

## UNU-MERIT Working Paper Series

**#2012-033**

### **Institutions and long-run growth performance: An analytic literature review of the institutional determinants of economic growth**

**Richard Bluhm and Adam Szirmai**

#### **Working Paper Series on Institutions and Economic Growth: IPD WP02**

This working paper is part of the research programme on ‘Institutions, Governance and Long-term Economic Growth’, a partnership between the French Development Agency (AFD) and the Maastricht Graduate School of Governance (Maastricht University – UNU-Merit). The research builds on the Institutional Profiles Database IPD, jointly developed by AFD and the French Ministry of the Economy since 2001.

**ISSN 1871-9872**

**Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)**

email: [info@merit.unu.edu](mailto:info@merit.unu.edu) | website: <http://www.merit.unu.edu>

**Maastricht Graduate School of Governance (MGSoG)**

email: [info-governance@maastrichtuniversity.nl](mailto:info-governance@maastrichtuniversity.nl) | website: <http://mgsog.merit.unu.edu>

Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands

Tel: (31) (43) 388 4400, Fax: (31) (43) 388 4499

**AFD-MGSoG/UNU-Merit Working Paper Series on  
« Institutions, Governance and Long term Growth »**

In 2010, the French Development Agency (AFD) initiated a partnership with the Maastricht Graduate School of Governance (Maastricht University - UNU-Merit) with a view to exploring the conceptual and econometric relationships between institutions and long-term growth. As a development bank with a long-term lending horizon, AFD is particularly interested in better understanding the determinants of countries' long term economic, social, and political trajectory.

AFD has thus developed a programme on “Institutions, Governance, and Long-term Growth” dealing with the five following dimensions:

- (i) Measuring institutions and discussing the meaning of such measures, notably through the Institutional Profiles Database;
- (ii) Testing the econometric relationship between institutional measures and long term growth;
- (iii) Exploring through a series of country case studies the historical relationship between processes of economic accumulation, forms of political organisation, and social cohesion;
- (iv) Discussing conceptual frameworks for making sense of the interaction between political, social and economic forces in the process of development;
- (v) Developing methodologies for political economy analyses.

The MGSoG/UNU-Merit team is involved in the five dimensions with a particular focus on the first two. Its primary objective is to explore the Institutional Profiles Database jointly developed by AFD and the French Ministry of the Economy since 2001. Institutional Profiles Database is unique by its scope (about 350 elementary questions pertaining to all institutional dimensions covering 148 countries in 2012), its entirely free access, and its ambition to incorporate the most recent theoretical advances in the field of political economy.

The present series intends to convey the results of our ongoing research, and in so doing to reflect the wealth of issues that can be fruitfully addressed from an “institutionalist” perspective. We hope that readers will find these papers stimulating and useful to develop their own understanding and research.

Nicolas Meisel (AFD)  
Adam Szirmai (MGSoG/UNU-Merit)

For more information on the programme, please visit our websites:

<http://www.maastrichtuniversity.nl/web/Schools/MGSoG/ProjectPages/InstitutionalProfilesDatabase.htm>  
<http://www.afd.fr/lang/en/home/recherche/themes-recherches/institutions-gouvernance-croissance>

# *Institutions and long-run growth performance*

*An analytic literature review of the institutional determinants of economic growth*

Richard Bluhm<sup>a</sup> and Adam Szirmai<sup>b</sup>

January 2012

## *Abstract*

This paper provides an analytic review of selected contributions to the study of institutions and economic growth. We review the contributions to the study of institutional determinants of long-run growth by Engerman and Sokoloff, and Acemoglu, Johnson and Robinson. We discuss the work of Rodrik and others who focus on institutions and institutional reform and take steps towards bridging the gap between the study of long-run and short-run growth performances. In addition, we review two new theoretical frameworks by North, Wallis and Weingast and Khan that relate the structure of institutions to short-run volatility and long-run growth trends. We survey a wide array of supplementary econometric evidence and criticisms relating to each of these key contributions. Special attention is given to identifying the underlying causal relationships, the empirical methods and the kind of data used to test theories and hypotheses found in the literature. Further, we compare the findings in different strands of the literature using a sources-of-growth framework which distinguishes between ultimate, intermediate and proximate causes of growth and development.

*Keywords:* growth, institutions, inequality, development

*JEL Classification:* O43, O30, O11

---

<sup>a</sup> Maastricht Graduate School of Governance, Maastricht University, Keizer Karelplein 19, 6211 TC, Maastricht, The Netherlands, e-mail: [richard.bluhm@maastrichtuniversity.nl](mailto:richard.bluhm@maastrichtuniversity.nl)

<sup>b</sup> UNU-MERIT and Maastricht Graduate School of Governance, Keizer Karelplein 19, 6211 TC, Maastricht, The Netherlands, e-mail: [szirmai@merit.unu.edu](mailto:szirmai@merit.unu.edu)

We gratefully acknowledge financial support from the Agence Française de Développement (AFD). This paper has benefitted substantially from comments by Nicholas Meisel, Fikret Adaman, Kıvanç Karaman, and Ayça Akarçay-Gürbüz, and discussions with Bart Verspagen, Thomas Ziesemer, Kaj Thomsson, as well as the participants of the AFD ‘Institutions and Growth’ workshop in Paris and similar workshops in Maastricht. The findings, interpretations and conclusions expressed in this paper are solely that of the authors and do not necessarily represent policies or views of the Maastricht Graduate School of Governance, UNU-MERIT, AFD and/or other affiliated institutions. All remaining errors are those of the authors.

## TABLE OF CONTENTS

1.	Introduction.....	2
2.	The Framework .....	4
3.	The Review .....	6
3.1	The Washington consensus, institutions and inequality .....	6
3.1.1	The original Washington Consensus .....	6
3.1.2	Disappointment, criticism and new policy avenues.....	8
3.2	The Engerman and Sokoloff hypothesis .....	11
3.2.1	Institutions of (in)equality .....	13
3.2.2	The model.....	14
3.2.3	Criticism and econometric evidence .....	16
3.3	The critical junctures hypothesis .....	22
3.3.1	Colonial origins .....	22
3.3.2	Reversal of fortune.....	24
3.3.3	The model.....	28
3.3.4	Critical junctures or modernization?.....	31
3.3.5	Criticism and additional evidence .....	32
3.4	Long-run to short-run growth .....	36
3.4.1	Long-run growth .....	36
3.4.2	The long-run model.....	39
3.4.3	Growth collapses, external shocks, and growth accelerations .....	40
3.4.4	Binding constraints and growth diagnostics.....	46
3.4.5	A unified framework? .....	50
3.5	Power, rents and economic growth.....	51
3.5.1	Limited and Open Access Orders .....	51
3.5.2	Political settlements .....	65
4.	Synthesis .....	74
4.1	Ultimate sources of growth .....	75
4.2	Intermediate sources of growth .....	78
4.3	Proximate sources of growth.....	79
4.4	Socio-economic outcomes.....	80
5.	Conclusion .....	82
	References .....	85

# 1. INTRODUCTION

There are vast differences in per capita incomes between the most developed countries and the least developed countries. In 1820, the average living standard between the West and all other regions differed by about factor two. In 2007, high income countries were on average 21 times wealthier than their low income counterparts.<sup>1</sup> Such income gaps are the result of differences in long-run growth rates. Similarly, the degree of income inequality varies sharply from the developed world to the developing world, as well as among countries and entire regions. Latin America is the world's most unequal region, while many countries in East Asia are much more economically equal but often unequal in terms of political access. In Africa, income inequality is smaller than in Latin America but levels of poverty remain high and political exclusion is a widespread phenomenon. What factors cause these differences in economic growth and development?

While neoclassical growth theory has proved a powerful device for understanding the proximate sources of growth, empirical investigations have shown that much of growth remains unexplained by factor accumulation.<sup>2</sup> In new growth theory and evolutionary theory, technology is endogenous and the question shifts to what determines the rate of technical change. Moreover, a large literature in economic history has focused on explaining the emergence of modern economic growth and the take-off of the West in the 19<sup>th</sup> century. In different ways, these literatures gave rise to a search for 'ultimate sources' of growth that can explain the causes of factor accumulation, technical change or the 'great divergence' (Pomeranz, 2001) and thus opened the door to institutional analysis. Hence, we argue that development must be viewed in a historical perspective linking institutions, but also political and economic inequalities, with growth outcomes today.

In this paper, we provide an analytic review of the recent literature on the relationships between institutions and economic growth. Following North (1990), institutions are defined as the humanly devised constraints that structure human interaction, and provide the rules of the game. To an important extent, institutions determine the scope and degrees of freedom for policy making. Together with policies and culture, institutions provide the incentives which guide the behavior of economic actors. The different authors discussed in this review have their own definitions of institutions, but these are largely captured by this broader definition.

We first place institutions among the many factors that influence economic growth using a framework developed by Szirmai (2005, 2008, 2012a). Next, we discuss how the Washington Consensus resulted in a broader reform agenda incorporating institutions and inequality. The review itself focuses on selected key contributions to the study of long-run growth (Engerman and Sokoloff, 1997, 2002; Acemoglu, Johnson and Robinson, 2001, 2002) and their critics (e.g. Glaeser et al., 2004; Albouy, 2008; Austin, 2008; Nunn, 2008), as well as studies emphasizing the different effects of institutions as ultimate and proximate sources of growth (Rodrik, 1999; Rodrik, Subramanian and Trebbi, 2004; Hausmann, Pritchett and

---

<sup>1</sup> Based GDP per capita data from Maddison (2010) for 1820 and the Penn World Table 6.3 for 2007.

<sup>2</sup> There is plenty of historical and contemporary evidence in favor of this assertion (for example, Solow, 1956, 1957; Denison, 1967; Abramovitz, 1989; Solow, 1991; Easterly and Levine, 2001; Hoff and Stiglitz, 2001).

Rodrik, 2005) and new theories aiming to combine growth and political stability with the structure of institutions (North, Wallis and Weingast, 2009; Khan, 2010). In our analysis of these works, we focus on the underlying causal relationships, methods and data used.

