented earlier in equations (9) and (13) with settler mortality serving as an instrument for institutions, but including either urbanization or the natural log of population density in 1500 as additional explanatory variables. Their results show that both measures of early prosperity become insignificant once institutions are endogenously determined, while in all specifications the coefficient of institutions remains relatively large, positive and highly significant.[29] They conclude that this strongly suggests that the strength of property rights institutions accounts for the reversal.

To further investigate why this change in relative incomes occurred during the late 18th and early 19th century, they hypothesize that countries with better property rights protection and more inclusive institutions were better able to capitalize on the opportunity to industrialize. According to Acemoglu et al. (2002), three mechanisms could have potentially barred countries with elite institutions and low property rights from industrializing quickly. Insecure property rights for non-elites could have prevented sufficient entrepreneurial investments, elites could have intentionally blocked industrial investments as the returns would have benefitted non-elites, and new technologies might bring about political discontent or threats to elite power. Using panel data and either a country's industrial output or per capita income as their dependent variable, they test this hypothesis in two ways. First, they use estimates of UK industrial output as a proxy for the opportunity to industrialize and interact it with their measure of institutions.[30] Second, they instrument their institutions measure with an

---

et al. (2002), it can be argued that a continuation of already existing extractive institutions does not change the status quo and renders colonialism irrelevant as an institutional intervention. These two can only be reconciled if industrialization is the key determinant of the reversal and colonialism matters only in shifting the institutional set-ups in some but not necessarily all colonies. In Acemoglu et al. (2002), this distinction is not always apparent.

[29] They report their results using three different measures of institutions, namely average protection against expropriation risk (1986-1995), constraint on the executive in 1990, and constraint on the executive in the first year after independence. The coefficient on these measures ranges from 0.37 to 0.88, depending on the measure used and the additional controls.

[30] The measure of institutions is "constraint on the executive" from Gurr's Polity III database, as it has a long

interaction of log settler mortality and UK industrial output.[31] For both strategies and dependent variables, the coefficient on the interaction term of institutions and UK industrial output is large and significant in most specifications (the magnitude is in the order of 0.132 to 0.206 for industrial output and 0.078 to 0.159 for log GDP per capita). They extend their models by allowing for an interaction between industrial output and geography (latitude) according to the time-variant geography hypothesis in equation (15). Its coefficient is insignificant by a large margin in all specifications.

3.2.3 The model

Combining theory and evidence from the two papers presented above, the model of Acemoglu et al. (2001, 2002) can be summarized schematically (see Figure 6 below). As a preliminary caveat, they explicitly acknowledge that such a parsimonious theory is only possible in the colonial context and that some of the reported relationships are very different – *if not the opposite* – in countries that were never colonized. Hence, colonialism is merely one of many critical junctures, albeit one of great significance.

Figure 6: The causal link of institutions and long-run growth
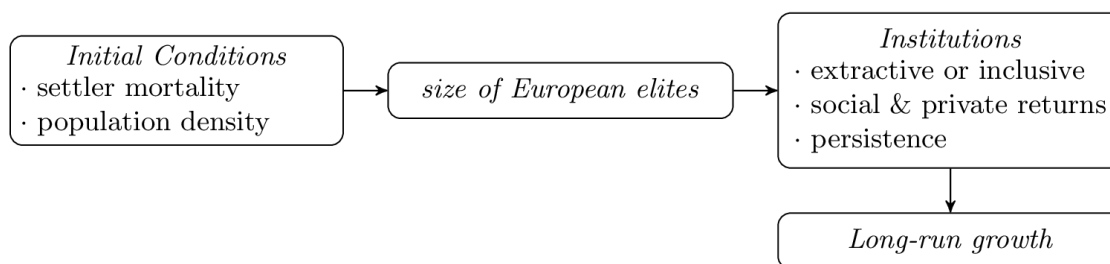
Source: authors' illustration.

The institutional structure and subsequent long-term growth outcomes in former colonies were severely affected by the initial conditions faced by the settlers. A dense indigenous population, relative prosperity and comparatively high settler mortality led to "extractive" institutional structures aimed at transferring surplus produce and rents to Europeans. These societies were characterized by a small European elite or appointed indigenous elite, exclusive institutions, few constraints on the executive and underdeveloped property rights for a majority of the population. Their political and economic systems relied on coercion, hierarchy, frequently even dictatorship and deeply enshrined inequalities. In contrast, in regions that were sparsely populated, relatively poor and endowed with a disease environment favorable to settlement, the resulting institutional structures were non-coercive, allowed broad access, stronger protection of property rights, and limited the powers of the executive. Geography matters, but only in determining the initial conditions which in turn shaped early institutions. It has no independent effect apart from predisposing entire regions to different institutional paths. Endogenous institutional dynamics maintained the adverse

---

time-series dimension going back to the first year of independence.

[31] Their panel model specification is as follows: $y_{i,t} = \mu_t + \delta_i + \phi \cdot \bar{R}_i \cdot U_t + \varepsilon_{i,t}$ and $\bar{R}_i$ is instrumented using $\ln M_i \cdot U_t$, where $y_{i,t}$ is either industrial production per capita or GDP per capita of country i in year t, μ_t are time effects, δ_i are country effects, $\bar{R}_i$ is the average of institutions across all t, U_i is UK industrial output per capita, and M_i is the log of settler mortality.

characteristics of early colonial institutions throughout time, as elites had few incentives to change the underlying institutional structure for fear of losing power, or engaged in efforts to maintain power even when the structure of the political institutions changed. Their model is very similar to Engerman and Sokoloff (1997), but stresses mortality and initial density rather than factor endowments as determinants of the size of European settlements. Further, Acemoglu et al. (2001, 2002) emphasize the distribution of political power more than the distribution of economic resources in their explanations of the causal mechanism.

Acemoglu and Robinson (2000a,b, 2006, 2008) constructed several formal models to corroborate the mechanisms mentioned above and to expand their theoretical reach beyond former colonies. To illustrate the issue of persistence, Acemoglu and Robinson (2008) present a model in which citizens and elites are engaged in a contest for their favorite institutional structure (democracy and non-democracy). The model's main result is that democratic reform altering the *de jure* power of elites vis-à-vis citizens may be partially or entirely offset by efforts of the elites to invest more in *de facto* political power. In some cases, the greater advantage of citizens in democracy may even lead to such intense counterbalancing efforts by the elites (through bribes and other mechanisms), that the democratic arrangement is economically less efficient than non-democracy. Acemoglu and Robinson (2008) call this captured democracy – a state in which the political institutions are "pro-citizen" but the economic institutions are designed to serve the interests of the elite. In their model, only simultaneous political and economic reforms reducing the gains of elites from controlling political institutions make adverse outcomes considerably less likely.

Further, Acemoglu and Robinson (2000a) explore the conditions under which political elites will block technological progress. They argue that it is not just the erosion of economic rents for elites that motivates their resistance to technological progress, but the threat of losing political power. In their view, the economically powerful cannot block new technologies if they do not have political power, whereas those who have political power and expect to remain powerful have no incentives to block progress. Only those who have political power and fear losing it have an interest in, and the means for, blocking technological advances. They block progress in an effort to reduce uncertainty, because there is no credible commitment to compensate those that lose power after a change of the economic structure. Acemoglu and Robinson (2000a) apply this logic to the different rates of industrialization in Britain and Germany versus Austria-Hungary and Russia. In Britain and Germany, landed interests anticipated continued political influence and did not oppose industrialization even though it would reduce their economic rents. In Austria-Hungary and Russia, on the contrary, the landed elites regarded railroads and industry as a threat to political power.

To explore why elites extend the franchise and contribute to democratization even in the face of potentially losing power, Acemoglu and Robinson (2000b) formalize the trade-off between the threat of revolution and piecewise concessions of power. Franchise extension acts as a credible commitment towards future redistribution to the citizenry. The threat of social unrest depends on the degree of organization among the poor and a society's level of inequality. If the poor are too well organized, maybe contrary to intuition, they will be able to frequently pose a threat of revolution. Hence, they are powerful enough to credibly ensure future

24

redistribution to themselves. If the poor are well enough organized to pose a threat to the regime but not enough to do so continuously and the society is highly unequal, then social unrest is more likely and democratization becomes the only mechanism credibly guaranteeing future redistribution. At some levels of inequality, temporary distribution may momentarily stave off the threat of revolution. However, countries with continuously low inequality are slow to democratize, or will not become democratic at all, as the demand for redistribution is not high.

According to Acemoglu and Robinson (2000b), Germany, for example, met rising inequality and the threat of social unrest by expanding the welfare state supported by a large socialist party ensuring the credibility of redistribution. Only the shock of the First World War increased inequality and created social unrest to a point that democratization was inevitable. Consequently, Germany exhibited a delayed pattern of franchise extension. Britain, in contrast, was continuously faced with the threat of revolution by the middle and lower classes and temporary redistribution was not a credible option. To maintain political power, the elites extended the franchise in multiple waves to the middle classes. Acemoglu and Robinson (2000b) use these results and additional evidence from Britain, Germany, Sweden and France to give a new meaning to the Kuznets curve. Rising inequality is accompanied by the threat of revolution, which in many cases can only be met by extending the franchise in order to credibly assure future redistribution. In all four countries, there is some scant evidence that inequality peaked roughly at the same time as the franchise was extended and declined thereafter.

3.2.4   *Modernization or critical junctures?*

In a classic work, Lipset (1959) argues that certain prerequisites are necessary for democracy to arise, such as higher levels of income, broad education, and a capitalist economy. He identifies income, industrialization, education and urbanization as highly correlated with democracy, but is cautious to not impose linear causality for any one factor but assumes multivariate causality (Lipset, 1959, p. 105). In the social sciences, modernization theory has many facets. Economists often associate either with deterministic stage-theories of development (such as Rostow, 1959) or the simplified proposition that rising levels of income and/or education cause democratization.

Acemoglu, Johnson, Robinson and Yared (2007, 2008) provide cross-country evidence challenging modernization theory and argue that their critical junctures approach is better suited as a theory of democratization and development. Motivated by a large body of research and statistical evidence since the 1960s linking democracy to income levels, Acemoglu et al. (2007, 2008) are interested in the direction of the causal relationship. In fact, modernization theory is diametrically opposed to their own theory, which holds that institutions (including democracy) cause development and not vice versa.

Acemoglu et al. (2007, 2008) argue that previous studies[32] have based their conclusions on cross-sectional correlations only and do not establish causality. They present an extended

---

[32] Examples of such studies are Barro (1999) and Przeworski, Alvarez, Cheibub and Limongi (2000).

research design able to cope with serial correlation and reverse causality. To reduce serial correlation in their panel data ranging from 1960 to 2000, their estimates are not just based on annual, but also five year, ten year and twenty year intervals. They prefer fixed-effects, Anderson and Hsiao (1982), and Generalized Methods of Moments (GMM) estimators over simple pooled ordinary least squares (OLS) estimators to reduce unobserved effects and to better approximate the true relationship. Further, to identify the causal relationship and endogenize income per capita, they construct two dynamic panel instruments (lagged savings rate and trade-weighted world income). They estimate the effects of changes in income on changes in democracy rather than drawing conclusions from correlated levels only. The rationale for this is strong, as any post-WWII sample is likely to estimate a high between-country correlation of income and democracy. Today, most of the richest countries are also the most democratic. Fixed effects estimators instead focus on within-country variation over time. They test different variations of the following specification (with and without instruments):

$$d_{i,t} = \alpha d_{i,t-1} + \beta y_{i,t-1} + \mathbf{x}'_{i,t-1}\gamma + \mu_t + \delta_i + \varepsilon_{i,t} \qquad (16)$$

where $d_{i,t}$ is the democracy score of country i in year t, $d_{i,t-1}$ is the first lag in the democracy score to capture mean reversion, $y_{i,t-1}$ is the first lag of income, $\mathbf{x}'_{i,t-1}$ is a vector of covariates, μ_t is a set of time effects, δ_i is a set of country effects, and $\varepsilon_{i,t}$ is the country-time specific error term. Notation is in lower cases to represent first differences.

Opposite to the results of earlier studies, Acemoglu et al. (2007) find that there is no causal effect running from income to democracy. In all cases, they first report the pooled OLS estimates without fixed effects and find a positive coefficient on income,[33] corresponding with the existing paradigm in the literature. However, when controlling for fixed effects, the coefficient on income becomes very small and insignificant. The Anderson-Hsiao and GMM estimates even change the direction of the relationship. With both democracy measures the effect is negative, large, and insignificant in most of the specifications. These results are robust to sample changes and additional controls such as education, which enters insignificantly. The instrumental variables estimates using either the lagged savings rate or trade-weighted world income further corroborate that there is no causal effect from income to democracy. Almost all two-stage least squares or GMM estimates with either instrument result in a negative or insignificant coefficient on lagged income. They supplement this analysis by investigating a 500-year sample with simple pooled OLS, while controlling for historical factors (such as log population density, early institutions and the date of independence). Here too, their most comprehensive specification is able to remove any significant remaining partial correlation between income and democracy. Further, in the companion paper, Acemoglu et al. (2007) develop a double-hazard model of democratic transition which also fails to establish an effect of income on democracy.

---

[33] Using the Freedom House measure of democracy the coefficient of lagged income is 0.072 with a standard error of 0.010. Using the Polity measure of democracy the coefficient lagged income is 0.053 with a standard error of 0.010. Both results refer to the five-year panel including the first lag of democracy as an additional control.

3.2.5 Criticism and additional evidence

The empirical and theoretical explorations of Acemoglu et al. (2001, 2002) attracted several criticisms directed both at their methodology and theory. In the following, we concentrate on the criticisms of the settler mortality data and instrumental variables method posed by Albouy (2004, 2006, 2008) and the criticism of theory and method by Glaeser et al. (2004). These two pointed out some of the most pressing issues in the Acemoglu et al. (2001) research and other studies using similar techniques or data, but are certainly not the only voices critical of their contribution (e.g. Przeworski, 2004a,b).

Albouy (2004) seriously questions the coding and construction of the settler mortality series. He argues the data lack "geographical relevance, statistical precision, or comparability across countries" Albouy (2004, p. 2). Geographical relevance refers to the fact that Acemoglu et al. (2001) imputed mortality rates for missing observations based on data from other neighboring countries. Out of the 64 countries present in the original sample in Acemoglu et al. (2001), only 36 have unique and distinct mortality rates which originated in their geographical region. According to Albouy (2004), Acemoglu et al. (2001) use inconsistent and statistically imprecise rules in selecting mortality rates, particularly in terms of time (first or later rate), unit (soldiers, bishops, or laborers) and weighting of multiple data points. He argues that the mortality rates are also not comparable across countries, as Acemoglu et al. (2001) mix rates from European soldiers on military campaigns with rates of soldiers in barracks. For Albouy (2004), peace in the 19th century is positively correlated with income levels and the confounding of these two rates makes settler mortality endogenous to the specification.

Albouy (2004) constructs two alternative series based either on soldiers in barracks or on campaign, and compares the original Acemoglu et al. (2001) model with his data. The first stage significance of his adjusted settler mortality instrument is much lower than the original, leading to the "weak instrument" problem.[34] Using clustered AR standard errors (Anderson and Rubin, 1949) rather than traditional standard errors, he shows that once the weak instrument problem is accounted for, the confidence intervals on the estimated effect of institutions become unreasonably large and often include zero, negative infinity and/or positive infinity in many specifications. He also shows that when using the original data series with additional controls such as continent dummies and latitude, or mean temperature and minimum rainfall, the first stage relationship becomes insignificant and the second-stage AR confidence interval unbounded. Albouy (2004) concludes that while the theory may be credible, the empirical effect of institutions cannot be substantiated with the current settler mortality series.

The criticism of Albouy (2004) resulted in two rebuttals by Acemoglu, Johnson and Robinson (2005, 2006) and further investigations by Albouy (2006, 2008). Acemoglu et al. (2005, 2006) maintain in a point-by-point discussion of Albouy's modifications that their coding was not unreasonable or inconsistent and present new evidence supporting the

---

[34] The weak instruments problem mainly refers to the predictive strength of the instrument. If the instrument is not able to isolate substantial exogenous variation in the instrumented variable, then the estimator will be biased towards the OLS estimator and sometimes have a confidence interval as large as the entire real line.

mortality rates they used. Much of the dispute concerns the assignment of mortality rates to countries in Africa and Latin America. To circumvent the issue of assignments, Acemoglu et al. (2005, 2006) emphasize that their results become even stronger when excluding all African observations. However, Albouy (2008) responds that this statement is then only based on 11 unique observations. Acemoglu et al. (2005, 2006) argue that Albouy's distinction between soldiers in barracks or campaigns is not helpful, as it mixes very small campaigns and large warfare in the same variable. Instead, they argue that their approach of selecting the first available peacetime mortality rate has been applied consistently. They dismiss many of the modifications done by Albouy on the basis that he is selecting later mortality rates which are lower due to improvements in medicine and are not relevant proxies for early potential settler mortality. Further, Acemoglu et al. (2006) argue that Albouy's alterations imply that Africa was a healthier place for Europeans than much of Europe. To underline this point, Acemoglu et al. (2005, 2006) modify Albouy's data and show that with a few – in their view necessary – corrections, all of their original results are restored or even amplified. The debate focuses on many more individual coding issues which will not be discussed here; neither do we aim to adjudicate between the two positions.

Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004) criticize three methodological and conceptual issues in the research of Acemoglu et al. (2001, 2002). First, they argue that all of the dominant indicators of institutions used in the literature are outcome measures and do not truly reflect "deep" institutional constraints. Second, they show that settler mortality and population density are highly correlated with other factors affecting GDP today, such as education or the disease environment, which in their view invalidates their use as instruments. They suggest that the settlers might have not brought only their institutions, but more fundamentally their higher levels of human capital. Third, they show panel evidence suggesting that lagged education predicts better institutions and conclude that the modernization hypothesis is a better reflection of reality.

The validity of the indicators used for identification is often a priori assumed and not addressed further in the empirical literature. Glaeser et al. (2004) regard this as problematic and show that the standard indicators (risk of expropriation, government effectiveness, and constraints on the executive) are only weakly correlated with more structural legal indicators, such as judicial independence, plurality, and proportional legislation. Instead, these indicators are mostly based on subjective assessments, exhibit high volatility, and reflect short-term electoral outcomes rather that deep institutional structures. In their view, if these indicators measure short-term outcomes, they cannot be used for causal inference in any study of long-term growth, as they do not reflect structural features but merely perceptions that are positively correlated with GDP levels.

The problem of instrument validity is a common cause for debate in all studies using an instrumental variables approach. As the exclusion restriction is not directly testable,[35] the

---

[35] Adopting an indirect least squares (ILS) representation similar to that in Albouy (2008), the implied system is: (1) $Y_i = \alpha_0 + \alpha_1 R_i + \varepsilon_i$, and (2) $R_i = \beta_0 + \beta_1 Z_i + v_i$. We can solve and replace (1) by its reduced form: $Y_i = \pi_0 + \pi_1 Z_i + \eta_i$. The estimator is $\hat{\alpha}_1 \approx \frac{\pi_1}{\beta_1} = \frac{\alpha_1 \beta_1}{\beta_1} = \alpha_1$. Instrument relevance refers to the requirement that $\beta_1 \neq 0$. Instrument exogeneity or excludability refers to the fact that $\eta_i = [\alpha_1 v_i + \varepsilon_i]$, so $E[v_i|Z_i] = 0$ by

theoretical argument about their validity is decisive. Glaeser et al. (2004) suggest that the instruments are systematically correlated to other factors affecting development outcome.[36] If the "neo-Europes" are richer today due to higher aggregate human capital of the colonial settlers, then low settler mortality is associated with high human capital today, invalidating the exclusion restriction for instrumental variables. They also examine the correlations between setter mortality (and log population density) and their structural legal indicators, which is weak.

Interestingly, their third criticism, which argues that education predicts better institutions, also elicited a direct reply by Acemoglu et al. (2005). In their response, Acemoglu et al. (2005) show that in the original panel regressions of Glaeser et al. (2004) the effect of education becomes insignificant, small and negative, once time fixed-effects are included. In their view, this is due to other omitted factors driving the relationship in the specification without time effects, which falsely led Glaeser et al. (2004) to conclude that there is an effect. In fact, this false conclusion just reflects a general upward trend in the country scores on the institutions indicators and increases in school enrolment occurring over the recent decades. Acemoglu et al. (2005) interpret the results of their re-specification as a confirmation of their critical junctures hypothesis.

These criticisms raise two important and generalizable points which the subsequent empirical literature on institutions and growth needs to address. First, new instrumental variables need to be demonstrated as robust, valid, and relevant, as well as motivated by a detailed description on how the underlying data was constructed. Second, the indicators used to proxy for certain institutional characteristics need to be discussed and firmly established to actually measure the underlying theoretical construct being examined.

3.3   Long-run to short-run growth

The research of Rodrik and coauthors bridges the gap between long-run studies of growth and a policy-relevant discussion of contemporary growth. He has written on the determinants of long-run growth (Rodrik, Subramanian and Trebbi, 2004), growth collapses (Rodrik, 1999), growth accelerations (Hausmann, Pritchett and Rodrik, 2005) and developed a growth diagnostics framework (Rodrik, 2005; Hausmann, Rodrik and Velasco, 2005; Rodrik, 2010) for identifying country specific "binding-constraints". In the following, we review these contributions and attempt to extract a framework linking the determinants of long-term, medium-term and short-term growth.

3.3.1   Long-run growth

Similar to our differentiation between proximate, intermediate and ultimate sources of growth, Rodrik, Subramanian and Trebbi (2004) argue that "growth theory has traditionally focused on physical and human capital accumulation, and in its endogenous variant, on technological change. But accumulation and technological change are at best *proximate causes* of economic growth" (emphasis added, pp. 132-133). They identify three competing

---

construction and $E[\varepsilon_i | Z_i] = 0$ only by assumption.

[36] This point was first raised in Djankov, Glaeser, La Porta, Lopez-de Silanes and Shleifer (2003).

hypothesis of the determinants of long-run growth which have been put forth in the literature: geography (e.g. Diamond, 1997; Sachs, 2001), international trade or economic integration (e.g Frankel and Romer, 1999; Sachs and Warner, 1995), and institutions (e.g. North, 1990; Acemoglu et al., 2001, 2002).

Rodrik et al. (2004) investigate the causal relationships hypothesized by these theories and assess the relative importance of each respective factor. In all theories there are causal interdependencies, such as intensive trade requiring certain institutional prerequisites, or higher income levels leading to both higher trade volumes and (positively) changed institutions. However, each of these theories does claim to identify the main cause of long-run development. To find the most pertinent causal mechanism, Rodrik et al. (2004) use instrumental variables for all endogenous regressors and show that the quality of institutions matters considerably more than the direct effects of trade or geography. Their identification strategy builds on two, then recent, innovations in the literature. First, using the approach of Acemoglu et al. (2001), they instrument for the quality of (legal) institutions today with settler mortality during colonization. Second, as suggested by Frankel and Romer (1999), actual international trade (imports and exports) as a percentage of GDP is instrumented with the results of a gravity equation predicting bilateral trade flows.[37] Geography is exogenous.

In their model, institutions are indirectly linearly dependent on geography and trade, and trade is dependent on institutions and geography. They test their main equation of interest (17), by endogenizing the quality of institutions with equation (18) and a country's trade integration with equation (19):

$$\ln y_i = \mu + \alpha R_i + \beta N_i + \gamma G_i + \varepsilon_i \tag{17}$$
$$R_i = \lambda_R + \alpha_R \ln M_i + \beta_R B_i + \gamma_R G_i + v_{R_i} \tag{18}$$
$$N_i = \lambda_N + \alpha_N \ln M_i + \beta_N B_i + \gamma_N G_i + v_{N_i} \tag{19}$$

where y_i is the log of GDP per capita, R_i is a measure of institutions (namely, rule of law), N_i is the trade share of GDP, G_i the measure of geography (distance to the equator), M_i is log settler mortality, and B_i is the constructed trade share (from the gravity equation estimates of Frankel and Romer, 1999). The exclusion restrictions are that M_i and B_i do not independently enter equation (17).

Rodrik et al. (2004) report the results for three samples sizes. The first sample consists of 64 countries, as in the original Acemoglu et al. (2001) study and uses settler mortality as an instrument for institutions. The second sample is an extended version of the first, consisting of 79 countries and incorporating newer settler mortality data. The third sample of 134 countries uses the fraction of population speaking English and the fraction of the population

---

[37] Frankel and Romer (1999) construct trade flows by extending the following empirical model of bilateral trade with many more geographic variables: $ln(\tau_{i,j}/y_i) = \alpha_0 + \alpha_1 D_{i,j} + \alpha_2 \ln S_i + \alpha_3 \ln S_j + \varepsilon_{i,j}$, where $\tau_{i,j}$ is the bilateral trade between countries i and j (exports plus imports), $D_{i,j}$ is the physical distance between the two countries, and S_i and S_j are measures of country size. Frankel and Romer (1999) drop observations where no bilateral trade is recorded, which is also a challenge to theoretical models of international trade in general. Recently, Helpman, Melitz and Rubinstein (2008) developed a theoretical model and a corresponding two-step estimation procedure which can incorporate zero trade flows.

speaking other European languages (from Hall and Jones, 1999) as alternative instruments for institutions. In all cases, institutions are approximated by an indicator assessing the strength of the "rule of law" (from Kaufmann, Kraay and Zoido-Lobatón, 2002) and geography is measured as distance to the equator. Rodrik et al. (2004) prefer the second sample, as they consider the Acemoglu et al. (2001) instrument more theoretically plausible (than using the linguistic measures) and the linguistic instruments in the third sample do pass the over-identifying restrictions.[38]

Their key result is that "the quality of institutions trumps everything else" (Rodrik et al., 2004, p. 135). In all samples, the specification, which includes the endogenously determined variables and the exogenous geography measure, yields insignificant and negative coefficients for the direct effects of trade and geography, but highly significant and very large coefficients for the direct effects of institutions. They also calculate the total impact by combining the direct effects and indirect effects from additional regressions modeling the linear dependencies. To estimate the entire system of simultaneous effects (apart from the feedback effects from income), they specify two additional instrumental variables regressions. Here, we show only the reduced form of relationships between institutions, trade and geography, and between trade, institutions and geography, respectively:

$$R_i = \delta_0 + \delta_1 B_i + \delta_2 G_i + v_i^a \qquad (20)$$

$$N_i = \zeta_0 + \zeta_1 R_i + \zeta_2 G_i + v_i^b \qquad (21)$$

To estimate the total effect of each variable, they separately apply a unit shock to the error terms of the trade and institutions equations.[39] A unit shock to the institutions equation has a total effect of 1.85 on log incomes, which would create a 5-fold difference in dollar incomes. A similar shock to the trade equation has a total effect on log income of 0.09. The effect of institutions is thus more than 20 times higher than that of trade. When considering only significant coefficients, then the instrumental variables estimate of the direct effect of institutions is equal to the total effect of institutions, which is 198 log points – a more than 6-fold increase in per capita income. The effect of geography remains large with a total effect on income of 149 log points. However, this effect is driven by the large indirect influence that geography has on institutions. They also estimate the same specifications using income per worker, capital per worker, human capital per worker and total factor productivity as dependent variables (from Hall and Jones, 1999). In each case, institutions have a large positive effect which is significant at the 99% level or higher, while in most cases the coefficients on international trade and geography are insignificant or just significant, negative and comparatively small. In sum, Rodrik et al. (2004) find that trade integration has a negligible influence on incomes, geography mainly affects incomes indirectly through

---

[38] If there is more than one instrument for one endogenous regressor, the model is over-identified. A test of over-identifying restrictions (e.g. Sargan test) tests that the residuals from an IV specification are uncorrelated with a set of exogenous instruments. However, these tests are widely known to have low power.