We limited the scope of this review to recent and macro-oriented studies of institutions and growth.<sup>3</sup> Most of the included works – except for North et al. (2009) and Khan (2010), who add new theoretical perspectives – draw on extensive empirical evidence and have been widely discussed in the literature. Naturally, selection also implies that we exclude influential studies focusing on other aspects of the political economy of institutions, such as Persson and Tabellini's (2003) study of the economic effects of constitutions, Dixit's (2004) work on the governance of transactions, or Greif's (2006) analysis of self-enforcing institutions.

There is now a growing consensus that institutions matter for growth, but disagreement about how exactly, 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, 1910) and post-WWII institutionalism (e.g. Gruchy, 1947, 1978; Kapp, 1950; Hirschman, 1958; Myrdal, 1968) have offered varying approaches to the study of institutions. New Institutional Economics with its transaction cost approach integrated institutions with neoclassical analysis (e.g. North and Thomas, 1973; North, 1990; Williamson, 1985). Building on this neo-institutional literature, recent theories and econometric studies suggest varying *causal* mechanisms of how institutions developed and determine long-run growth.

Neither early institutionalisms nor neoclassical economics emphasized within-country income inequality. In early neoclassical economics, inequality was often seen as promoting growth through higher savings by the rich and positive incentive effects.<sup>4</sup> For example, Kuznets (1955) hypothesized that inequality first rises and then falls as a by-product of development. Similarly, the earlier institutions literature mainly focused on factors other than inequality, such as the advantages of technological backwardness (Gerschenkron, 1962), escaping low-equilibrium traps (Rosenstein-Rodan, 1943), or forward and backward linkages (Hirschman, 1958).<sup>5</sup> Since these early contributions, the effects economic and/or political inequalities have become integral parts of current research linking institutions and growth.

In order to highlight how institutions and inequalities interact in affecting economic growth, our review of the literature is guided by three themes. (1) Are differences in institutional arrangements a fundamental determinant of differences in long-run economic growth since 1500? (2) Does political and economic inequality affect growth by influencing the nature of institutions? (3) To what extent do past institutions determine current institutions?

The remainder of this paper is organized as follows. Section 2 briefly presents our framework. Section 3 reviews each group of authors in the terms of their own theory and models, stylizes the causal channels, and presents some of the most important criticisms. It

---

<sup>3</sup> Aron (2000) provides an excellent review of the state of empirical literature up until the late 1990s.

<sup>4</sup> There is a tradition of this line of reasoning. For example, Lewis (1954) models development with unlimited supplies of labor, where exploitation of cheap labor and inequality are directly linked to growth through the assumption that the rich are responsible for savings, investment and accumulation.

<sup>5</sup> The exception was Myrdal (1968), who saw inequality (i.e. class and status) as an obstacle for growth.

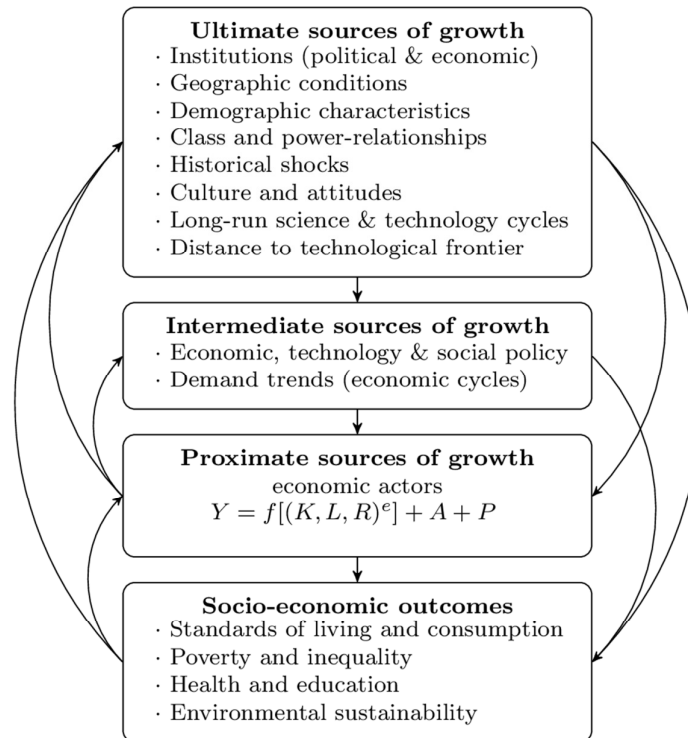
has five subsections. Section 3.1 illustrates how the Washington Consensus contributed to a renewed focus on institutions. Sections 3.2 and 3.3 address the ‘factor endowments’ and ‘critical junctures’ hypotheses, respectively. Section 3.4 highlights the difference between the empirics of long-run and short-run growth. Section 3.5 reviews two new approaches linking rent-seeking and growth. Section 4, the synthesis, discusses the key findings of the review building on the framework presented in the beginning. 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) and developed further in Szirmai (2012a), who reviews classical development theories on the basis of this framework.

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 although they sometimes include variables capturing proximate and intermediate causality. On the other hand, studies of growth in the short and medium run mainly refer to intermediate and proximate causes and may attempt to link these with contemporary socio-economic outcomes.

FIGURE 1: THE SOURCES OF GROWTH FRAMEWORK



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, GDP per capita) and socio-economic outcomes. In the following, we briefly review the components of this framework.<sup>6</sup>

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.<sup>7</sup> 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 run and short run 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

---

<sup>6</sup> See Szirmai (2008) and also the forthcoming second edition of “The dynamics of socio-economic development” by Szirmai (2012b) for a much more detailed discussion of the framework.

<sup>7</sup> 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.



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 role of historical shocks like 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 true even though many contemporary outcomes can be positively or negatively altered even in the absence of growth and is not intended to downplay the feedback channels that connect outcomes to proximate and ultimate causes. The degree to which productive capacity is transformed into social outcomes depends on the nature of social and economic policy (intermediate causality), institutions, and initial levels of social inequality (ultimate causality). For example, the interaction between outcomes (inequality) and ultimate causes (geography, initial inequality and institutions) is the basis of the theory of Engerman and Sokoloff (1997), whose contribution we review in Section 3.2.

### **3. THE REVIEW**

#### **3.1 THE WASHINGTON CONSENSUS, INSTITUTIONS AND INEQUALITY**

The rediscovery of institutions in the modern development literature can be traced to two sources. On the one hand, it relates to continuous innovations in the economic history and institutional economics literatures which pioneered the transaction cost approach to institutions, as referred to in the introduction. On the other hand, a renewed interest in institutional constraints and the quality of governance arose from the disappointment with the results of the Washington Consensus reforms in the 1990s. In this section, we briefly outline the original content of the Washington Consensus and discuss how it resulted in a broader reform agenda incorporating institutions and distributional concerns.

##### **3.1.1 THE ORIGINAL WASHINGTON CONSENSUS**

The Washington Consensus was first formulated by Williamson (1990) in the wake of the Latin-American debt crisis. Contrary to the many meanings this term later took on, Williamson (1990) simply attempted to devise a list of policies on which influential political players and economists at the U.S. Treasury and the Bretton Woods institutions could agree – in other words, policies that commanded a consensus.<sup>8</sup> The following policies were on the list:

---

<sup>8</sup> In many ways, the term ‘Washington Consensus’ was a misnomer from the beginning. As Williamson (2002) points out he used the term “to frame a conference whose ulterior purpose was to persuade Washington that Latin America was engaged in serious reform, not to furnish a policy agenda for Latin America”. These policies originated in Latin America and summarized the broad lessons learned from stagnation in the 1980s. The

1. *Fiscal discipline*: allow only for minimal short-term deficits, balance the budget in the medium-term and keep the debt to GNP ratio from rising.
2. *Public expenditure priorities*: reduce subsidies that provide disincentives and reprioritize spending towards primary education, primary health care and infrastructure.
3. *Tax reform*: broaden the tax base and implement more moderate marginal tax rates.
4. *Liberalize interest rates*: establish market-determined and moderately positive real interest rates to discourage capital flight and encourage savings.
5. *Exchange rates*: ensure competitive exchange rates, in line with equilibrium exchange rates.
6. *Trade policy*: liberalize imports; allow only very limited temporary infant industry protection and only moderate tariffs (10-20%).
7. *Foreign direct investment*: remove barriers to inward FDI to promote inflows of capital, skills and technology.
8. *Privatization*: sell state-owned companies, as private management is often more effective and sales relieve the pressure on public budgets.
9. *Deregulation*: remove barriers to entry and exit of firms into industry and individuals into the labor market. Complex regulation only fuels rent-seeking and corruption.
10. *Property rights*: strengthen private property rights, a key pillar of any capitalist system.

The reform agenda was inspired by criticisms of the import substitution policies and government interventions in markets that characterized development policy from the 1950s to the 1970s. Research increasingly identified quantitative restrictions as the driving force behind rent-seeking, slow growth and political crisis (e.g. Krueger, 1974). In this respect, the consensus exemplified a fundamental paradigm shift away from the structuralist perspective. As Krueger points out it was “generally believed that import substitution at a minimum outlived its usefulness and that liberalization of trade and payments is crucial for both industrialization and economic development” (1997, p. 1). Structuralists believed that markets would not function in developing countries in the same ways as in the advanced economies. There were a variety of structural constraints (infrastructure, limited and fragmented markets, insufficient information, poverty traps) which called for extensive state intervention to promote growth and development. The structural adjustment policies of the 1990s stood structuralism on its head. They identified government interventions and policy induced distortions as the main structural constraints to the function of markets and economic growth (Szirmai, 2005). Williamson (1990) summarizes the three core ideas behind the new consensus as macroeconomic discipline, a market-based economy, and outward-orientation. Stated more succinctly, the policy package these ideas implied became known as “stabilize, privatize, and liberalize” (Rodrik, 2006, p. 973).