[39] Here, the term shock simply refers to a change and not shock as understood in growth terms. They actually solve the implied system of simultaneous equations (of standardized variables) and recover the parameters for each specified interrelationship. Then they calculate the effects of changing one variable, ceteris paribus, which is equivalent to "shocking" that equation's error term.

institutions, and the quality of institutions has both the largest direct and total effects on per capita incomes.

Rodrik et al. (2004) argue that instrumentation strategies should not be confused with theory building and testing, referring particularly to the contribution of Acemoglu et al. (2001) on which their research builds. For Rodrik et al. (2004), the proposition that colonialism was a key determinant of the modern between-country income distribution cannot account for the similar spread of incomes in countries that were never colonized. They illustrate this point by reporting the standard deviation of log incomes in former colonies (1.01) and non-colonies (0.89). Further, they argue that although Acemoglu et al. (2001) have identified a successful and valid instrumentation strategy, this does not require settler mortality to play a large role in the causal relationship. They underline this argument with an analogy. Angrist and Krueger (1991) identified when a person is born within a year (i.e. the quarter of birth) as a possible instrument for estimating the effect of schooling on earnings. They show that because compulsory schooling goes from age 6 to 16 exactly, children born early in a year have the opportunity to drop out with less schooling than those born later in the year. Using this source of exogenous variation they can recover a consistent estimate of the returns to schooling. However, this strategy does not amount to a quarter of birth related theory of earnings or a direct test of such a theory. Similarly, according to Rodrik et al. (2004), the Acemoglu et al. (2001) strategy does not directly test a theory of colonial origins of development.

3.3.2   *The long-run model*

Following Rodrik et al. (2004), the long-run model can be summarized as shown in Figure 7 below. They concentrate on "deep" determinants of growth or, in our terms, ultimate sources of growth, and allow for interrelationships between all endogenous variables. Institutions affect the income level and higher levels of income affect national institutions. Trade integration (nominal trade over nominal GDP) can directly affect income and higher income can result in more trade integration. Trade also affects institutions, for example by demanding greater organizational capacity or safety nets as compensation for increasing openness (see Rodrik, 1998b), and better institutions can aid in deepening economic integration. Only geography is entirely exogenous and potentially influences institutions (e.g. through tropical diseases), economic integration (e.g. through proximity to trading partners) and the income level (e.g. directly by determining underdevelopment in the tropics).

Figure 7: The "deep" determinants of growth
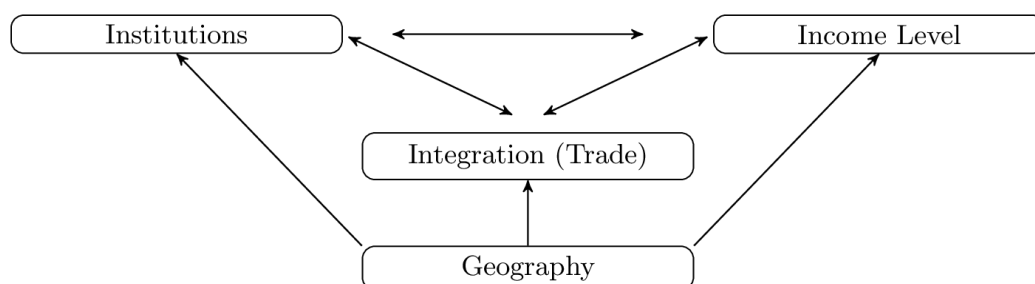
Source: Rodrik, Subramanian and Trebbi (2004).

32

The one-way and two-way arrows above represent all possible relationships among the elements in their multivariate framework. The theory behind these directions comes from the previously cited literature and this model must rather be interpreted as a metamodel in which all these theories fit, rather than an original theory on its own. For example, a simplified version of modernization theory is represented in the feedback channel from income level to institutions, although Rodrik et al. (2004) are primarily concerned with the opposite relationship. They show instrumental variables estimates of all of the interrelationships apart from the income to institutions and income to trade feedback channels (for lack of an instrument for income). Summarizing the results, they find that institutions have by far the largest effect on long-run growth, trade integration has no direct effect, and geography exerts only a strong indirect influence on income through institutions and to a much lesser extent through trade integration. Trade does not exert any effect on institutions, but better institutions feedback positively to economic openness. Hence, Rodrik et al. (2004) stress that causality mainly runs from institutions to income and that there is a strong indirect effect of geography on income via institutions, while all other relationships matter less.

Rigobon and Rodrik (2005) test a very similar model which allows for reverse interrelationships among all included variables. The main difference to the previous model is that institutions are split into rule of law and democracy, rather than just one proxy.[40] Further, instead of using IVs, they employ a novel identification through heteroscedasticity method pioneered by Rigobon (2003). Overall, the results are very similar to Rodrik et al. (2004). Both institutional measures positively predict income, but the effect of rule of law is much more significant than democracy, both economically and statistically. Openness has negative effects on income and greater distance from the equator (geography) positively affects income, democracy and institutions. The main addition of this research to Rodrik et al. (2004) is that the reverse effects of income on institutions or trade are significant but comparatively small, while democracy and rule of law are positively interdependent.

3.3.3 Growth collapses, external shocks, and growth accelerations

Much of the research present so far has concentrated on differences in contemporary levels of GDP per capita, which is academically relevant but of limited use for current policy aimed at stimulating and sustaining growth. To illustrate the difference, we can conceive of the level of GDP per capita and indicators of quality institutions as stock variables which consist of the cumulative sum of *flow variables*, such as growth spurts or collapses and a multitude of policies/reforms (Rodrik et al., 2004). Hence, level regressions measure the cumulative impact of all historical growth-enhancing or growth-constraining policies. It is obvious that the theory and evidence of the determinants of long-term growth vis-à-vis short/medium term growth yield very different insights. According to Rodrik et al. (2004), the policy implications of the long-run literature for short-run growth are non-existent or even harmful when misinterpreted, while investigations linking growth in the short run to institutional characteristics have yet to produce robust and relevant results. To explore the roots of

---

[40] They proxy for the rule of law with the corresponding indicator from the World Governance Indicators (see, for example Kaufmann et al., 2002) and for democracy with the composite indicator from Polity IV (Marshall and Jaggers, 2003).

contemporary growth further, Rodrik argues that we should distinguish among growth collapses, growth accelerations and sustained growth – recognizing that each of these can relate differently to institutions and policies (e.g. Rodrik, 1999; Rodrik et al., 2004; Hausmann et al., 2005). We review the evidence in favor of such a distinction in the following.

In "Where did all the growth go?", Rodrik (1999) focuses on explaining how average growth rates and total factor productivity growth rates in Latin America, the Middle East and East Asia were comparatively high until the mid-1970s, but collapsed in the first two regions thereafter. He argues that the so-called East Asian miracle[41] prior to the Asian financial crisis of 1997-98 can be explained by the total factor productivity declines and dismal growth performance in the Middle East and Latin America after 1973. For Rodrik (1999), the mystery is not the so-called miracle in East Asia, but the relative decline elsewhere.

To explain the growth collapse, Rodrik (1999) proposes to conceptualize the economic turbulence of the 1970s not as merely an effect of external shocks[42] (changes in the terms of trade, wars, and the oil crisis) but as an interaction between external shocks, latent social conflict and conflict management institutions. Specifically, he understands social conflict as a coordination failure among social groups deciding on how to divide a shrinking (negative shock) or growing (positive shock) economic base. In his simple model, groups can either cooperate, which is equal to maintaining the initial distribution applied to the new resource base, or fight, which is aimed at increasing their expected shares. In the latter strategy, latent social conflict turns into open conflict. Open conflict bears with it a cost to the economy and thus further reduces the resource base. Rodrik (1999) argues that the latter behavior arises especially in highly polarized or ethnically fragmented societies (high conflict potential), and/or when the returns to winning are high because the successful exclusion of competing parties is likely (weak conflict management institutions). Differences in growth performance are a function of total shocks experienced in the 1970s, which in turn can be heuristically expressed as:[43]

$$total\ shock = trade\ shock_{70s} \cdot \frac{latent\ conflict}{conflict\ management\ institutions}$$

Rodrik (1999) operationalizes the dependent variable as the differential in per capita growth from 1975-1989 and per capita growth from 1960 to 1975. External shocks in the 1970s are measured by the standard deviation of the first log-differences in the terms of trade from 1971 to 1980 multiplied with the total share of trade in GDP from 1970 to 1974.[44] Rodrik (1999)

---

[41] The East Asian miracle was commonly thought to include rapid increases in productivity. This paradigm was popularly challenged by Krugman (1994), who, building on the work of Young (1994, 1995) and others, has argued that East Asia grew so rapidly mainly due to one-off increases in capital and labor inputs. In retrospect, strong growth in East Asia did not end with the East Asian financial crisis; these economies have continued to grow after the crisis but on average slower than before.

[42] Here we mean shock in the sense of an abrupt and large change, and not a unit change as before.

[43] Essentially, this relationship is a summary of the results of a simple formal model provided in the working paper version of this article (Rodrik, 1998a).

[44] Interestingly, this is merely a measure of the change in the terms of trade rather than a measure of terms of trade shocks. The term shock implies that the measure should capture large changes only.

constructs two sets of conflict and institutions measures. His preferred set is the Gini coefficient of inequality (from Deininger and Squire, 1996) as an indicator of latent conflict and the ICRG composite indicator of the quality of government institutions (from Knack and Keefer, 1995) as an indicator of the quality of conflict institutions. An alternative set uses ethnolinguistic fragmentation (from Mauro, 1995) and a composite indicator of democracy (Freedom House). Rodrik (1999) first tests an additive linear specification, which can be generalized as follows:

$$d_i = \psi + \alpha S_i + \beta F_i + \gamma R_i + \mathbf{X}'_i \delta + \varepsilon_i \tag{22}$$

where d_i is the growth differential between two periods, S_i is a measure of external shocks, F_i is a measure of latent conflict, R_i is a measure of conflict management institutions and $\mathbf{X}'_i$ is the transpose of a vector of covariates (including growth in the previous period, the log of GDP at the break year, and regional dummies) with a corresponding vector of coefficients δ.

Rodrik (1999) finds strong evidence confirming the theory outlined before. All regressions of the growth differential on the explanatory variables include regional dummies, growth of GDP per capita from 1960 to 1975, and the log of GDP per capita in 1975 to account for both the effect of convergence or mean reversion. Including the external shocks measure in addition yields a highly significant coefficient of -0.17. When inequality is added to the specification its coefficient is highly significant and negative (-0.12), while the shock measure remains significant. Interestingly, when the quality of government institutions is added to the regression, the coefficients of external shocks and inequality become insignificant and close to zero. He interprets this as direct evidence of the prescriptions arising from his model, that is, well-developed social conflict management institutions ensure that the distribution of economic resources remains free of opportunistic behavior by certain groups. As a consequence, the output reducing effects of shocks and latent conflict become virtually irrelevant. When using the alternative indicators but leaving the measure of trade shocks unaltered, ethnic fractionalization (conflict) and democracy (conflict management) are both significant and very similar in magnitude but with opposite signs, which suggests that ethnic conflict matters even when controlling for the quality of institutions.

In a second estimation, Rodrik (1999) uses the growth differential as before and the growth differential after the break year (from Pritchett, 1998) as dependent variables. The break year refers to the point of deviation from previous trend growth. Instead of including measures of shocks, conflict and institutions separately, Rodrik (1999) constructs composite measures of social conflict similar to the heuristic equation shown above. The modified model is more in line with the multiplicative effects proposed in his theory and can be represented as:

$$d_i = \psi + \pi(S_i F_i (1 - R_i)) + \mathbf{X}'_i \delta + \varepsilon_i \tag{23}$$

where notation is as before and we additionally assume all measures of institutions to be standardized between zero and one.

The terms of trade variable remains his preferred measure of external shocks. He tests four combinations. The first uses ethno-linguistic fragmentation and democracy, the second, the Gini coefficient for high data quality countries and democracy, the third, all available inequality data and the ICRG institutions measure, and the fourth, the proportion of people not speaking the country's language at home and democracy. In all of the specifications, these measures have highly significant, negative and large coefficients (ranging from -0.77 to -1.65). The results indicate that a one standard deviation change to the conflict indicator corresponds to 0.75 to 1.65 percent lower growth per year relative to the growth performance before.

As latent conflict and institutions can be operationalized in many ways, Rodrik (1999) extends this specification with additional indicators, such as the murder rate, a measure of trust, racial tension, and social spending. Generally, the pattern and results remain robust to these alternatives. Interestingly, when further expanding the specifications to include conventional explanations such as openness to trade, debt to GDP, import tariffs, and government consumption of GDP, their coefficient are all insignificant. Rodrik (1999) also constructs an index of "bad policy" consisting of the inflation rate and black market premia for foreign currency after 1975. This index is strongly correlated with the growth differentials and, in turn, all of his social conflict measures and measures of conflict management institutions are associated with the index in the expected direction. He concludes that participatory politics, democratic institutions, rule of law and social insurance all contribute to macroeconomic stability and resistance to external shocks.