Although Williamson (1990) did not link the reform agenda to specific authors, they were largely supported and inspired by the mainstream economics literature (e.g. Balassa, 1982; Sachs, 1985; Edwards, 1989). However, it is rarely noted that Williamson (1990) also

---

Washington institutions (the U.S. Treasury, the IMF and the World Bank) only later adopted these lessons, in part, due to Williamson’s paper.

expressed his astonishment that all of these ideas exclusively relied on mainstream economics and did not incorporate many of the then recent advances in the development literature on poverty-traps (e.g. Schleifer et al., 1989), endogenous growth (e.g. Romer, 1986), or two-sector models (e.g. Uzawa, 1966; Lucas, 1988). This list of omissions can easily be extended to, *inter alia*, include the discovery of the role of externalities and information asymmetries (Greenwald and Stiglitz, 1986) or efficiency wage theory (Shapiro and Stiglitz, 1984). Especially the latter two contributions showed that markets are inefficient if they are incomplete or information is imperfect, and that labor market rigidities (i.e. unemployment) can occur endogenously.

As Williamson (2008) later pointed out, the consensus represented the lowest common denominator among policy makers, and he further makes the qualification that it did not fully capture the policies the U.S. Treasury and Bretton-Woods institutions would eventually pursue especially in terms of financial liberalization and exchange rate regimes.<sup>9</sup> Moreover, the reform package subsequently became associated with ‘market fundamentalism’ or neoliberalism – interpretations which Williamson (1999, 2002, 2008) repeatedly stressed had no basis in his original formulation. We agree with Naim (2000) that this development is most likely best understood as a search for a new ideology after the collapse of the Soviet economies which led the term to acquire a life of its own. In order to avoid a broader ideological discussion, we restrict this review to the narrower definition of ten specific policy instruments.

### 3.1.2 DISAPPOINTMENT, CRITICISM AND NEW POLICY AVENUES

Approximately a decade into the implementation of the reform agenda, it became increasingly clear that it was not delivering the promised results. Especially in Latin America, where the policies were more rigorously applied than any other region, economic growth in the 1980s and 1990s was well below the growth rates under import substitution and several severe financial crises reversed earlier accomplishments (Ortiz, 2003). Growth in the region as a whole compared negatively with high growth in the East Asian economies, whose approach to reform often involved industrial policy, state-intervention and more unorthodox policy instruments (Rodrik, 1996). Similarly, the transition trajectories of Eastern European economies were strongly influenced by a particular version of the Washington Consensus emphasizing ‘shock’ privatization, liberalization and price adjustments. As in Latin America, the outcomes were very heterogeneous, ranging from successful reforms in Poland to the collapse of the entire Russian economy. All across post-Soviet Eastern Europe the initial falls in output greatly exceeded the reformer’s expectations, leading defenders of rapid privatization and price adjustments to call for more radical reforms (Sachs, 1991, 1992) and

---

<sup>9</sup> For example, Williamson (1990) argued for competitive interest rates based on an intermediate exchange rate regime, which did not necessarily imply completely pegging the exchange rate to another currency (or set of currencies) or floating it freely. Later, however, Williamson (2008) notes that the IMF and World Bank soon promoted a dual policy of firmly fixed or free-floating exchange rates. Similarly, Williamson (2002) argues that the Argentine collapse was caused by a grossly over-valuated exchange rate and disregard of fiscal prudence, both of which were direct violations of the original Washington Consensus recommendations. More recent research points out the growth benefits of currency undervaluation for developing countries (e.g. Rodrik, 2008).

critics to point out the agenda's complete disregard for institutional development and history of a country (Menshikov, 1993; Murrell, 1995).

In a well-known essay titled *Washington Consensus or Washington Confusion?*, Naim (2000) describes the failure of the reform agenda as the consequence of an approach that was both too narrow and which neglected the significance of institutional constraints. Exchange rate liberalization, fiscal reform and openness to FDI had limited effects when corruption reigned, tax authorities could not enforce collection, and an unreliable justice-system, undereducated work force and insufficient public infrastructure discouraged investors (Naim, 2000, p. 93). Williamson (2002) partially shares this assessment and concedes that his list underemphasized the role of financial crisis and institutions. In the economics literature, these developments motivated but also coincided with the so-called 'institutional turn'. Since the mid-1990s, a growing number of econometric studies identified large effects of corruption and institutions on economic growth (Mauro, 1995; Knack and Keefer, 1995; Hall and Jones, 1999; Acemoglu et al., 2001). These findings spurred a follow-up literature that reexamined the effects of openness to trade, exchange rate overvaluation and inflation on growth, often finding little support of an independent effect of policies once differences in institutions were accounted for (e.g. Easterly and Levine, 2003). The rediscovery of institutions shed considerable doubt on the empirical foundations of the policies propagated by the Washington Consensus. However, while this turn towards institutions remains academically relevant most early studies did not include much in the way of "useful ideas on how to implement institutional reforms" (Naim 2000, p. 94).

Interestingly, the original Washington consensus paid no attention to policies explicitly targeting poverty and inequality. Especially from the perspective of persistent and high poverty and inequality in Latin America this may seem like a startling omission. Two explanations are often advanced to explain why inequality was not part of the initial reform package. First, as argued by Williamson (2002), a focus on distributive issues simply did not command a consensus and there was a general expectation that the income distribution would not fundamentally change as a result of reforms. Second, as argued by Krugman (2008), most models actually predicted a stable or even improving income distribution arising from trade liberalization. Hence, inequality was not considered to be a problem. The gains from trade openness were expected to drive GDP growth by 2% annually benefitting everyone and raising wages while depressing capital returns.<sup>10</sup> In fact, contrary to expectations, inequality in Latin America actually increased in 1990s and increased trade openness failed to deliver the intended effects. Robust explanations of this phenomenon are difficult to establish, but Krugman (2008) locates the likely causes in financial liberalization which advantaged only parts of society and tax policies that continued to be consumption-based and regressive. As with the renewed attention for institutions, studies in the mid-1990s began to highlight the importance of inequality for economic growth in general (e.g. Alesina and Rodrik, 1994) and

---

<sup>10</sup> Krugman (2008) points out that this income convergence is one of the main prescriptions of the Heckscher-Ohlin model of international trade and the associated Stolper-Samuelson theorem which predicts that the price of the abundant production factor rises as the relative price of the output produced by it increases. As Latin America in the 1990s was considered labor abundant, wage inequality should have declined as a result of trade openness.

identified initial inequality as a driving factor behind the dissimilar growth experiences of East Asia and Latin America in particular (Rodrik, 1994).

The disappointments with the consensus provoked two very different sets of responses.<sup>11</sup> The first response was a rejection of market oriented and neo-liberal policies and a rediscovery of structuralist ideas. Stiglitz (2002) blamed much of the crisis and instability of the 1990s in Latin America, Eastern Europe and Africa on dogmatic market-oriented policies, especially in the field of financial liberalization. Chang (2002) attacked Washington consensus policies for ‘kicking away the ladder’ and reducing the scope for industrial and innovation policies in latecomer economies. Cimoli, Dosi and Stiglitz (2009) argued in favor of interventionist policies and renewed protectionism. Other participants in the debate criticized the way the reform agenda was implemented by the international financial institutions but also failed to outline coherent alternatives (as discussed in Santiso, 2004). These criticisms of the Washington consensus emphasize directing attention towards institutional constraints to development and solutions that allow for institutional diversity.

The second type of response continued to press for market oriented reform, but concluded that the initial reform agenda did not go far enough. The reforms of the 1990s were retrospectively conceptualized as ‘first-generation’ reforms whose benefits can only be fully realized if an extended set of deeper ‘second-generation’ reform is successfully implemented (Kuczynski and Williamson, 2003; Krueger, 2002). Second-generation reforms stressed institutions and improved governance, though there was little agreement on what establishing good institutions actually entailed. Other authors called for a broadening of the reform agenda. For example, Stiglitz (2008) highlights that privatization was quickly augmented with competition policy for natural monopolies and as distributive issues came to the forefront social policy and social protection was added to the agenda. In practice, post-Washington Consensus development policy incorporates both deepening and broadening.

In short, we can interpret the rediscovery of institutions, and to a lesser extent the renewed interest in distributive issues, as a product of two trends. The first trend consists of an autonomous rediscovery of the importance of institutions in the economics and economic history literatures, which resulted in new theories of how institutions affect development (e.g. North, 1990), coupled with the advent of econometric studies underlining a significant and sizable effect of institutions on growth.<sup>12</sup> The second trend is a renewed focus of international organizations, policy makers and development researchers on the role of institutions, governance and rent-seeking in development, aimed at understanding and overcoming the heterogeneous outcomes of previous market-oriented reforms. To a considerable extent, the

---

<sup>11</sup> This distinction has become very common in the literature, but varies slightly across authors. For example, Rodrik (2006) identifies a school of thought arguing for deepening of the reforms and another school arguing for broadening of reforms. Santiso (2004) identifies radical rejection, broadening and deepening. Stiglitz (2008) refers to a Washington consensus plus which entailed broadening and a Washington consensus plus plus, which entailed deepening and institutional reforms.

<sup>12</sup> The rediscovery took place within economics. Other disciplines such as anthropology, sociology and political science had always emphasized the key role of institutions. However, until recently mainstream economics ignored many of the advances and basic findings of related disciplines.

modern debate on good governance and institutional reform was thus born out of the failure of the Washington Consensus.

However, Rodrik (2006) cautions that if this development is taken at face-value, it may result in a form of ‘institutional fundamentalism’ or excessive focus on good governance that is not justified by the econometric finding that ‘institutions matter’. Radical changes in institutions only occur at historical junctures, such as the collapse of the Soviet Union or the establishment of Maoist China, and even then institutional path-dependence may influence the transition outcomes. Cross-country studies comparing the effects of institutions on levels of GDP per capita in, for example, Chile and Nigeria have little or nothing to say about how one country could acquire the institutions of another. This is an important distinction to make, especially at the beginning of an analytic literature review on institutions and growth. The difference between the insight that institutions matter for growth and its implications for institutional reform is an important theme to which we return in Section 3.4.

### 3.2 THE ENGERMAN AND SOKOLOFF HYPOTHESIS

In a series of papers focusing on the divergent developmental experiences of the New World<sup>13</sup>, Engerman and Sokoloff developed a controversial theory which has received considerable attention in the modern literature. They focus on very long-run growth outcomes and regard economic inequality as the principal factor shaping the institutions that account for the take-off of North America throughout the course of the 19<sup>th</sup> century and the subsequent relative decline of South America.