Hausmann, Pritchett and Rodrik (2005) investigate growth accelerations to add to the evidence on growth collapses and growth differentials after the mid-1970s. They employ a novel approach compared to the previous literature, which has concentrated heavily on level regressions or panel data econometrics and mainly came to the conclusion that openness, sound money and property rights matter. Since growth is highly volatile and countries experience growth, stagnation or decline at dissimilar points in time, shifts in the underlying trend for each country can be more informative then evidence based on average growth performance. Hausmann et al. (2005) argue that both neo-classical and endogenous growth theory evolve around the idea of shifting growth paths, comprised of accelerations to a new steady state in the former, or permanent growth accelerations in the latter. Their approach captures these shifts and allows for non-linear relationships, such as a country emerging from a poverty trap, while another remains stuck in a low-level equilibrium. Ultimately, it also links the research to policy relevant questions, such as: how is growth ignited and how is it sustained?

Hausmann et al. (2005) define three conditions which identify growth accelerations. First, average growth during an acceleration episode must be rapid, that is greater or equal to 3.5% per annum. Second, the growth rate must be at least 2% per annum higher than in the previous growth episode and, third, total output after the growth acceleration must exceed the pre-episode maximum level of output. An episode refers to eight years. These criteria are applied as forward-looking and backward-looking comparisons, where the eight years subsequent to a break year are compared with the eight years before, and then the next

possible break year is examined. The third criterion compares the level of output at the end of the growth acceleration with all the available years before the break year. These conditions are crucial as they serve to distinguish pure post-recession recoveries from actual changes towards higher trend growth. Countries can have multiple and overlapping growth accelerations, as long as these are five years apart. Hausmann et al. (2005) use spline regressions to identify the start of accelerations if there is more than one year as a candidate.[45] They use data from the Penn World Tables from 1950 to 1999, hence the first episode can begin in 1957 and the last in 1992.

Table 2: Frequency of accelerated growth episodes

Decade	Region						Total	Eps.	Obs.
	Asia	Africa	Middle East	Europe	Latin America	Other			
1950s	11.11%	5.26%	22.22%	12.82%	3.77%	10.00%	8.78%	12	148
1960s	6.12%	3.49%	5.26%	0.76%	2.78%	6.90%	3.44%	23	668
1970s	3.36%	2.46%	6.06%	0.00%	2.81%	1.89%	2.49%	23	922
1980s	5.30%	0.56%	1.12%	2.78%	0.97%	0.00%	1.62%	16	990
1990s	3.13%	1.10%	0.00%	4.26%	5.45%	4.76%	2.96%	8	270
Total	4.90%	1.87%	4.08%	2.34%	2.53%	2.89%	2.77%	83	2998
Eps.	18	20	10	12	17	6	83		
Obs.	429	965	245	513	673	173	2998		

Source: Hausmann, Pritchett and Rodrik (2005).

This filter results in 83 growth accelerations which includes the well-known growth accelerations (e.g. countries in East Asia during the late 1980s and early 1990s, China in 1978 or Brazil in 1967), but also 20 growth spurts in sub-Saharan Africa and 10 growth accelerations in the Middle East and North Africa. The magnitude of the average acceleration using their filter is very high. The median and average growth per annum is 4% and 4.7%, respectively. As a result, output was on average about 40% higher at the end of an episode than before. When computing the unconditional probability of acceleration per decade, Hausmann et al. (2005) find that the results differ strongly by decade and region (see Table 2 above). However, the number of observations also varies by time and region. If early data availability is correlated with experiencing accelerations, then we must contend that these tabulations obviously exhibit an upward bias.

For 69 of these 83 episodes, Hausmann et al. (2005) have data for the 8 years subsequent to the growth acceleration, which allows an assessment on whether this growth performance was sustained in the longer-term. Interestingly, 23.2% of previous accelerations were followed by negative growth, 23.3% by slow growth (less than 2% per annum), and 53.6% by rapid growth. Of the identified episodes, African countries tended to have negative growth before and after growth accelerations, while Asian countries dominate the group of countries

---

[45] Specifically, they select episodes by identifying the highest F-statistic from multiple spline regressions with the knot (or break) at the candidate year. Spline regression models allow for discontinuities in the underlying data, while the model F-statistic can be used to compare similar specifications.

with high growth prior to and following growth accelerations. Naturally, all of these results strongly depend on the parameters used to identify accelerated growth episodes.[46]

Hausmann et al. (2005) examine the correlates of growth accelerations in various ways and find that growth accelerations seem to be accompanied by more investment, more exports and a devaluation of the exchange rate. However, these correlations could merely capture the filter's inability to remove growth rebounds after macroeconomic crises. We concentrate on the results of their probit analysis here, as it represents their most elaborate attempt to identify structure in the data. Their dependent variable is a dummy variable taking on a value of unity for three years centered on the beginning of a growth acceleration, and zero otherwise. They include explanatory variables capturing favorable terms of trade, positive or negative changes in the Polity IV scores on regime change,[47] the death of an incumbent leader,[48] recent armed conflict or civil wars,[49] economic liberalization,[50] and financial liberalization.[51]

Only a few of the variables emerge as consistently significant. Surprisingly, negative regime changes have a positive impact on igniting growth, while positive regime changes remain insignificant throughout. Favorable terms of trade help to ignite a growth spurt and incumbent leaders that die in office while only holding a short tenure negatively affect growth. Financial liberalization is highly significant and has the largest coefficient of all the estimated variables. Economic liberalization is mostly insignificant, just as wars and civil wars do not have distinguishable effects on growth. In general, all of these specifications have low explanatory power and do not explain more than 8% of the variance. They compare these results to those of alternative estimation methods using probit regression with country-clustered standard errors, censored tobit regression, modified logit regressions to cope with rare-occurrence bias, random-effects probit, and a linear probability model (LPM). In all cases, the results remain remarkably similar in terms of significance and magnitude.

A puzzling result is why changes towards autocracy should positively predict growth accelerations, while changes towards democracy have a negligible effect. Likewise, economic liberalization does not matter much, while financial liberalization does. Hausmann et al. (2005) argue that this result can only be understood if we further differentiate between growth acceleration and sustained growth. To corroborate this point, they re-estimate the previous specifications but differentiate the dependent variable into a sustained and unsustained growth. Sustained growth is defined as growth in excess of 2% per annum in the ten years

---

[46] For example, taking a 5-year horizon results in 137 growth episodes and taking a 10-year horizon results in just 37 episodes. If the rapid growth threshold is raised to 4% per annum, then 68 episodes are identified, and if it is lowered to 3%, then 90 episodes are identified.

[47] From Marshall and Jaggers (2003). Regime change is defined as a 3 unit change in the underlying Polity IV score. The variables are coded unity for a total of five years following a move to greater democracy or authoritarianism.

[48] A dummy variable with a five-year unity value beginning with an incumbent leader's death.

[49] Defined as unity over five years since the end of a civil war, otherwise zero. Similarly, a separate dummy variable is defined for armed conflicts in general.

[50] Capturing a transition to openness similarly to Sachs and Warner (1995), which is also defined as a five year dummy.

[51] A five year dummy from the date financial liberalization occurred with the starting date from the working paper version of Bekaert, Harvey and Lundblad (2005).

after the acceleration episode (8 + 10 = 18 years in total) and growth falling below 2% per annum in the same time horizon is categorized as unsustained. The results diverge sharply. Terms of trade shocks are only significant for unsustained growth, economic liberalization is strongly associated with sustained growth, positive regime changes are significant for sustained but not unsustained episodes, negative regime changes remain significant and positive throughout, and financial liberalization is only related to unsustained growth episodes.

Many of the above results are more intuitive. Terms of trade shocks and financial liberalization strengthen exports or increase foreign capital/domestic returns, but are highly volatile. Economic liberalization, if understood as deep structural reform, seems a precursor of sustained growth and not linked to immediate growth changes. Positive regime changes now matter for medium-term growth which could be related to broader participation in economic opportunities. However, moves to autocracy still positively predict growth and offset the effect of positive changes. Hausmann et al. (2005) do not attempt to conceptualize this specific result, although it can be can be interpreted as indicating that autocratic leaders often establish temporary stability in unstable nation states, which in turn creates enough security for short and medium term growth takeoffs. As long run studies suggest the opposite, this is an indication of the non-linear relationships between growth performance, time and regime types. Hausmann et al. (2005) conclude that the determinants of growth accelerations are not well-identified in these specifications, as they too often incorrectly predict the outcome.

They point to two main results that emerge from this research. First, igniting growth is relatively easy as seen in the rather large number of strong growth spurts. Second, these accelerations are not preceded or well-predicted by changes in political structures, economic reforms or other institutional changes and appear rather to be driven by idiosyncratic factors. We explore the concepts of accelerations in the short and medium-term further by focusing on Rodrik's growth diagnostics framework, which focuses on country-specific "binding constraints", and by subsequently presenting a provisional framework uniting the evidence from the preceding sections.

3.3.4 Binding constraints and growth diagnostics

For Rodrik (2005, 2008, 2010) the lack of variables that can be robustly linked to growth accelerations on average is not very surprising. The plethora of growth models of contemporary growth are evidence of the diverse factors that bring about modern development. Various growth models hold reliable prescriptions, however, each variant only holds under strictly defined conditions. In other words, "Raul Prebisch, Anna Krueger, and Jeffrey Sachs are all correct – at different times and under specific circumstances" (Rodrik, 2010, p. 35). This view especially evolved after the unsatisfactory results of the Washington consensus, which reduced the vector of possible growth strategies to a clearly defined list of quintessential reform strategies, applicable everywhere, to be undertaken as fast as possible, and without much consideration for the country context. The emerging long-term growth literature and advances in endogenous growth models both contributed to the intellectual dismissal of the consensus and the tacit admission of a much more complex reality. This gave

rise to a larger literature on policy reform in a second-best context rather than in ideal type situations.

Hausmann, Rodrik and Velasco (2005, 2008) provide a meta-framework of growth policy analysis and strategies for igniting growth in the short-run. Their key idea is that in a second-best economy, which is virtually the reality everywhere, there is an interplay between any specific distortion and all other distortions. In any reform scenario, not only the direct impact of reducing or removing the targeted distortion must be considered but also the changing interrelationships with all other distortions. Among this universe of distortions, there are certain "binding constraints", i.e. those with the most profound growth debilitating effects which ought to be targeted first. They present a stylized model for conducting "growth diagnostics" (the activity of identifying binding constraints) which, following Hausmann et al. (2005, 2008) can be formally summarized as follows:

$$\frac{du}{d\tau_j} = -\lambda_j + \sum_i \lambda_i \frac{\partial[\mu_i^s(\tau,\ldots)-\mu_i^p(\tau,\ldots)]}{\partial\tau_j} \qquad (24)$$

where u is the welfare of an average member of the economy, τ_j or τ_j is the tax-wedge or distortion on activity j or i with $(i \neq j)$, λ_j or λ_i is the direct cost/benefit of distortion j or i, $\mu_i^s(\tau,\ldots)$ is the social value of activity i after all taxes and all distortions, and $\mu_i^p(\tau,\ldots)$ is the corresponding private valuation.

The framework captures the simple idea of intertwined and differently sized distortions while remaining suitable to incorporate almost any growth model. For example, inadequacies of certain institutions linked to any activity could be considered part of the distortion or a separate condition driving the wedge between private and social valuations. We can break down equation (24) into three distinct parts. The outcome is simply the change in welfare of the average member given a change in the distortion j. The first term is the direct change in welfare of altering the distortion of activity j, i.e. a reduction increases welfare. The second term, however, is the cumulative interaction effect of changing the distortion j with the distortions on all other activities. In other words, the weighted sum of gaps in private and social valuations given a change in distortion j.

The implications are obvious. If the effect of the second term is larger than the first, it is possible that the interaction effects completely offset the welfare gain from the (distortion reducing) reform or even lead to a net welfare loss. Likewise, it is easy to see that ideal-type reforms only consider λ_j and ignore the cumulative or second-best effect of the summation over λ_i's. So what are binding constraints? Essentially nothing else than very large direct effects (λ_j's) which according to Hausmann et al. (2005, 2008) also implies that the indirect effects might not outweigh a reduction in the constraint. They evaluate the merits of five stylized approaches to reform, including their own:

(1) "Wholesale Reform": ideally desirable, but nearly impossible to carry out, as it requires perfect knowledge of all distortions and perfect execution.

(2) "As much as you can": a potentially dangerous policy that can be welfare decreasing when the second-best effects are neglected.

(3) "Second-best reform": ideal piecewise approach, but not feasible as it requires the perfect knowledge of all interaction effects.

(4) "Target largest distortions (τ_i)": largest wedge is not necessarily largest problem for growth and it requires the knowledge of all distortions (arising from market and government failures).

(5) "Binding constraints": feasible, eliminate the distortions with largest first-order welfare increasing effects than assumed second-order welfare decreasing interaction effects.