TABLE 1: GDP PER CAPITA IN SELECTED NEW WORLD ECONOMIES

	<i>GDP per capita relative to the U.S.</i>			
	<i>1700</i>	<i>1800</i>	<i>1900</i>	<i>1997</i>
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 17<sup>th</sup> 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 (see Table 1 above). Engerman and Sokoloff (1997, 2002) argue that the North was initially not economically attractive to early colonizers. Only later this trend was reversed, owing to differences in the institutional

<sup>13</sup> See Engerman and Sokoloff (1997, 2002, 2005), Sokoloff and Engerman (2000), and the additional sources cited in the text.

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 encouraged human capital accumulation.<sup>14</sup> 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, differing institutional developments are a credible explanation for both timing and scale of the reversal in relative incomes between the northern and southern colonies over 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 soil quality 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 small European elites governing a largely poor and enslaved population that proved detrimental to the emerging institutional structures. In the North, the relatively 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<sup>15</sup> 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).

---

<sup>14</sup> The price of homogeneity among the population of European descent was the complete marginalization and even elimination of the indigenous Indian population.

<sup>15</sup> 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.

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 endowed with minerals and had 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. Not least due to the immigration restrictions enforced by the Spanish, these colonies also ended up with relatively fewer Europeans.<sup>16</sup> 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 (dense) native population.<sup>17</sup> 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).

### 3.2.1 INSTITUTIONS OF (IN)EQUALITY

Central to their hypothesis is that these factor endowments and patterns of colonial rule resulted in institutions which first created and then maintained an unequal distribution of wealth, human capital, and access to economic and political opportunities. Relying on qualitative evidence, they give examples of how these inequalities expressed themselves in six 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). Latest by the mid-1800s, the

---

<sup>16</sup> 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.

<sup>17</sup> 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.



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 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 promoting small land holdings – 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, on the contrary, landholdings were highly concentrated and large land ownership 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. 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.

### 3.2.2 THE MODEL

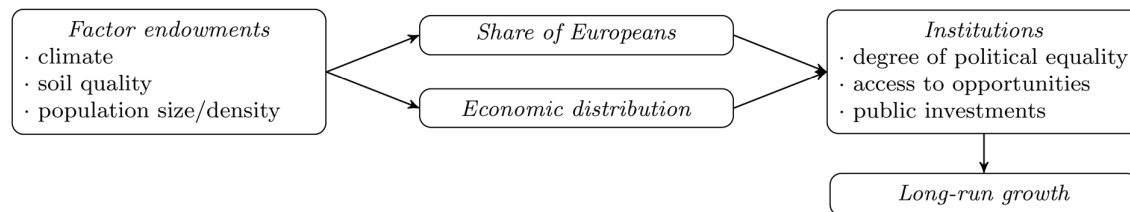
To summarize, the causal mechanism proposed by Engerman and Sokoloff (1997, 2002) is a combination of exogenous factors predetermining a development path, based on greater or lesser economic and political inequality, and endogenous institutional dynamics that maintain path-dependence over time. Exogenous 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 together with 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 where 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 along differing paths.

Countries with homogeneous populations<sup>18</sup>, 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 better 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 LINKS BETWEEN INEQUALITY, INSTITUTIONS AND LONG-RUN GROWTH



Source: own 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 where 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 driving the emergence of institutions that favor the owners of scarce factors of production.

<sup>18</sup> 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 to ethnic or racial homogeneity, where the European settlers have succeeded in marginalizing or even eliminating the indigenous population. In many 'heterogeneous' countries in South America relatively more of the indigenous population has survived.

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 followed a different logic (Fieldhouse, 1982). Colonizers were intervening less in the existing institutional structures of indigenous societies. Further, Africa was colonized comparatively late in the 19<sup>th</sup> century but has been a net slave exporter since the 15<sup>th</sup> century. 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 theories in which access to institutions (political equality) is a key feature of ‘high quality institutions’.

### 3.2.3 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:

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

$$I_i = \zeta + \beta \ln W_i + \mathbf{X}'_i \delta + v_i \quad (2)$$

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<sup>19</sup> declines by one standard deviation and schooling<sup>20</sup> 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 of nearly 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.<sup>21</sup> 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 causes economic inequality today could be driving these relationships, without necessarily having originated in economic inequality during colonial times.

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

---

<sup>19</sup> 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 Kaufman, Kraay and Zoido-Lobaton (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.

<sup>20</sup> Schooling is measured as the average of secondary enrollment rates from 1998-2003 with data from the World Development Indicators (WDI).

<sup>21</sup> Interestingly, the exclusion of the Americas increases the negative effects of the instrumented inequality coefficient as opposed to weakening the relationship.

in 2000. His baseline model begins with the assumed reduced form relationship between slavery, population density and income directly:

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

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.<sup>22</sup>

The effect of slavery on institutions is illustrated in the partial correlation plot below.<sup>23</sup> 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

<sup>22</sup> 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).

<sup>23</sup> 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.

modifies the specification in two ways to differentiate between plantation and non-plantation slavery (4) and to distinguish by size of slave holding (5):

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

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

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 the total population, (M) as the ratio for medium size slave holdings and (H) as the ratio for large slave holdings.

Differentiating between non-plantation and large plantation slavery 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 plantation 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.<sup>24</sup> 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 from individual U.S. counties and states 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:

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

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

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 U.S. 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, Nunn (2008a) 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 positively related to income per capita. For Nunn (2008a), these findings contradict the Engerman and Sokoloff (1997) hypothesis, which states that inequality negatively influences

---

<sup>24</sup> 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.

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 United States were determined nationally after 1864 and through competition with other states in the union. As a result, the U.S. south became more competitive and open than its counterparts with a stronger legacy of slavery in South America (Engerman and Sokoloff, 2002, p. 86). Therefore, causal inference based on the U.S. 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<sup>25</sup>, 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 subject to 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:

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

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.

---

<sup>25</sup> Ethnic fractionalization is an obstacle to development in its own right (Easterly and Levine, 1997).

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<sup>26</sup>, but robustly correlated with higher poverty. For medium-term outcomes, the only robust link is a negative association between 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. Overall land inequality has a negative sign but frequently remains insignificant. The land Gini is positively associated with the outcome variables, yet often insignificantly. Political 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<sup>27</sup>, 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. Last, while case studies add to our understanding of the processes involved they also introduce considerable complexity into the reasoning, which challenges parsimonious theory.

---

<sup>26</sup> 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.

<sup>27</sup> This concept draws on a previous paper by Acemoglu, Robinson and Verdier (2004) and incorporates insights from Bates (1981).



### 3.3 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 key criticisms of their approach and theory put forth by Albouy (2008), Glaeser, La Porta, Lopez-de Silanes and Shleifer (2004), and Austin (2008).

#### 3.3.1 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.<sup>28</sup> 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

---

<sup>28</sup> 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.

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, 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), their model can be summarized as a system of equations:

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

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

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

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

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 origin,  $M_i$  is the log mortality rate of the settlers,  $\mathbf{X}'_i$  is the transpose of a vector of covariates affecting all variables. Only  $M_i$ ,  $R_i$ ,  $y_i$  and parts of  $S_i$  and  $\mathbf{X}'_i$  are actually observed.

The advent of European colonialism can be regarded 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<sup>29</sup>. 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:

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

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

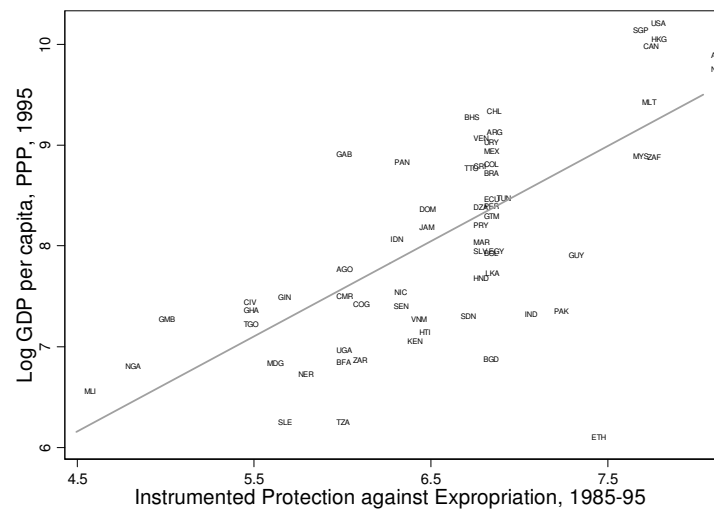
---

<sup>29</sup> 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.

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 decisive determinant of current institutions (e.g. La Porta, Lopez de Silanes, Shleifer and Vishny, 1999). However, Acemoglu et al. (2001) find that colonial identity 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 colonies only. Moreover, 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



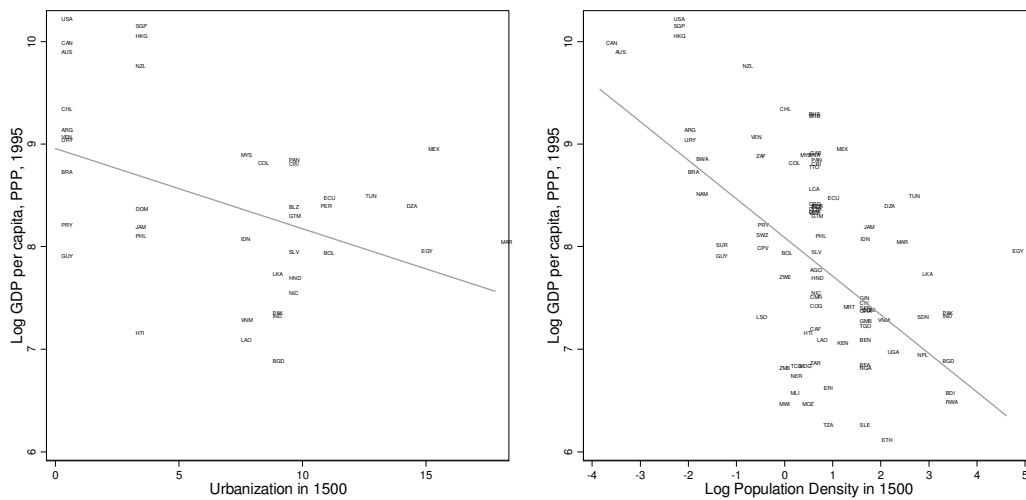
Source: computed using data from Acemoglu et al. (2001)

### 3.3.2 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.<sup>30</sup> 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 urbanization and income are strongly correlated when considering both cross-sectional and panel data since 1913. However, in theory population density is less robustly 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 to keep up with population growth.<sup>31</sup> 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.