There is considerable uncertainty inherent in all of these reform strategies, but the essential argument of Hausmann et al. (2005, 2008) is that the binding constraints approach requires the least amount of information, which can in most instances be estimated or guessed rather than perfectly rank-ordered, and this approach to reform has a smaller potential of harming rather than improving the situation. However, this assertion is not entirely obvious, since second-order effects are hard to estimate in any real world scenario and if misjudged can nullify or reverse any attempt at reform no matter what the strategy. Hausmann et al. (2005, 2008) further acknowledge that direct identification of the most directly welfare improving reform is not possible either and suggest instead to systematically analyze the proximate determinants of growth, find underperforming variables and their associated distortions. This approach, which they call "growth diagnostics", can be summarized in a decision tree beginning with the determinants of a balanced growth path in standard neoclassical analysis (Figure 8 below).

The stepwise approach follows from evaluating the components of the balanced-growth equilibrium and at each step questioning which variables affect their performance. We can follow this process by breaking down equation (25) into separate components:
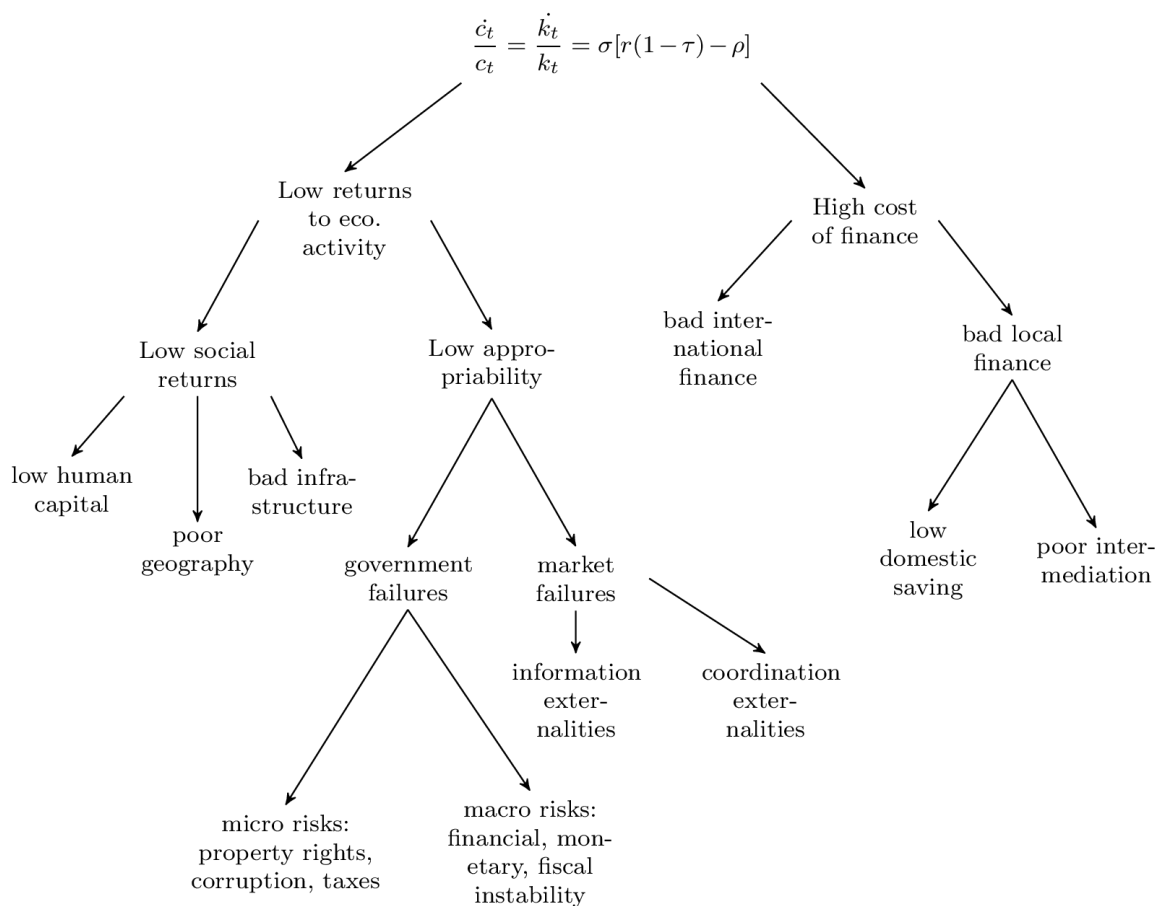
$$\frac{\dot{c}_t}{c_t} = \frac{\dot{k}_t}{k_t} = \sigma[r(1 - \tau) - \rho], \quad with \; r = r(\alpha, \theta, x) \tag{25}$$

where c is consumption, k is capital, σ is the intertemporal elasticity of consumption, r is the return on capital, τ is the tax on capital (formal/informal), and ρ is the world interest rate. Further, r depends on total factor productivity (α), an index of externalities (θ) and the availability of complementary factors of production (x).

Two terms are essential for growth diagnostics: (1) $r(1 - \tau)$, which is the private return to domestic investments, and (2) ρ, which is the cost of finance. High cost of financing domestic investments might be due to a high international assessment of country risks, a high regulatory burden or unattractive FDI positions, among others. Likewise, local capital markets may be underdeveloped and exhibit increased volatility, which in turn is negatively assessed in international capital markets. If private returns are low, this might be due to low social returns or low appropriability. These are essentially defined by four variables: (1) high τ – high taxes, inefficient tax systems, or high risk of expropriation, (2) high θ – large externalities, coordination failures and spillover effects, (3) low α – low productivity, low level of technology, etc., and (4) low x – low human capital stock, underdeveloped

infrastructure, and transport costs in the wider sense. The sub-nodes of the decision tree represent a number of factors that influence these four variables.

Figure 8: Growth diagnostics

$$\frac{\dot{c}_t}{c_t} = \frac{\dot{k}_t}{k_t} = \sigma[r(1-\tau)-\rho]$$

Low returns to eco. activity

High cost of finance

Low social returns

Low appro-priability

bad inter-national finance

bad local finance

low human capital

bad infra-structure

government failures

market failures

low domestic saving

poor inter-mediation

poor geography

information exter-nalities

coordination exter-nalities

micro risks: property rights, corruption, taxes

macro risks: financial, mon-etary, fiscal instability

*Source*: Hausmann, Rodrik and Velasco (2008).

The framework succeeds in combining many macroeconomic and microeconomic interactions, while hierarchically organizing the basic conclusions of a large amount of modern economic theory on factor accumulation, learning and spillovers, externalities, institutions, financial markets, taxation, and government or market failures, and more. We present it here mainly with the purpose of showing that proximate sources of growth, which determine growth outcomes in the medium-term and short-run, depend on many more variables than just factor accumulation often going beyond what evidence from cross-country regressions can reveal. In a broader sense, the research of Rodrik and collaborating authors make clear that for growth theory, evidence and policy analysis the time-frame matters crucially for the results the research will produce and determines their relevance to policy-making.

3.3.5 A unified framework?

Figure 9 is a schematic representation capturing most of the evidence and theory examined in the preceding sections on Rodrik's research. We distinguish between long-run growth paths

and growth in the medium/short-term. Long-run growth paths are to some extent deterministically influenced geographic conditions and their effects on the quality of institutions, but also by a country's ability to build institutions that protect property rights, allow for participatory politics, and create a strong rule of law. In contrast, growth in the medium-term and short-term depends on many traditional factors identified by neoclassical economics, such as factor accumulation, but also modern institutions, external shocks, conflicts, and growth policy in the broadest sense. For simplicity of illustration, we omit possible feedback paths (for example, back from the diverging growth performances to medium-term and long-term factors).

Figure 9: From long-run to short-term growth
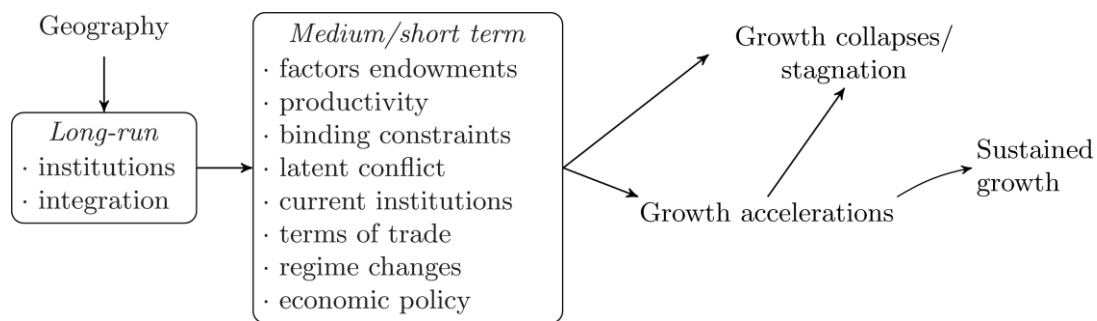
Source: authors' illustration.

First, the interplay between long-term factors has been well-established by the work reviewed earlier and in Rodrik et al. (2004). Institutions are the most crucial of the long-run determinants. In fact, changing institutions in a positive manner can overcome the deterministic influence of geography, increase trade volumes and even capture positively reinforcing effects running back from income levels. More interestingly, these long-run development paths in part determine the current state of institutions, technology (productivity), accumulated human and physical capital, and to a lesser extent the degree of latent social conflict.

Second, for the medium and short-term, we combine all the major insights of Rodrik (1999), Hausmann et al. (2005), and Hausmann et al. (2008). Factor endowments (such as physical and human capital) matter just as much as "binding constraints", which restrict the productive potential and link the factors endowments with the components of the growth diagnostics framework. Hausmann et al. (2005) have shown that is easy to ignite growth, which can happen through policy changes, changes in factor proportions, institutional changes and many idiosyncratic factors that cannot be captured by estimating cross-country averages. However, it is much harder to sustain growth. Rodrik (1999) has offered explanations linking growth performance in the medium-term to latent social conflict and a country's capacity to mitigate the effects of external shocks and the resulting distributive fights in socially fragmented environments through well-developed conflict management institutions. Hausmann et al. (2005) convincingly show that the institutional requirements and factors determining sustained episodes are very different from those affecting accelerations. Moreover, the lack of robust relationships underlines that many other idiosyncratic factors drive the different growth performances.

Last, Rodrik (2000) emphasizes that institutional functions does not directly prescribe institutional forms. For example, he argues that although property rights are among the most fundamental institutions affecting growth, they must not necessarily be equivalent to ownerships rights. Control rights might reduce the gap between social and private returns without any formal transfer of ownership. Many more examples of such a diversity of successful but intrinsically different institutional forms can be found.[52]

In sum, while the link of institutions to long-run level of per capita income is well-established (subject to some econometric objections regarding the instrumentation strategies), the evidence linking growth to short and medium-term outcomes is much less robust. While some might interpret this lack of a strong relationship as merely an empirical obstacle, it is equally plausible that it is due to a multitude of very different country-specific institutional and non-institutional factors involved in igniting, sustaining and collapsing growth performances.
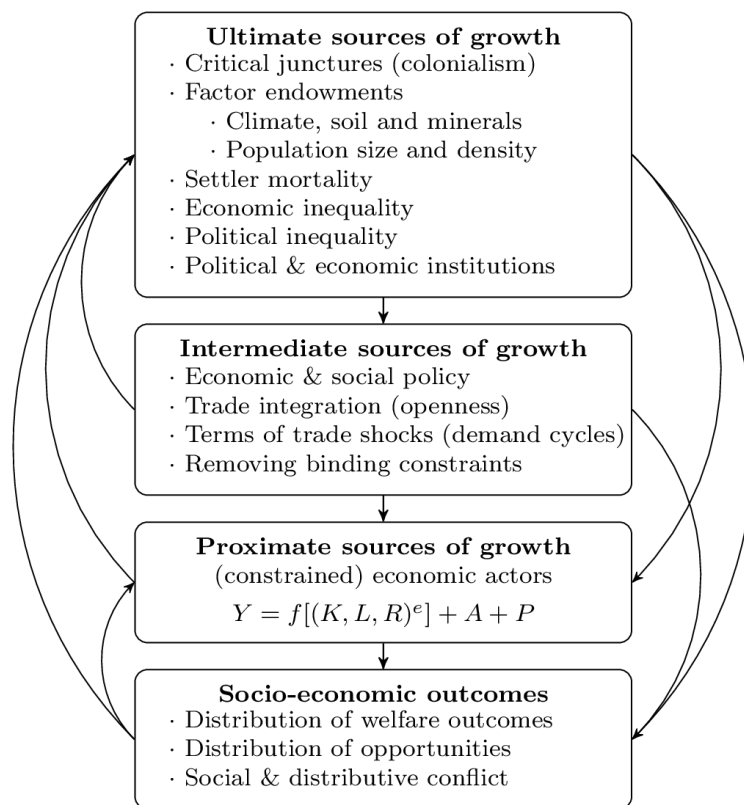
4.   SYNTHESIS

In this section, we place the findings of the literature in the sources-of-growth framework presented in the beginning of this paper. We use this framework to highlight the similarities and differences of the theories and variables examined. Figure 10 below presents a modified version of the framework including only the factors examined in this review of the literature.

However, before we contrast the theories in terms of sources of growth and socio-economic outcomes, two remarks need to be made. First, some long-run factors, such as technology cycles or the distance to the technological frontier, have not been discussed by the contributions reviewed in this paper but nevertheless remain relevant to modern economic growth (e.g. see Comin, Easterly and Gong, 2010). Second, theories referring to changes in culture and attitudes as drivers of long-run growth are not prominent in the recent debates, with a few exceptions.[53] At best they have been partially incorporated into the incentive structures provided by institutions.

---

[52] See Rodrik (2000) for a more detailed discussion. Rodrik (2000) defines five major institutional functions – namely, property rights, regulation (of market failures), macroeconomic stability, social security, and conflict management – and shows how in different countries these functions are fulfilled by very different institutions and/or institutional configurations.