FIGURE 5: URBANIZATION/ POPULATION DENSITY IN 1500 AND GDP PER CAPITA IN 1995



Source: computed using data from Acemoglu, Johnson and Robinson (2002).

<sup>30</sup> 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.

<sup>31</sup> Specifically, due to better hygiene etc. higher living standards led to more successful pregnancies and more surviving children. Before 1800 more prosperity 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 historical institutions literature is focused on how property rights, innovation and efficient production created the conditions for escaping these preindustrial dynamics.

Acemoglu et al. (2002) find that urbanization in 1500 and income per capita in 1995 are significantly negatively correlated (see Figure 5 above), although this relationship is visibly driven by the four major British offshoots (USA, CAN, AUS, NZL). 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, which just about accounts for the income difference between, for example, Uruguay and Guatemala in 1995. They repeat the analysis using log population density in 1500, its coefficient is -0.38 and also 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*, seemingly 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 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 hypothesis, which have been put forth in different variants by many authors (e.g. Lewis, 1978; Myrdal, 1968; Diamond, 1997; Sachs, 2001), 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.<sup>32</sup> Acemoglu et al. (2002) succinctly summarize the two hypotheses as follows:

$$Y_{i,t} = \alpha_0 + \alpha_1 \cdot G_i + v_t + \varepsilon_{i,t} \quad (14)$$

$$Y_{i,t} = \alpha_0 + \alpha_1 \cdot G_{i,t} + \alpha_2 \cdot T_t \cdot G_{i,t} + v_t + \varepsilon_{i,t} \quad (15)$$

---

<sup>32</sup> 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). Most 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.

where  $Y_{i,t}$  is GDP per capita in country  $i$  and time  $t$ ,  $G_i$  are time-invariant geographic characteristics (e.g. latitude),  $T_t$  is the state of technology at time  $t$ ,  $G_{i,t}$  are time-invariant and time-variant geographic characteristics (e.g. temperature or coal reserves),  $v_t$  is a general time effect, and  $e_{i,t}$  are country-time specific effects. The simple version (14) concentrates on  $\alpha_1$ , while the sophisticated version (15) argues that  $\alpha_2$  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)<sup>33</sup> and inclusive institutions (enabling broad participation in productive opportunities). Secure property rights are not sufficient for lasting development when they only apply to wealth-owning or political elites. 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).<sup>34</sup>

Extending their earlier work, they 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 share of Europeans also translated into a social stratification similar to their countries of origin where even lower strata would demand rights comparable to those present in their country of origin. Interestingly, the different types of settlers, such as aristocrats in Latin America versus indentured servants, puritans and pilgrims in the U.S. do not feature in their explanation. Moreover, they do not attribute a distinct role to the differences in (weapons) technology between Europeans and native populations, which varied sharply from the early colonization of the Americas 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 logarithm 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

---

<sup>33</sup> This is the main point of North and Thomas (1973).

<sup>34</sup> 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 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.

specifications the coefficient of institutions remains relatively large, positive and highly significant.<sup>35</sup> 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 18<sup>th</sup> and early 19<sup>th</sup> 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.<sup>36</sup> Second, they instrument their institutions measure with an interaction of log settler mortality and UK industrial output.<sup>37</sup> Using 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.

### 3.3.3 THE MODEL

FIGURE 6: THE CAUSAL LINK OF INSTITUTIONS AND LONG-RUN GROWTH



Source: own illustration.

<sup>35</sup> 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.

<sup>36</sup> The measure of institutions is 'constraint on the executive' from the Polity III database, as it has a long time-series dimension going back to the first year of independence.

<sup>37</sup> 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  $t$ ,  $\mu_t$  are time effects,  $\delta_i$  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.

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 many of the reported relationships do not apply, are very different, or the opposite in countries that were never colonized. Hence, colonialism is merely one of many critical junctures, albeit one of great significance.

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 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 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) construct 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 aimed at 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 not just the possible erosion of economic rents for elites motivates their resistance to technological progress, but primarily 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 19<sup>th</sup> century 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. However, Moore (1966) has famously argued that the fates of the landed elites in these countries were somewhat different; in Britain they were already ousted by the “bourgeois revolution” and in late 19<sup>th</sup> century Russia by the “peasant revolution”, while only in Germany and Japan they were able to “revolutionize from above”.

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 piecemeal concessions of power. Franchise extension acts as a credible commitment towards future redistribution to non-elites. 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 redistribution to themselves. If the poor are well enough organized to pose a threat to the regime but not well-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 with the support of 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 creating a ‘political Kuznets curve’.

### 3.3.4 CRITICAL JUNCTURES OR MODERNIZATION?

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. Lipset 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 (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 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 linking democracy to income levels since the 1960s, 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 claims that economic institutions cause development and under certain conditions also democracy, and not vice versa.

Acemoglu et al. (2007, 2008) argue that previous studies<sup>38</sup> have based their conclusions on cross-sectional correlations only and do not establish causality. They present an extended 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):

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

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

---

<sup>38</sup> Examples of such studies are Barro (1999) and the classic book by Przeworski, Alvarez, Cheibub and Limongi (2000).

of covariates,  $\mu_t$  is a set of time effects,  $\delta_i$  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.

Contrary to earlier studies, Acemoglu et al. (2007) find no causal effect of income on democracy. In all cases, they first report the pooled OLS estimates without fixed effects and find a positive coefficient on income<sup>39</sup>, 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. Using different democracy measures the effect of income 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 a 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.

### 3.3.5 CRITICISM AND ADDITIONAL EVIDENCE

The empirical and theoretical explorations of Acemoglu et al. (2001, 2002) attracted several criticisms directed both at their methodology and underlying theory. In the following, we concentrate on problems with the settler mortality data and instrumental variables method pointed out by Albouy (2004, 2006, 2008), the criticism of theory and method by Glaeser et al. (2004), and Austin's (2008) arguments against the "compression" of economic history. Taken together, these comments point the most pressing issues in the research of Acemoglu et al. (2001) and other studies using similar techniques or data, but they 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), they 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. Further, Acemoglu et al. (2001)

---

<sup>39</sup> 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.

mix rates from European soldiers on military campaigns with rates of soldiers in barracks. Albouy (2004) suggests that peace in the 19th century is correlated with income levels, thus confounding of these two rates makes settler mortality endogenous to the specification.

To remedy these concerns, Albouy (2004) constructs two alternative series based either on soldiers in barracks or on campaign, and compares the original 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.<sup>40</sup> 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. He concludes that while the theory is plausible, the empirical effect of institutions cannot be substantiated with the current settler mortality series.

Albouy’s concerns elicited multiple responses. Acemoglu et al. (2005, 2006) maintain in a point-by-point discussion of his modifications that their coding was consistent and present new evidence supporting the mortality rates they used. Much of the dispute concerns the assignment of mortality rates to countries in Africa and Latin America. To partially circumvent this issue, 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) also argue that Albouy’s distinction between soldiers in barracks or on 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 other 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) indicate 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. This lengthy debate focuses on many more individual coding issues which will not be discussed here and while these are serious data issues any reader should be aware of, we do not aim to adjudicate between the two positions.

Departing from a more general vantage point, Glaeser et al. (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

---

<sup>40</sup> 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.

today, such as educational attainment 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 also 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 expert assessments, exhibit high volatility, and reflect short-term electoral outcomes rather than 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 restrictions are not directly testable<sup>41</sup>, the theoretical argument about instrument validity is decisive. Glaeser et al. (2004) suggest that the instruments are systematically correlated to other factors affecting development outcome.<sup>42</sup> 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 settler mortality and indicators of legal structures, which are weak.

Interestingly, their third point – i.e. education predicts better institutions – also elicited another 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. They argue that this is due to other omitted factors driving the relationship in the specification without time effects, which incorrectly led Glaeser et al. (2004) to conclude that there is an effect. In fact, this conclusion may just reflect a general upward trend in the country scores on the institutions indicators and increases in school enrolment occurring over the recent decades. Not surprisingly, Acemoglu et al. (2005) interpret the results of their re-specification as a confirmation of their critical junctures hypothesis.

Austin (2008) critically analyzes the ‘reversal of fortunes’ hypothesis from the perspective of economic history with a focus on Africa. Acemoglu et al.’s (2002) finding of a reversal in relative prosperity among former colonies crucially depends on the quality of their historical

---

<sup>41</sup> 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 construction and  $E[\varepsilon_i | Z_i] = 0$  only by assumption.

<sup>42</sup> This point was first raised in Djankov, Glaeser, La Porta, Lopez-de Silanes and Shleifer (2003).

data. Just as the mortality figures, their proxies for initial prosperity are based on statistics that were assembled by other authors from a variety of sources. Austin (2008) points out that for Sub-Saharan Africa the population estimates are extremely crude backward projections based on colonial enumerations undertaken in the late 19<sup>th</sup> and early 20<sup>th</sup> century, which questions their reliability. Interestingly, Austin (2008) still confirms that Africa was more densely populated and urban than the ‘neo-Europes’ but cautions that such statements must inevitably rely more on circumstantial evidence (e.g. the implied population effects of land abundance and imported crops) rather than sound data.

More fundamentally, Austin (2008) suggests that regressions which span nearly five centuries of economic development and history require far-reaching assumptions towards how these periods relate. A key pillar of the theory by Acemoglu et al. (2001, 2002) is the assertion that Europeans imposed extractive institutions wherever they did not settle in sufficiently large numbers. However, most of Africa was colonized late and in some cases very briefly, implying that European colonizers must have been able to impose extractive institutions easily and potentially even before official colonization (for example, through slave trade). Austin (2008) notes several problems with this reading of history. On the one hand, even within Africa there was a legal distinction between settler and non-settler colonies, which did not evolve along the number of European settlers but their total land use and affected how these colonies developed. On the other hand, while colonialists intervened in traditional property rights and production structures, these interventions were all subject to African influence and sometimes reversed entirely by indigenous perseverance. For example, Austin (2008) argues that the prominent British ‘West Africa Lands Policy’ which maintained traditional ownership rights in non-settler colonies was born out of the realization that the indigenous African cash crop production was working well within these communal property rights, and ran counter to the initial British policy of championing private ownership. In other words, Acemoglu et al. (2001, 2002) focus only on the colonizers and neglect the colonized, although in Africa the latter outnumbered the former by at least three orders of magnitude.

Another key assumption in the theory and empirical models of Acemoglu et al. (2001, 2002) is that rent-seeking (extraction by European colonizers) and economic growth are mutually exclusive. Contending this notion, Austin (2008) suggests that not all kinds of rent-seeking are necessarily detrimental to growth and cites the example of the Caribbean island colonies’ economic success until the abolition of slavery. However, Acemoglu et al.’s (2002) argument is more refined; they point out that the negative effects of extractive institutions only became apparent with the industrial revolution, when physical and human capital became relevant inputs to production. Nevertheless, the relationship between rents and economic growth is likely to depend on who receives the rent, the type of rents, and what they are used for – themes to which we return in Section 3.5.

Together these criticisms raise three important and general points which the subsequent empirical literature on institutions and growth should 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 of interest. Third, empirical modeling spanning several centuries of history may invite appealingly parsimonious conclusions, but its theoretical implications are less straightforward and such studies do not necessarily disprove other potentially relevant causal mechanisms.

### 3.4 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. They have 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-run, medium-run and short-run growth.

#### 3.4.1 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 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 growth. 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 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.<sup>43</sup> Geography is exogenous.

---

<sup>43</sup> 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 the two countries, and  $S_i$  and  $S_j$  are measures of country size. Frankel and Romer (1999) drop observations where no bilateral trade

In their model, institutions are also 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 level of trade integration with equation (19):

$$\ln y_i = \mu + \alpha R_i + \beta N_i + \gamma G_i + \varepsilon_i \quad (17)$$

$$R_i = \lambda_R + \alpha_R \ln M_i + \beta_R B_i + \gamma_R G_i + v_{R_i} \quad (18)$$

$$N_i = \lambda_N + \alpha_N \ln M_i + \beta_N B_i + \gamma_N G_i + v_{N_i} \quad (19)$$

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,  $G_i$  a 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 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-Lobaton, 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 the using the linguistic measures) and the linguistic instruments in the third sample do pass the over-identifying restrictions.<sup>44</sup>

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 relationships between institutions, trade and geography, and between trade, institutions and geography, respectively:

---

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.

<sup>44</sup> If there is more than one instrument for one endogenous variable, then 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 known to have low power.



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

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

To estimate the total effect of each variable, they separately apply a unit shock to the error terms of the trade and institutions equations.<sup>45</sup> 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 institutions, and the quality of institutions has both the largest direct and total effect on per capita incomes.

In addition, 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 – or any – role in the causal relationship. They underline this argument with an analogy. Angrist and Krueger (1991) use when a person is born within a given year (the quarter of birth) as an instrument for estimating the effect of years of schooling on subsequent 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 an unbiased estimate of the returns to schooling. However, this does not imply a quarter of birth related theory of earnings. Similarly, according to Rodrik et al. (2004), Acemoglu et al.’s (2001) strategy does not directly test a theory of colonial origins. Instrumental variables thus require a theory or ad hoc explanation justifying that they are exogenous, but precisely this underlying link is not tested within the same framework.

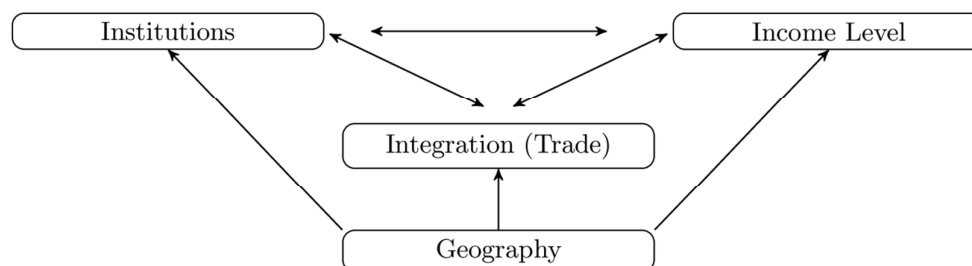
---

<sup>45</sup> 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.

### 3.4.2 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 impact on income, and higher income can result in more trade integration. Trade also influences 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).

The one-way and two-way arrows below 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 meta-model 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 comparatively less.

Rigobon and Rodrik (2005) test a very similar model which allows for reverse dependencies 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.<sup>46</sup> Further, instead of using

<sup>46</sup> They proxy for the rule of law with the corresponding indicator from the World Governance Indicators (see,

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 statistically much more significant than democracy. 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 and trade are significant but comparatively small, while democracy and rule of law are positively interdependent.

Since Rodrik et al. (2004) also use instrumental variables techniques based on the settler mortality instrument, their work is subject to the criticisms we discussed in the previous section. In addition, while their approach of comparing different ‘fundamental causes’ of growth is interesting and novel, it also introduces several new problems into their research design. The instruments for institutions (settler mortality) and trade openness (predicted trade shares) are both at least influenced or directly derived from geographic characteristics (e.g. proximity to Europe, climate, distance between trading partners). This potentially conflates the true effect of geography, which could operate through these variables in different ways than suggested by the analysis. For example, Rodrik et al. (2004) are not able to account for versions of the ‘sophisticated geography hypothesis’, which requires variation over time. Hence, it is not clear how successful their system of equations is in actually differentiating between the direct and indirect effects of trade, institutions and geography (for a more similar but more elaborate argument along these lines see Dixit, 2007).

### 3.4.3 GROWTH COLLAPSES, EXTERNAL SHOCKS, AND GROWTH ACCELERATIONS

Much of the research presented 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<sup>47</sup> and indicators of quality institutions as level 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). Level regressions measure the cumulative impact of all historical growth-enhancing or growth-constraining policies. Hence, it is obvious that 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 contemporary growth further, Rodrik argues that we should distinguish between 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.

---

for example Kaufmann et al., 2002) and for democracy with the composite indicator from Polity IV (Marshall and Jaggers, 2003).

<sup>47</sup> Although per definition GDP is a flow not a stock. However, it builds on the output potential of an economy and can therefore be interpreted as a level or stock which is the sum of past growth in output potential.

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<sup>48</sup> 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 miracle in East Asia, but the relative decline elsewhere.

To explain these growth collapses, Rodrik (1999) proposes to conceptualize the economic turbulence of the 1970s not as merely an effect of external shocks<sup>49</sup> (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:<sup>50</sup>

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

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.<sup>51</sup> Rodrik (1999) 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

---

<sup>48</sup> 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.

<sup>49</sup> Here we mean shock in the sense of an abrupt and large change, and not a unit change as before.

<sup>50</sup> Essentially, this relationship is a summary of the results of a formal model provided in the working paper version of the 1999 article (Rodrik, 1998a).

<sup>51</sup> 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.

(Freedom House). Rodrik (1999) first tests an additive linear specification, which can be generalized as follows:

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

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).

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, but when the quality of 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:

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

where notation is as before and all measures of institutions are standardized to lie within the unit interval [0,1].

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 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.

In many ways, Rodrik's (1999) paper is an important early contribution to a rapidly growing literature on trend-breaks and advanced an interesting hypothesis regarding the interaction of institutions with inequality and growth outcomes. By that virtue, it also leaves many relevant econometric issues unaddressed. While Rodrik (1999) makes no casual claims, he still derives strong conclusions from partial correlation plots based on simple cross-sectional regressions. Moreover, income inequality data in general is well-known to be notoriously flawed; a situation which is only slowly improving with more wide-spread data collection and harmonization efforts. Last, Rodrik's index of total shocks combines two variables (inequality and democracy) whose individual effects on growth are still debated in the literature without formally testing the added explanatory power of interactions between them.

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 potentially be more informative than evidence based on average growth performance (Pritchett, 1998). Hausmann et al. (2005) argue that both neo-classical and endogenous growth theory evolve around the idea of shifting growth paths, comprising 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, their research aim to answer 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 regressions with trend-breaks to identify the start of accelerations if there is more than one year as a candidate. They use GDP per capita data from the Penn World Tables from 1950 to 1999; hence, the first episode can begin in 1957 and the last in 1992.

Their filter results in 83 growth accelerations which include 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 below). However, the number of observations also varies by time and region. If early data availability is correlated with experiencing accelerations, then these we must contend that these tabulations obviously exhibit an upward bias.

TABLE 2: FREQUENCY OF ACCELERATED GROWTH EPISODES

Decade	<i>Region</i>						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, p. 310).

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

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.<sup>52</sup>

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, 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<sup>53</sup>, the death of an incumbent leader<sup>54</sup>, recent armed conflict or civil wars<sup>55</sup>, economic liberalization<sup>56</sup>, and financial liberalization<sup>57</sup>.

Only 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 these results can only be understood after differentiating further between sustained and unsustained acceleration and re-estimate the previous specifications. Sustained growth is defined as growth in excess of 2% per annum in the ten years after the acceleration episode (8 + 10 = 18 years in total) and growth falling below 2% per annum in 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

---

<sup>52</sup> For example, taking a 5-year horizon results in 137 growth episodes and taking a 10-year horizon result in just 37 episodes. If the rapid growth threshold is raised to 4% per annum, then 68 episodes are identified, and if it's lowered to 3%, then 90 episodes are identified.

<sup>53</sup> From Marshall and Jaggers (2003). Hausman et al. (2005) define a regime changes as a 3 unit change in the Polity IV score, but they mistakenly code this variable for *any* change (see Jong-A-Pin and De Haan, 2011).

<sup>54</sup> A dummy variable with a five-year unity value beginning with an incumbent leader's death.

<sup>55</sup> 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.

<sup>56</sup> Defined as a five year dummy capturing a transition to openness based on Sachs and Warner (1995).

<sup>57</sup> 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).



sustained growth, positive regime changes are significant for sustained but not unsustained episodes, negative regime changes remain significant, and financial liberalization is only related to unsustained growth episodes. The signs of these effects are all positive.