[53] Clark (2007) is one of the few contemporary economists stressing the influence of culture together with technology and Malthusian dynamics on long-run growth. He argues there is the "popular misconception [that] the preindustrial world is of a cowering mass of peasants ruled by a small, violent, and stupid upper class that extracted from them all surplus beyond what was needed for subsistence and so gave no incentives for trade, investment, or improvement in technology" (Clark, 2007, pp. 145). This argument rests on showing how Britain in 1300 and other earlier civilizations had a system of incentives and sufficiently stable rule of law in place. Clark (2007) focuses on simple and selective comparisons of macroeconomic indicators on prices, taxes, public debt and proxies for property rights. He downplays the role played by the transformation of political institutions occurring in the 18th and 19th centuries in igniting economic growth. Very recently, there has been a resurgent interest in the study of the economics of culture and some empirical studies do in fact find effects of culture on long-run growth, which stand to be more widely confirmed (Spolaore and Wacziarg, 2009; Tabellini, 2010). Another exception is provided by Harrison (1985) and Harrison and Huntington (2000).

Figure 10: Modified Sources-of-Growth Framework

Ultimate sources of growth
· Critical junctures (colonialism)
· Factor endowments
    · Climate, soil and minerals
    · Population size and density
· Settler mortality
· Economic inequality
· Political inequality
· Political & economic institutions

Intermediate sources of growth
· Economic & social policy
· Trade integration (openness)
· Terms of trade shocks (demand cycles)
· Removing binding constraints

Proximate sources of growth
(constrained) economic actors

$$Y = f[(K, L, R)^e] + A + P$$

Socio-economic outcomes
· Distribution of welfare outcomes
· Distribution of opportunities
· Social & distributive conflict

Source: authors' illustration.

4.1   Ultimate sources of growth

The core of Engerman and Sokoloff's theory on development among former colonies focuses on ultimate sources of growth and development, that is, a range of factors interacting in shaping institutions in the long run. The timing of the historical shock of colonization and its consequences were driven by two types of factor endowments. First, geographic conditions (mineral resources, climate and soil quality) determined the commodities which could be most profitably produced or harvested by the colonizers. These can be broadly grouped into plantation or mining commodities with economies of scale, such as sugar and certain minerals, and small-scale farming commodities with limited or no economies of scale, like wheat. Through factor endowments, geography defined how attractive a region was for early colonizers as a whole and what type of settlement would come about. Second, demographic characteristics (native population size and density) then determined the availability of unskilled labor, the need to import slaves or contract workers to produce commodities with economies of scale, and subsequently the ratio of arriving European settlers to the non-European population. The unequal distribution of skills created economic inequality in favor of the scarce production factor skilled labor, which in the Americas was largely synonymous to differentiating between natives/slaves/contract laborers and people of European descent. The high degree of economic inequality in South America resulted in a dualistic political economy with a dominant class of European descendant elites and the rest of the

45

population.[54] The resulting distribution of economic power then became institutionalized and was reflected in political inequalities – for example, through limited access to the ballot box and slow extension of the franchise in South America and the Caribbean. Hence, geographic and demographic conditions feature prominently in defining institutions, which in turn affected growth and very unequal social outcomes through a combination of intermediate social and economic policies and proximate sources of growth.

Acemoglu et al. (2001) argue that local disease environments favorable or unfavorable to European mortality affected the size of European settlements, the shape of institutions which the settlers built, and, as a result, the long-run growth outcomes in former colonies. Regions in which Europeans expected high mortality rates received fewer European migrants and inherited extractive institutions created by small elites. By contrast, non-extractive institutions emerged in regions where Europeans could easily settle. For Acemoglu et al. (2001), expected settler mortality is hence a central part of the theory of colonial institutions and a convenient tool for econometric identification. The "critical juncture" of European colonialism then led to what Acemoglu et al. (2002) call a "reversal of fortune" among those countries and regions that were relatively highly developed at an early stage and other regions that were initially less highly developed. They link the start of the relative change in GDP per capita between colonies to the onset of the industrial revolution in 19th century. According to Acemoglu et al. (2002), former colonies with non-extractive institutions and well-protected private property took advantage of the opportunity to industrialize quickly, while powerful elites, extractive institutions and adverse incentives for non-elites barred development in extractive ex-colonies.

Like Engerman and Sokoloff (1997, 2002), Acemoglu et al. (2001, 2002) focus on the factors shaping institutions in the long-run and emphasize the preeminence of historical shocks, or critical junctures, in defining the shape of institutions. However, the two theories differ in three respects. First, Acemoglu et al. (2001, 2002) emphasize the importance of constraints to European colonization over factor endowments. Second, they attribute a weaker role to geography. Third, it is political inequality, not economic inequality, which determines the nature of growth-obstructing institutions.

In their theory, geography exerts its influence through the local disease environment, not independently. Its role is exclusively indirect via the diseases that cause settler mortality. Over time, the indigenous populations developed partial immunity to local diseases. European settlers, on the contrary, faced almost certain death in some regions such as tropical Africa. Settler mortality interacted with population density, which itself has two meanings – one substantive and one methodological. On the one hand, European settlement was easier in regions with sparse and dispersed populations, but this effect is less strong than the mortality effect. On the other hand, population density serves as a proxy for early per capita income and helps to show how the initially relatively richer regions of settlement became poorer over

---

[54] In their theory, the owners of scarce skills are favored, which leads to economic inequality. When this economic inequality becomes institutionalized, the elites gain effective control over the entire resource base of the economy. This comes close to a class-based theory of development. However, the term "class" is conspicuously absent in the modern institutional literature discussed in this paper.

time and initially relatively poor regions of settlement became richer (reversal of fortune). Settler mortality and population density determined the size of European elites and the distribution of political power which subsequently shaped political and economic institutions. Institutions do not emerge semi-deterministically from factor endowments, but as a byproduct of the chances of Europeans to settle permanently or, in other words, as a byproduct of the degree to which they were building institutions for themselves. Stronger property rights and limits on executive power emerged where Europeans had settled in larger numbers. While most colonies were developing some forms of property rights and checks on power, the real pay-off for economic growth occurred only later during industrialization and only in those regions where the rights of larger segments of the population were protected, rather than just those of small elites.

The analysis of Rodrik et al. (2004) does not offer a unique theory of long-run growth of its own. Rather it examines competing explanations of long-run growth. By analyzing the effects of institutions (Acemoglu et al., 2001), trade integration (Frankel and Romer, 1999) and geography (Sachs, 2001; McArthur and Sachs, 2001) on long-run growth, they show how the influence of institutions exceeds that of all other factors. However, they introduce an important qualification, namely that institutions include a substantial indirect effect of geography which is not attributable to settler mortality. Compared to the other theoretical approaches, Rodrik et al. (2004) do not provide an explicit theory of how geography influences institutions.

While in theory restricted to former colonies, these explorations of ultimate causality helped to empirically identify the importance of more inclusive and more egalitarian institutions for long-run growth and outlined many of the factors involved in shaping them.[55] They provide substantial credibility to the link between institutional characteristics and later industrialization and the emergence of modern economic growth. It is, however, important not to overgeneralize the impact attached to any specific factor. For example, general population dynamics are a central component of the preindustrial economic dynamics, but a strict interpretation of the theories presented here would refer only to initial population size and density during colonization. Similarly, even theories stressing a deterministic influence of geographic factors such as factor endowments, location or climate on long-run growth leave substantial room for improving growth dynamics today, and the respective roles of political and economic inequality depend strongly on the country and time-specific political economy (Acemoglu and Robinson, 2000b; Acemoglu et al., 2007).

---

[55] The theories seem to suggest an almost linear relationship between institutionalized inequality and long run growth performance. This is highly misleading. Historically, industrialization has been associated with rapidly increasing inequality as some segments of the population and some segments of the economy forge ahead of others. Also since 1982, the inequalities of incomes and wealth have been increasing exponentially in the most advanced economies. The relationship between institutionalized inequality and long-run growth performance only holds in the very long run in cross-country comparisons. It states that countries with a historical legacy of institutionalized inequality will have lower per capita incomes today, while countries with a more egalitarian institutional legacy tend to have higher levels of per capita income.

4.2   Intermediate sources of growth

For Engerman and Sokoloff (1997, 2002) the intermediate sources of growth, i.e. economic policies, technology policies, political reforms and social policies in the broadest sense, are to some degree a function of historically developed institutional structures. Countries with egalitarian institutions based on comparatively homogeneous populations (e.g. the US and Canada) extended the franchise relatively fast, provided universal schooling and lowered access barriers to credit, ownership of land and protection of intellectual property. Countries with inegalitarian institutions, on the contrary, pursued policies of slow franchise extension, limited schooling and had high barriers to intellectual property rights and credit. Acemoglu et al. (2001, 2002) agree with this. They link policies to the distribution of political power between elites and masses. In politically unequal societies, the ruling elites have few incentives to invest in new technologies if these can threaten the basis of their power. In addition, their theory suggests that redistribution and franchise extension can serve as means of staving off social conflict or revolution and were, historically, often not directly aimed at increasing productive capacity (Acemoglu and Robinson, 2000b).

There are many other intermediate sources of growth. Rodrik gives special attention to trade openness. The degree of integration in international trade interacts positively with more inclusive institutions (signified by the arrow back to ultimate causes). But it only weakly affects growth independently (Rodrik et al., 2004; Rigobon and Rodrik, 2005).[56] Trade is in part shaped by institutions and in part exogenously determined by geographic location, among other factors. While trade is often considered a "deep" determinant of growth, Rodrik finds its effects are most pertinent in the medium and short term. Thus, Rodrik (1999) links growth collapses to an interaction between declines in the terms of trade (demand trends and openness), ultimate sources of growth and socioeconomic outcomes. Similarly, Hausmann et al. (2005) show that growth accelerations are, amongst others, affected by positive trade shocks and financial liberalization, while sustained growth is associated with economic liberalization and positive regime changes. We can conceive of changes in the terms of trade changes, financial liberalization and economic reforms as intermediate sources of growth (demand trends and economic policies in the broadest sense), which certainly affect modern growth rates and stability, as well as socio-economic outcomes.

Furthermore, the removal of "binding constraints", which is so prominent in the Hausmann-Rodrik-Velasco approach, can be seen as a typical intermediate source of growth. Removing binding constraints as a reform strategy links the ultimate sources of growth to the complex structure of national economies today. Analyzing binding constraints puts institutions into perspective and highlights that there is no uniform approach to improving institutions and governance, as these are historically shaped and embedded in a second-best economy. According to Hausmann et al. (2005), successful growth policy needs to consider country-specific interactions between the targeted reform and the incentives provided by the prevailing political and economic structure. This approach suggests a diversity of institutional solutions to improve on factors inhibiting the proximate sources of growth. In our framework,

---

[56] We rely on Rodrik et al. (2004) to have effectively overturned the initial results of Frankel and Romer (1999) indicating a very large and independent effect of trade integration.

it belongs to the intermediate sources of growth, as reforms are short and medium term interventions. The removal of binding constraints can create growth accelerations in the short run. It provides a window of opportunity for more far-reaching institutional and policy reforms which are required to sustain growth over longer periods.

4.3   Proximate sources of growth

The proximate sources of growth are directly affected by both the institutional structure and the intermediate sources of growth. For Engerman and Sokoloff (1997, 2002), this becomes visible especially during industrialization, where previously economic successful heterogeneous societies with high degrees of institutionalized inequality became relatively less productive than more homogeneous societies. The importance of natural resources and cheap unskilled labor declined during industrialization, while the skill-premium increased greatly. The differing institutional arrangements with regard to education, access to land and credit, and patents dramatically affected aggregate efficiency. More homogeneous societies with low access barriers to economic activity tended to pull ahead of heterogeneous elite-ruled societies.[57] In addition, inegalitarian colonial institutions were associated with colonial drain. Continued colonial plunder of economic surpluses further disadvantaged the more unequal colonies.

The outcomes of proximate causality feed back into the institutional structure, as the underlying initial distribution of rents between elites and the rest of society resulted in more egalitarian final long-run outcomes in initially relatively homogeneous societies, while inequality remain unchanged or become even larger in more heterogeneous elite-ruled societies. This effect is so persistent, that both Engerman and Sokoloff (1997, 2002) and Acemoglu et al. (2001, 2002) find that historical institutional structures still influence the proximate sources of growth today. However, while Engerman and Sokoloff (1997, 2002) stress the interplay of institutions, a wide range of socio-economic outcomes and the proximate sources of growth, Acemoglu et al. (2001, 2002) argue more narrowly along the lines of North and Thomas (1973). Strong property rights foster economic activity and align social with private returns, as long as these rights apply to a large proportion of all economic actors.[58] Institutions, as an ultimate source of growth, are directly linked to proximate causes of growth such as productivity, technology, and capital accumulation. For both theories, the increasing skill-premium, low barriers to accessing institutions and well-defined property rights explain the "reversal of fortune" among former colonies occurring with onset of industrialization and remain linked to the proximate sources of growth ever since.

---

[57] Interestingly, this decline for reasons of not promoting domestic institutions seems to have been common knowledge at the time. For example, in 1795, Immanuel Kant writes "[t]he worst, or from the standpoint of ethical judgment the best, of all this is that no satisfaction is derived from all this violence, that all these trading companies stand on the verge of ruin, that the Sugar Islands, that seat of the most horrible and deliberate slavery, yield no real profit, but only have their use indirectly and for no very praiseworthy object – namely, that of furnishing men to be trained as sailors for the men-of-war and thereby contributing to the carrying on of war in Europe" (Kant, 1903[1795], pp. 141–142).