Many of the above results are more intuitive. Terms of trade shocks and financial liberalization can 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 is 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 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. Alternatively, as argued by Meisel and Ould Aoudia (2008) but also North et al. (2009) and Khan (2010), regime changes could be secondary to the political settlement they contain, which may be more autocratic but politically viable and growth-enhancing at the same time. Hausmann et al. (2005) conclude on a cautionary note by warning that the determinants of growth accelerations are not well-identified in their specifications, as their models too often incorrectly predict the outcome.

The positive effects negative regime changes are arguably the most controversial result of the ‘growth accelerations’ perspective advanced by Hausman et al. (2005). To address this issue in more detail, Jong-A-Pin and De Haan (2011) reevaluate the links between political regime changes, economic liberalization and growth accelerations. They argue that the effect of negative regime changes in Hausmann et al. (2005) was driven by a coding error and disappears once it is corrected, but they in turn find evidence that democratic regime changes reduce the probability of growth accelerations. Further, they show that economic liberalization precedes growth accelerations on average and that regime durability (the time a regime did not substantively change) is negatively associated with accelerations. However, both the original study by Hausmann et al. (2005) and most subsequent similar research take few – if any – measures to limit the endogeneity of the included variables and remain very sensitive to routine revisions in the underlying GDP data (Johnson et al., 2009).

In sum, two main results emerge from this line of research. First, igniting growth is relatively easy as can be seen in the relatively rather large number of strong growth spurts across several decades and continents. Second, these accelerations are not necessarily preceded or well-predicted by changes in political structures, economic reforms or other institutional changes, which yield low explanatory power, and thus appear to be primarily driven by idiosyncratic factors. In the next section, we present one possible approach to analyzing what these country-specific constraints of short-run and medium-run growth could be.

#### 3.4.4 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 testament 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, Anne Krueger, and Jeffrey Sachs are all correct – at different times and under specific circumstances” (Rodrik, 2010, p. 35). This view especially evolved after the dissatisfactory results of the Washington consensus, which reduced the vector of possible growth strategies to a clearly defined list of quintessential reform strategies, which we discussed in Section 3.1. The emerging long-run 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, such as the binding constraints framework.

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 interaction 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:

$$\frac{du}{d\tau_j} = -\lambda_j + \sum_i \lambda_i \frac{\partial[\mu_i^S(\tau, \dots) - \mu_i^P(\tau, \dots)]}{\partial \tau_j} \quad (24)$$

where  $u$  is the welfare of an average member of the economy,  $\tau_j$  or  $\tau_j$  is the tax-wedge or distortion on activity  $j$  or  $i$  with ( $i \neq j$ ),  $\lambda_j$  or  $\lambda_i$  is the direct cost/benefit of distortion  $j$  or  $i$ ,  $\mu_i^S(\tau, \dots)$  is the social value of activity  $i$  after all taxes and all distortions, and  $\mu_i^P(\tau, \dots)$  is the corresponding private valuation.

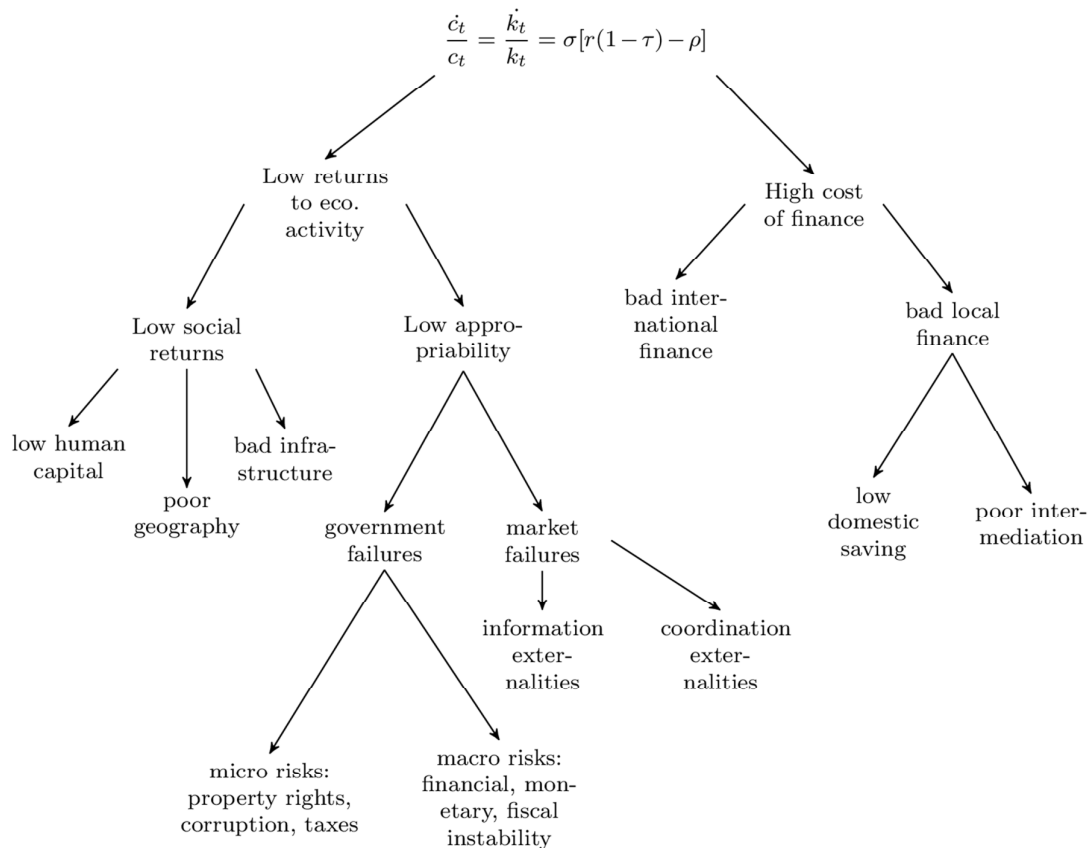
The framework captures the simple idea of intertwined and differently sized distortions while remaining suitable to incorporate almost any growth model. For example, inadequacies of certain institutions linked to any activity could be considered part of the distortion or a separate condition driving 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 affect 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 straightforward to derive. 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  $\lambda_j$  and ignore the cumulative or second-best effect of the summation over  $\lambda_i$ 's. So what are binding constraints? Essentially nothing else than very

large direct effects ( $\lambda_j$ 's) which according to Hausmann et al. (2005, 2008) also implies that the indirect effects might not outweigh a reduction in the constraint. To illustrate why this matters, 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 ( $\tau_i$ )”: 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.

FIGURE 8: GROWTH DIAGNOSTICS



Source: Hausmann, Rodrik and Velasco (2008).

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 it 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 the strategy. Hausmann et al. (2005, 2008) further acknowledge that identifying 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. They summarize this ‘growth diagnostics’ approach with a decision tree (Figure 8).

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:

$$\frac{\dot{c}_t}{c_t} = \frac{\dot{k}_t}{k_t} = \sigma[r(1 - \tau) - \rho], \quad \text{with } r = r(\alpha, \theta, x) \quad (25)$$

where  $c$  is consumption,  $k$  is capital,  $\sigma$  is the intertemporal elasticity of consumption,  $r$  is the return on capital,  $\tau$  is the tax on capital (formal/informal), and  $\rho$  is the world interest rate. Further,  $r$  depends on total factor productivity ( $\alpha$ ), an index of externalities ( $\theta$ ) 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)  $\rho$ , which is the cost of finance. A high cost of financing 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. Essentially, the framework characterizes countries as either savings or investment constrained. Four variables determine a country’s situation: (1) high  $\tau$  – high taxes, inefficient tax systems, or high risk of expropriation, (2) high  $\theta$  – large externalities, coordination failures and spillover effects, (3) low  $\alpha$  – 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.

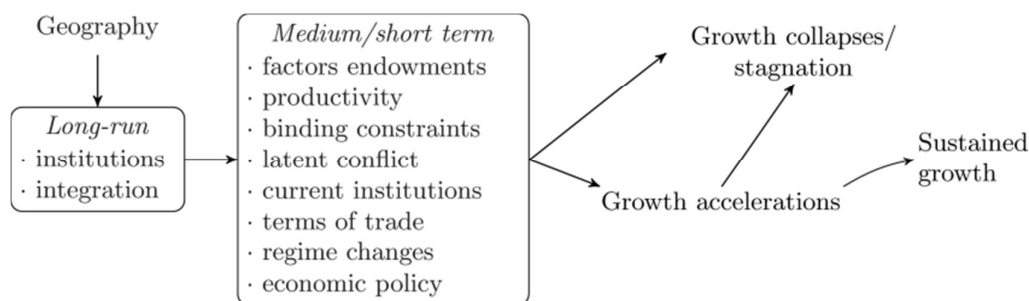
The framework succeeds in combining many macroeconomic and microeconomic interactions, while hierarchically organizing the basic conclusions of a large amount of modern economic theory on factor accumulation, learning and spillovers, externalities, institutions, financial markets, taxation, and government or market failures, and more. We present it here mainly with 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

makes clear that for growth theory, evidence and policy analysis the time-frame matters crucially for the results the research will produce and their relevance to policy-making.

### 3.4.5 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: own illustration.

First, the interdependencies among long-term factors have been well-established by the reviewed literature. For Rodrik et al. (2004) in particular, institutions are the most crucial of all long-run determinants. In fact, they imply that 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. Moreover, these long-run development paths in part determine the current state of institutions, the technology (productivity) gap, 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 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 of economic actors and thus link their factors endowments with the components of the growth diagnostics framework. Hausmann et al. (2005) have shown that it is easy to ignite growth, which can happen through policy changes, changes in factor proportions, institutional changes and many factors not 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.

Last, Rodrik (2000) emphasizes that institutional functions do 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 ownership 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.<sup>58</sup>

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. The latter interpretation is clearly more positive compared to the long-run literature, which highlights deterministic development paths and – if taken literally – suggests that it is very difficult to escape history.