[58] Acemoglu and Johnson (2005) empirically confirm this narrow focus in an empirical study of the long-run effects of property rights versus contracting institutions. In this paper, they find that well-defined property rights have first-order effects on growth in the long run, investment and financial development, while contracting institutions only exhibit positive and significant second-order effects on financial intermediation.

We can also interpret the concept of "binding constraints" not only as a reform strategy but as constraints placed on the economic actors, in effect directly limiting their economic choices. These constraints can be of many kinds. Highly interventionist and bureaucratic economic policies can be a constraint on entrepreneurship. Bad macroeconomic policies can lead to macroeconomic instability which adversely affects all actors' economic incentives. Constraints can arise from institutional features, such as weak protection of property rights which limits invention, but also from too strong protection of property rights which limits imitation by enterprises in follower economies. Constraints can arise from insecure private property rights which inhibit the actors' capacity to appropriate returns from economic activity. Constraints can lie in socioeconomic outcomes, such as low human capital which as a proximate source of growth reduces an economy's productive capacity. Finally, constraints can also be exogenous and related to adverse geographic conditions such as landlockness or lack of infrastructure.

4.4 Socio-economic outcomes

In the theory of Engerman and Sokoloff (1997, 2002), inequalities in welfare outcomes and opportunities to take part in economic activities are intrinsically linked to the development of institutions and long-run growth outcomes. Especially in conjunction with historical shocks, initial socio-economic differences in outcomes play a large role in defining subsequent institutional structures. As we have seen, the ultimate sources of growth and development determine the long-run shape of the distribution of income and opportunities to market participation available to a majority of the population of a country. The intermediate sources, such as redistributive or schooling policies, can modify social outcomes and could thus act as a counterweight to long-run institutional influences. However, as noted before, policy itself is often a function of institutional structures, so degrees of freedom in policy are not unlimited. The proximate sources of growth finally determine growth rates and levels of GDP per capita in the short-run, and thus directly affect social outcomes.

Social outcomes feed back into ultimate, intermediate and proximate sources of growth contributing to institutional path dependence. Economic inequality and inequality of opportunity help to maintain institutions of limited access. For example, people who cannot vote cannot redistribute income and wealth towards themselves without revolt, and lack of access to education will result in lower growth, technological change and economic efficiency. These links are weaker and less multifaceted in the theory of Acemoglu et al. (2001, 2002) than in Engerman and Sokoloff (1997, 2002). For Acemoglu et al. (2001, 2002) the persistence of institutions is rather exclusively determined by the distribution of political power. In other words, political inequality trumps economic inequality and is conceptualized not as an independent outcome, but a central characteristic of what defines 'good' institutions. Beyond this, welfare outcomes matter for the survival of non-democratic regimes, but they place less of an emphasis on the interaction between ultimate causes and social outcomes, as compared to Engerman and Sokoloff (1997, 2002).

Rodrik (1999) further connects socio-economic outcomes to growth performance in the short and medium term. His theory and evidence show how trade shocks interact with latent social conflict and harm growth especially in countries with underdeveloped conflict management

institutions (including limited democracy).[59] Social conflict expresses itself in a high degree of income inequality, high rates of crime, or other indicators of inequalities in welfare and opportunities. In this view, institutions for growth are also institutions of social cohesion and macroeconomic stability, which transcends a narrow focus on property rights institutions only (Rodrik, 2000).

5.   CONCLUSION

Overall, we find support in the literature for the three working hypotheses formulated at the beginning of this paper. With regard to the first hypothesis, institutional arrangements do in fact shape long-run economic growth. This conclusion has been underlined in theory by Engerman and Sokoloff (1997, 2002) and Acemoglu et al. (2001, 2002), but also draws on a wide-array of econometric support (Acemoglu et al., 2001, 2002; Rodrik et al., 2004; Easterly, 2007). Interestingly, slavery itself emerges as detrimental for long-run growth beyond its effect on inequality (Nunn, 2008a) and the mechanisms through which it negatively affects growth differ strongly from former slave importing countries in the New World to former slave exporting countries in Africa (Nunn, 2008b).

The second hypothesis, that political and economic inequality affects growth (a) indirectly via its effects on institutions and (b) directly via the proximate sources of growth also finds support in the literature. Econometric tests of Engerman and Sokoloff's theory reveal that structural economic inequality harms not only institutions and growth, but also limits investment in schooling (Easterly, 2007). Engerman and Sokoloff (1997, 2002) themselves present qualitative evidence for a wide range of institutions which is consistent with the second hypothesis.

Explaining the historically lagging development of schooling and literacy in Latin America and the Caribbean as compared to the USA and Canada, as the consequence of economically and politically exclusive institutions, shows how human capital accumulation was inhibited over time. This pattern of limited access to specific institutions extends to banking and patents, which are vital for increasing aggregate efficiency, capital accumulation and technological change. Engerman and Sokoloff (2005) maintain that these effects, in many cases, originated from economic inequality, which only later transformed itself into political inequality. On the other hand, Acemoglu et al. (2007) argue that political inequality matters more than economic inequality in most circumstances.

The third hypothesis directly addresses the issue of persistence of institutionalized political and economic inequality over time. All empirical studies and models discussed in this paper generally conclude that early institutions still affect institutions today (Acemoglu et al., 2001, 2002; Rodrik et al., 2004). By extension, present day inequalities in income are in part shaped by the historical development of institutions (Easterly, 2007). In fact, without time-persistent institutions, empirical studies of long-run growth would fail to corroborate most of the relationships discussed in this paper.

---

[59] The correlation between institutions and the volatility of growth performances is supported by plenty of additional research aiming to establish a causal relationship (e.g. Acemoglu, Johnson, Robinson and Thaicharoen, 2003; Mobarak, 2005).

Nevertheless, this line of research on institutions still faces major methodological, empirical and theoretical problems. For a large part, these arise from the attempt to reconcile a quest for parsimony with ambitious research agendas, while facing considerable empirical difficulties. For example, the empirical conclusions about the effects of institutions and policies on contemporary growth outcomes are at best still tentative. The analysis of growth accelerations rather than just growth rates, leads to unexpected and interesting findings, but why growth accelerations occur remains largely unexplained. Further, the effects of political and economic inequality remain ambiguous, because they are still imperfectly quantified or sometimes not empirically distinguished at all.

Some reject this type of institutional analysis entirely, on the grounds that institutions are completely endogenous (Przeworski, 2004a,b). However, all of the reviewed theories allow for institutional changes and simply argue that there is persistence to *some* degree and considerable evidence of, for example, the lasting effects of colonial heritage.

The role of geography and, to a lesser extent, diseases is still contested. A considerable branch of the long-run literature argues for both direct and indirect effects of geography on development (McArthur and Sachs, 2001; Bloom and Sachs, 1998; Sachs, 2001), whereas Rodrik et al. (2004) and Acemoglu et al. (2001, 2002) show that geography only indirectly affects growth through institutions. But even these indirect effects are very large. In line with such findings, Engerman and Sokoloff (1997, 2002) place a much higher emphasis on geographical factors as historical determinants of institutions than Acemoglu et al. (2001, 2002). Moreover, there are grounds to doubt the success of the instrumental variables strategies employed by Acemoglu et al. (2001, 2002) and Rodrik et al. (2004), as pointed out by Albouy (2004) and Glaeser et al. (2004). Likewise, Easterly's (2007) application of a colonial theory to the entire world is questionable.

In all of these analyses, institutions are still treated as an aggregate "black-box". Too often, the focus is exclusively on property rights related variables only, such as the risk of expropriation, the rule of law or constraints on the executive. A pure property rights focus neglects the interactions among different types of institutions in bringing about stability, social cohesion and safe investment climates.

Much still needs to be added to these investigations, such as studying the effects of institutional characteristics in more specific historical contexts in a Gerschenkronian tradition, investigating the interactions between specific institutional domains and technological advance, and developing a broader approach to institutions which is not exclusively focused on property rights only. Such approaches will require better measurement of a variety of institutional characteristics and better empirical methods to separate these into distinct and meaningful dimensions.

REFERENCES

Abramovitz, M. 1989. "Thinking about growth." in *Thinking about growth and other essays of economic growth and welfare*, Cambridge University Press.

Acemoglu, D. and J.A. Robinson. 2000. "Political losers as a barrier to economic development." *American Economic Review*, 90 (2): 126–130.

Acemoglu, D. and J.A. Robinson. 2000. "Why did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective." *Quarterly Journal of Economics*, 115 (4): 1167–1199.

Acemoglu, D. and J.A. Robinson. 2006. *Economic Origins of Democracy and Dictatorship*, Cambridge University Press.

Acemoglu, D. and J.A. Robinson. 2008. "Persistence of power, elites, and institutions." *The American Economic Review*, 98 (1): 267–293.

Acemoglu, D. and S. Johnson. 2005. "Unbundling institutions." *Journal of Political Economy*, 113 (5): 949–95.

Acemoglu, D., J.A. Robinson, and T. Verdier. 2004. "Alfred Marshall Lecture: Kleptocracy and divide-and-rule: a model of personal rule." *Journal of the European Economic Association*, 2 (2-3): 162–192.

Acemoglu, D., M.A. Bautista, P. Querubin, and J.A. Robinson. 2007. "Economic and political inequality in development: the case of Cundinamarca, Colombia." *NBER Working Paper.*

Acemoglu, D., S. Johnson, and J.A. Robinson. 2001. "The colonial origins of comparative development: An empirical investigation." *American Economic Review*, 91 (5): 1369–1401.

Acemoglu, D., S. Johnson, and J.A. Robinson. 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., S. Johnson, and J.A. Robinson. 2005. "A response to Albouy's "A re-examination based on improved settler mortality data"." *unpublished*, Massachusetts Institute of Technology.

Acemoglu, D., S. Johnson, and J.A. Robinson. 2006. "Reply to the Revised (May 2006) version of David Albouy's "The Colonial Origins of Comparative Development: an Investigation of the Settler Mortality Data"." *unpublished*, Massachusetts Institute of Technology.

Acemoglu, D., S. Johnson, J.A. Robinson, and Y. Thaicharoen. 2003. "Institutional causes, macroeconomic symptoms: volatility, crises and growth." *Journal of Monetary Economics*, 50 (1): 49 – 123.

Acemoglu, D., S. Johnson, J.A. Robinson, and P. Yared. 2005. "From education to democracy?" *American Economic Review*, 95 (2): 44–49.

Acemoglu, D., S. Johnson, J.A. Robinson, and P. Yared. 2007. "Re-evaluating the Modernization Hypothesis." *unpublished*, Harvard University.

Acemoglu, D., S. Johnson, J.A. Robinson, and P. Yared. 2008. "Income and democracy." *The American Economic Review*, 98 (3): 808–842.

Albouy, D. 2004. "The colonial origins of comparative development: a reexamination based on improved settler mortality data." *unpublished*, University of California, Berkeley.

Albouy, D. 2006. "The colonial origins of comparative development: an investigation of the settler mortality data." *unpublished*, University of California, Berkeley.

Albouy, D. 2008. "The colonial origins of comparative development: an investigation of the settler mortality data." *NBER working paper, 14130.*

Alesina, A. and D. Rodrik. 1994. "Distributive politics and economic growth." *The Quarterly Journal of Economics*, 109 (2): 465–490.

Anderson, T.W. and C. Hsiao. 1982. "Formulation and estimation of dynamic models using panel data." *Journal of Econometrics*, 18 (1): 47–82.

Anderson, T.W. and H. Rubin. 1949. "Estimation of the parameters of a single equation in a complete system of stochastic equations." *The Annals of Mathematical Statistics*, 20 (1): 46–63.

Angrist, J.D. and A.B. Krueger. 1991. "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics*, 106 (4): 979–1014.

Banerjee, A.V. and A.F. Newman. 1993. "Occupational Choice and the Process of Development." *The Journal of Political Economy*, 101 (2): 274–298.

Barro, R.J. 1999. "Determinants of democracy." *Journal of Political Economy*, 107 (6): 158–183.

Barro, R.J. 2000. "Inequality and Growth in a Panel of Countries." *Journal of Economic Growth*, 5 (1): 5–32.

Bates, R. 1981. *States and markets in tropical Africa: The political basis of agricultural policy.* Berkeley: University of California Press.

Bekaert, G., C.R. Harvey, and C. Lundblad. 2005. "Does financial liberalization spur growth?" *Journal of Financial Economics*, 77 (1): 3–55.

Bloom, D.E. and J.D. Sachs. 1998. "Geography, demography and economic growth in Africa." *Brookings Papers of Economic Activity*, 2.

Chenery, H., M.S. Ahluwalia, J.H. Duloy, C.L.G. Bell, and R. Jolly. 1974. *Redistribution with growth; policies to improve income distribution in developing countries in the context of economic growth*, Oxford: Oxford University Press.

Clark, G. 2007. *Farewell to Alms: A Short Economic History of the World*, Princeton: Princeton University Press.

Coatsworth, J.H. 1999. "Economic and Institutional Trajectories in 19th Century Latin America." in J.H. Coatsworth and A.M. Taylor, eds., *Latin America and the World Economy since 1800*, Cambridge, MA: Harvard University Press, pp. 23–54.

Comin, D., W. Easterly, and E. Gong. 2010. "Was the Wealth of Nations Determined in 1000 BC?" *American Economic Journal: Macroeconomics*, 2 (3): 65–97.

Commons, J.R. 1936. "Institutional economics." *The American Economic Review*, 26 (1): 237–249.

Curtin, P.D. 1989. *Death by migration: Europe's encounter with the tropical world in the nineteenth century*, Cambridge University Press.

Curtin, P.D. 1998. *Disease and empire: The health of European Troops in the Conquest of Africa*, Cambridge University Press.

Deininger, K. and L. Squire. 1996. "A new data set measuring income inequality." *The World Bank Economic Review*, 10 (3): 565.

Deininger, K. and L. Squire. 1998. "New ways of looking at old issues: inequality and growth." *Journal of Development Economics*, 57 (2): 259–287.

Denison, E.F. 1967. *Why growth rates differ: postwar experience in nine western countries*, Washington, D.C.: The Brookings Institution.

Diamond, J. 1997. *Guns, germs and steel: the fates of human societies*, W. W. Norton & Company.

Djankov, S., E. Glaeser, R. La Porta, F. Lopez-de-Silanes, and A. Shleifer. 2003. "The new comparative economics." *Journal of Comparative Economics*, 31 (4): 595–619.

Easterly, W. 2007. "Inequality does cause underdevelopment: Insights from a new instrument." *Journal of Development Economics*, 84 (2): 755–776.

Easterly, W. and R. Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *The Quarterly Journal of Economics*, 112 (4): 1203–1250.

Easterly, W. and R. Levine. 2001. "What have we learned from a decade of empirical research on growth? It's Not Factor Accumulation: Stylized Facts and Growth Models." *The World Bank Economic Review*, 15 (2): 177.

Engerman, S.L. and K.L. Sokoloff. 1997. "Factor endowments, institutions, and differential growth paths among new world economies." in S. Haber, ed., *How Latin America Fell Behind*, Stanford, CA: Stanford University Press, pp. 260–296.

Engerman, S.L. and K.L. Sokoloff. 2002. "Factor endowments, inequality, and paths of development among new world economies." *Economia*, 3 (1): 41–88.

Engerman, S.L. and K.L. Sokoloff. 2005. "Colonialism, inequality, and long-run paths of development." *NBER working paper.*

Engerman, S.L. and K.L. Sokoloff. 2005. "The evolution of suffrage institutions in the New World." *The Journal of Economic History*, 65 (04): 891–921.

Fieldhouse, D.K. 1982. *The colonial empires: a comparative survey from the eighteenth century*, Macmillan.

Forbes, K.J. 2000. "A Reassessment of the Relationship between Inequality and Growth." *American Economic Review*, 90 (4): 869–887.

Frankel, J.A. and D. Romer. 1999. "Does trade cause growth?" *American Economic Review*, 89 (3): 379–399.

Galor, O. and J. Zeira. 1993. "Income Distribution and Macroeconomics." *The Review of Economic Studies*, 60 (1): 35–52.

Galor, O. and O. Moav. 2004. "From Physical to Human Capital Accumulation: Inequality and the Process of Development." *The Review of Economic Studies*, 71 (4): 1001–1026.

Gerschenkron, A. 1962. *Economic backwardness in historical perspective*, Cambridge, MA: Harvard University Press.

Glaeser, E.L., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer. 2004. "Do institutions cause growth?" *Journal of Economic Growth*, 9 (3): 271–303.

Gruchy, A.G. 1947. *Modern economic thought: the American contribution*, Prentice-Hall.

Gruchy, A.G. 1978. "Institutional versus Orthodox Economics and the Interdisciplinary Approach." in T. Huppes, ed., *Economics and Sociology: Towards an Integration*, Leiden: Martinus Nijhoff, pp. 51–75.

Gutierrez, H. 1986. "La Mortalite des Eveques Latino-Americains aux XVIIe et XVIII Siecles." in *Annales de Demographie Historique*, pp. 29–39.

Hall, R.E. and C.I. Jones. 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *The Quarterly Journal of Economics*, 114 (1): 83–116.

Harrison, L.E. 1985. *Underdevelopment is a state of mind: The Latin American case*, Center for

International Affairs, Harvard University.

Harrison, L.E. and S.P. Huntington. 2000. *Culture matters: How values shape human progress*, Basic Books.

Hartz, L. 1964. *The Founding of New Societies: Studies in the History of the United States, Latin America, South Africa, Canada and Australia*. New York, Harcourt, Brace & World.

Hausmann, R., D. Rodrik, and A. Velasco. 2005. "Growth diagnostics." *unpublished*.

Hausmann, R., D. Rodrik, and A. Velasco. 2008. "Growth Diagnostics." in N. Serra and J.E. Stiglitz, eds., *The Washington Consensus Reconsidered*, Oxford University Press, USA, chapter 15, pp. 324–356.

Hausmann, R., L. Pritchett, and D. Rodrik. 2005. "Growth accelerations." *Journal of Economic Growth*, 10 (4): 303–329.

Helpman, E., M. Melitz, and Y. Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes." *The Quarterly Journal of Economics*, 123 (2): 441–487.

Hirschman, A.O. 1958. *The strategy of economic development*, New Haven, CT: Yale University Press.

Hoff, K. and J.E. Stiglitz. 2001. "Modern economic theory and development." in G.M. Meier and J.E. Stiglitz, eds., *Frontiers of development economics: The future in perspective*, The World Bank, pp. 389–444.

Kant, I. 1903[1795]. *Perpetual peace: a philosophical essay*, Allen & Unwin.

Kapp, K.W. 1950. The social costs of private enterprise, Harvard University Press.

Kaufmann, D., A. Kraay, and P. Zoido-Lobatón. 1999. "Governance matters." *World Bank Policy Research Working Paper No. 2196*.

Kaufmann, D., A. Kraay, and P. Zoido-Lobatón. 2002. "Governance matters II: updated indicators for 2000-01." *World Bank Policy Research Working Paper No. 2772*.

Khan, B.Z. and K.L. Sokoloff. 1998. "Patent institutions, industrial organization and early technological change: Britain and the United States, 1790-1850." *Technological Revolutions in Europe*, pp. 292–313.

Knack, S. and P. Keefer. 1995. "Institutions and economic performance: cross-country tests using alternative institutional measures." *Economics & Politics*, 7 (3): 207–227.

Krugman, P. 1994. "The myth of Asia's miracle." *Foreign Affairs*, 73 (6): 62–78.

Kuznets, S. 1955. "Economic growth and income inequality." *The American Economic Review*, 45 (1): 1–28.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny. 1999. "The quality of government." *Journal of Law, Economics, and organization*, 15 (1): 222.

Lewis, W.A. 1954. "Economic development with unlimited supplies of labour." *The Manchester School*, 22 (2): 139–191.

Lewis, W.A. 1978. *Growth and fluctuations*, 1870–1913, Allen & Unwin.

Lipset, S.M. 1959. "Some social requisites of democracy: Economic development and political legitimacy." *The American Political Science Review*, 53 (1): 69–105.

Maddison, A. 1987. "Growth and slowdown in advanced capitalist economies: techniques of quantitative assessment." *Journal of Economic Literature*, 25 (2): 649–698.

Maddison, A. 1988. "Ultimate and proximate growth causality: a critique of Mancur Olson on the rise and decline of nations." *Scandinavian Economic History Review*, 36 (2): 25–29.

Mariscal, E. and K.L. Sokoloff. 2000. "Schooling, Suffrage, and the Persistence of Inequality in the Americas, 1800–1945." in S. Haber, ed., *Political institutions and economic growth in Latin America: essays in policy, history, and political economy*, Stanford, CA: Hoover Institution Press, pp. 159–218.

Marshall, M.G. and K. Jaggers. 2003. "Polity IV Dataset (Computer file; version p4v2003)." Center for International Development and Conflict Management, College Park (MD), University of Maryland.

Mauro, P. 1995. "Corruption and growth." *The Quarterly Journal of Economics*, 110 (3): 681–712.

McArthur, J.W. and J.D. Sachs. 2001. "Institutions and Geography: Comment on Acemoglu, Johnson and Robinson 2000." *NBER working paper.*

Mitchell, W.C. 1910. "The rationality of economic activity." *The Journal of Political Economy*, 18 (3): 197–216.

Mobarak, A.M. 2005. "Democracy, volatility, and economic development." *Review of Economics and Statistics*, 87 (2): 348–361.

Myrdal, G. 1968. *Asian Drama: An Inquiry into the Poverty of Nation*, Vol. 1-3, New York: Pantheon.

North, D.C. 1990. *Institutions, institutional change, and economic performance*, Cambridge: Cambridge University Press.

North, D.C. and R.P. Thomas. 1973. *The Rise of the Western World*, a New Economic History, Cambridge University Press.

Nunn, N. 2008a. "Slavery, inequality, and economic development in the Americas: an examination of the Engerman-Sokoloff Hypothesis." in E. Helpman, ed., *Institutions and economic performance,* Cambridge, MA: Harvard University Press, pp. 148–180.

Nunn, N. 2008b. "The Long-Term Effects of Africa's Slave Trades." *Quarterly Journal of Economics,* 123 (1): 139–176.

Perotti, R. 1996. "Growth, income distribution, and democracy: What the data say." *Journal of Economic growth*, 1 (2): 149–187.

Persson, T. and G. Tabellini. 1994. "Is inequality harmful for growth?" *The American Economic Review*, 84 (3): 600–621.

Pritchett, L. 1998. "Patterns of Economic Growth: Hills, Plateaus, Mountains, and Plains." *World Bank Policy Research Working Paper No. 1947.*

Przeworski, A. 2004. "Institutions Matter?" *Government and Opposition*, 39 (4): 527–540.

Przeworski, A. 2004. "The last instance: Are institutions the primary cause of economic development?" *European Journal of sociology*, 45 (02): 165–188.

Przeworski, A., M. Alvarez, J.A. Cheibub, and F. Limongi. 2000. *Democracy and Development: Political Regimes and Economic Well-being in the World, 1950-1990*, New York: Cambridge University Press.

Rigobon, R. 2003. "Identification through heteroskedasticity." *Review of Economics and Statistics*, 85 (4): 777–792.

Rigobon, R. and D. Rodrik. 2005. "Rule of law, democracy, openness, and income." *Economics of Transition*, 13 (3): 533–564.

Rodrik, D. 1998a. "Where Did All The Growth Go? External Shocks, Social Conflict, and Growth Collapses." *NBER Working Paper 6350*.

Rodrik, D. 1998b. "Why do more open economies have bigger governments?" *Journal of Political Economy*, 106 (5): 997–1032.

Rodrik, D. 1999. "Where did all the growth go? External shocks, social conflict, and growth collapses." *Journal of Economic Growth*, 4 (4): 385–412.

Rodrik, D. 2000. "Institutions for high-quality growth: what they are and how to acquire them." *Studies in Comparative International Development (SCID)*, 35 (3): 3–31.

Rodrik, D. 2003. "Introduction." in D. Rodrik, ed., *In Search of Prosperity: Analytical Narratives on Economic Growth*, Princeton University Press, pp. 1–19.

Rodrik, D. 2005. "Growth strategies." *Handbook of Economic Growth*, 1, 967–1014.

Rodrik, D. 2008. "Second-best institutions." *NBER Working Paper*.

Rodrik, D. 2010. "Diagnostics before prescription." *The Journal of Economic Perspectives*, 24 (3): 33–44.

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

Rosenstein-Rodan, P.N. 1943. "Problems of industrialisation of eastern and south-eastern Europe." *The Economic Journal*, 53 (210): 202–211.

Rostow, W.W. 1959. "The stages of economic growth." *The Economic History Review*, 12 (1): 1–16.

Sachs, J.D. 2001. "Tropical underdevelopment." *NBER Working Paper*.

Sachs, J.D. and A. Warner. 1995. "Economic reform and the process of global integration." *Brookings papers on economic activity*, 1995 (1): 1–118.

Sen, A.K. 1960. *Choice of techniques: An aspect of the theory of planned economic development*, Blackwell.

Sokoloff, K.L. and E.M. Zolt. 2007. "Inequality and the Evolution of Institutions of Taxation: Evidence from the Economic History of the Americas." in S. Edwards, G. Esquivel, and G. Marquez, eds., *The decline of Latin American economies: growth, institutions, and crises,* Chicago, IL: University of Chicago Press.

Sokoloff, K.L. and S.L. Engerman. 2000. "History lessons: Institutions, factors endowments, and paths of development in the new world." *The Journal of Economic Perspectives*, 14 (3): 217–232.

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.

Solow, R.M. 1991. "New directions in growth theory." in B. Gahlen, H. Hesse, and H.J. Ramser, eds., *Wachstumstheorie und Wachstumspolitik: ein neuer Anlauf*, Mohr/Siebeck, pp. 3–17.

Spolaore, Enrico and Romain Wacziarg. 2009. "The Diffusion of Development." *The Quarterly Journal of Economics*, 124 (2): 469–529.

Szirmai, A. 2008. "Explaining success and failure in development." *UNU-MERIT Working Paper #2008-013*.

Szirmai, A. 2012. *Dynamics of socio-economic development: an introduction*, 2nd ed., Cambridge

University Press. forthcoming.

Tabellini, G. 2010. "Culture and institutions: economic development in the regions of Europe." *Journal of the European Economic Association*, 8 (4): 677–716.

Thirlwall, A.P. 2005. *Growth and Development with special reference to developing economies*. Palgrave, Macmillan.

Veblen, T. 1899. *The theory of the leisure class: An economic theory of institutions*. Macmillan.

Williamson, O.E. 1985. *The economic institutions of capitalism*, New York: Free Press.

Young, A. 1994. "Lessons from the East Asian NICS: A contrarian view." *European Economic Review*, 38 (3-4): 964–973.

Young, A. 1995. "The tyranny of numbers: confronting the statistical realities of the East Asian growth experience." *The Quarterly Journal of Economics*, 110 (3): 641–680.