### **3.5 POWER, RENTS AND ECONOMIC GROWTH**

In this section, we review two recent contributions by North, Wallis and Weingast (2009) and Khan (2010). These authors place violence and power at the heart of economic development and argue that economic rents are a crucial mechanism for creating social stability and establishing a political order. Contrary to the long-run literature reviewed earlier, these theories shift the focus away from property rights and egalitarian or extractive institutions towards intra-elite coalitions, the adaptive capacity of institutions and the nature of informal institutions as key determinants of growth and volatility. In this manner, they both link the diversity of contemporary growth experiences to the prevailing political structure. However, while North et al. (2009) emphasize how these relationships evolved historically, Khan (2010) focuses on the policy-problem of how to promote growth in the medium-term.

#### **3.5.1 LIMITED AND OPEN ACCESS ORDERS**

In an influential book, North et al. (2009) develop a framework for integrating the issue of monopolized control over violence into the study of socio-economic development and economic history. They distinguish modern societies where all citizens enjoy open access to economic and political organizations from closed-access societies in which the distribution of economic returns and access to organizations is heavily skewed towards elites. Their core argument consists of three elements: a pattern of contemporary cross-country differences in

---

<sup>58</sup> 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.

development, a new taxonomy to explain this pattern and historical narratives suggesting that this pattern holds throughout time. In the following, we review the first two, but make only limited reference to the historical examples.

TABLE 3: GROWTH, ORGANIZATIONS, DEMOCRACY AND GOVERNMENT SPENDING BY GDP

GDP per capita in 2000	GDP per capita growth (1950-2004)			Organizations (1973-2002) <sup>+</sup>		Democracy (2000)	Government (1995-2000)
	Years positive	Average when positive	Average when negative	Per million residents	Percent of world	Combined Polity IV score	Average share of GDP
\$20,000+ (no oil)	84%	3.88%	-2.33%	63.6	82.8%	9.42	53%
\$20,000+ (incl. oil)	81%	4.19%	-3.49%			6.31	
\$15,000-\$20,000	76%	5.59%	-4.25%	26.9	1.9%	6.86	33%
\$10,000-\$15,000	71%	5.27%	-4.07%	21.2	1.9%	5.67	40%
\$5,000-\$10,000	73%	5.25%	-4.59%	16.7	6.3%	3.70	33%
\$2,000-\$5,000	66%	5.39%	-4.75%	4.5	3.8%	1.18	27%
\$300-\$2,000	56%	5.37%	-5.38%	2.8	3.3%	0.31	31%

*Sources:* North, Wallis and Weingast (2009, pp. 3-10), own addition of average Polity IV score from Marshall, Jaggers and Gurr (2010). Original data for income from Heston, Summers, and Aten (2006) for the period from 1950 to 2004, for organizations from Coates, Heckelman, and Wilson (2007), and for government spending from the IMF Government Finance Statistics (GFS).

*Notes:* North et al. (2009) do not specify how they classify countries as major oil producers. For the added Polity column, we coded countries as major oil producers when net oil exports are equal to or more than one third of total exports. Whenever a cell spans two rows, no separate data was available. The number of countries available in each category differs per column and data source. <sup>+</sup> North et al. (2009) do not specify whether they are using averages or a single point in time from the Coates et al. (2007) data.

North et al. (2009) begin by contrasting the average growth performances of high income and lower income<sup>59</sup> countries from 1950 to 2004 (see Table 3 above). During this period, countries with a per capita income above \$20,000 in the year 2000 experienced moderate positive rates of growth during expansions and moderate negative rates of growth during contractions. In contrast, countries with a per capita income below \$20,000 experienced faster positive growth than the rich countries but had fewer years of positive growth and higher negative growth rates during contractions. These differences in growth patterns can be interpreted as follows. On the one hand, faster growth at lower income levels is consistent with the notion of catch-up (Hausmann et al., 2005a) and advantages of technological backwardness (Gerschenkron, 1962). On the other hand, the pattern of more and deeper growth contractions illustrates the vulnerability of low income countries to shocks. Poorer countries experience more frequent and deeper growth collapses (Rodrik, 1999).

Another salient feature of Table 3 is the positive correlation between the number of organizations per million residents and income per capita. The numbers refer only to formal trade and business organizations, which function as a crude proxy for overall organizational density. We can observe two discontinuities. First, countries with a per capita income above \$20,000 can support a large variety of organizations – more than twice as much as those with an income between \$15,000 and \$20,000. Second, countries with a per capita income below

<sup>59</sup> High income denotes a GDP per capita above \$20,000. Lower income refers to countries below this mark.

\$5,000 have less than five organizations per million residents, which is approximately four to five times less than in the income range from \$5,000 to \$20,000 and more than 14 times less than the richest countries. In fact, the richest nations have only 13% of the world population but are home to 82.8% of all organizations.

Democracy is also positively correlated with per capita income levels. 21 of the 30 richest countries in 2000 are non-oil producing countries. Except France and Singapore, all of these 21 are tied for the highest Polity IV ranking of democratic institutions (North et al. 2009, p. 4). In Table 3, we added the combined Polity IV scores<sup>60</sup> to the income groups used by North et al. (2009). The general pattern supports their contention that democracy and income are positively associated. The most developed countries also have the highest democracy scores. However, including the major oil producing countries<sup>61</sup> in the high income category results in much lower average Polity scores compared to the high income category without the major oil producers. Their Polity scores are among the worst of all countries in the database.<sup>62</sup>

Similarly, the degree of government expenditure relative to gross domestic product is higher at higher levels of income. This relationship does not hold when examining central government expenditures only, but emerges once sub-national governmental expenditures are included in total government expenditures. According to North et al. (2009), this suggests that subnational governments in high income countries are bigger because they provide more public goods (infrastructure, health, education, etc.) and support denser networks of subnational organizations. In Table 3, we find some support for this assertion. There is a substantial discontinuity between countries with a per capita income above \$20,000, which have an average government share of 53%, and all those below, which have an average total government share of 32% of GDP.<sup>63</sup>

For North et al. (2009), the contemporary pattern of developmental differences can be schematized in terms of the difference between *open access orders* and *limited access orders*. Open access orders are economically and politically highly developed. They experience few and mild episodes of negative growth, have active civil societies with many organizations, large decentralized governments, impersonal relationships, strong rule of law, well-protected property rights and a shared sense of equality. Limited access orders, on the contrary, are characterized by volatile growth patterns, polities without broad democratic consent, few organizations, small and centralized governments, personal relationships, weak and unequally enforced rule of law, insecure property rights and high levels of inequality.

---

<sup>60</sup> The combined Polity score ranges from -10 to +10 and is constructed by subtracting a country's autocracy score from its democracy score, both of which are on a range from 0 to 10. The revised combined Polity score, which we used in Table 2, assigns codes within the scale to interregnums and other transition periods which are otherwise coded outside of the 0 to 10 scale (Marshall, Gurr, and Jaggers, 2010, p. 17).

<sup>61</sup> We classified countries as major oil exporters if fuel exports made up at least one third of their total merchandise exports in 2000 based on the World Development Indicators 2011.

<sup>62</sup> Major oil producers with a GDP per capita above \$20,000 in 2000 are Bahrain, Oman, Kuwait, the United Arab Emirates and Qatar. Their combined Polity IV scores are -9, -9, -7, -8, and -10, respectively.

<sup>63</sup> This is a country-weighted average of all income groups below a per capita GDP of \$20,000. Computed using the country counts from North et al. (2009, p. 10).



Limited access orders were the prevailing social order for most of human history and still dominate the much of developing world today – hence, the authors also refer to them as *natural states*. North et al. (2009) claim that natural states follow an entirely different logic than open access orders. The natural state does not possess consolidated control over organized violence. Natural states are governed by elite factions who each control parts of the military, militias or other potentially violent groups. However, this system of factionalized rule is not an accident. It emerged as a pragmatic form of social organization aimed at limiting the losses occurring from repeated outbreaks of violence in forager societies. In order to create a measure of social stability, elite factions agree to share economic resources according to their violence potential and to limit access to key political and economic activities for non-elites by forming a ruling coalition. The monopolistic rents created in this process of limiting access provide incentives for the elites to abstain from violence. These incentives control violence and enable greater production as long as the distribution of political power remains largely unchanged.

According to North et al. (2009), open access orders solve the problem of controlling violence differently. As opposed to partially or fully limiting access to economic and political opportunities, they open access to economic organization and political activity entirely and make participation in both an impersonal right independent of the distribution of political power. While natural states aim to create stability directly through distributing rents, open access orders create stability indirectly by institutionalizing a peaceful process of rent creation through innovation and rent erosion through competition. However, establishing open access is not possible without first consolidating a society's violence potential under the control of a Weberian bureaucracy, establishing strong organizations and rule of law for elites. Otherwise, there is no guarantee that factions do not obtain the desired resources by force. Based on this reasoning, North et al. (2009) categorize history into three general types of social orders: foraging orders, natural states and open access orders. They further distinguish three types of natural states: fragile, basic and mature. Table 4 below summarizes the main characteristics of these social orders.

TABLE 4: TAXONOMY OF SOCIAL ORDERS

	Forager	Natural States with Limited Access			Open Access
		<i>Fragile</i>	<i>Basic</i>	<i>Mature</i>	
<i>Control of Violence</i>	individual	highly factional	factional	factional/consolidated	consolidated
<i>Elite Organizations</i>	-	few state	state only	state/ few private	public/ private
<i>Organizational Lifespan</i>	-	short	semi-durable	durable/perpetual	perpetual
<i>Rule of Law</i>	-	simple rules	coalition	all elites	universal
<i>Relationships</i>	personal	personal	personal	less personal	impersonal

Source: own illustration.

The *forager societies* before the Neolithic revolution (around 8000 BC) were primitive societies, locked in a Hobbesian state of dispersed and individualized violence. They receive

no further attention in the framework. From the first agricultural transformation around 8000 BC until the 19<sup>th</sup> century the world consisted primarily of natural states.

*Fragile natural states* are highly unstable social orders, subject to repeated outbreaks of internal and external violence. There is a fragile balance between the distribution of political power and economic interests in the ruling coalition. Rule of law is a mix of very simple rules in private and public law on which patrons can base their decisions. Public law combines the personal identity and social function of individuals. Legal rules are neither equal across groups nor equally applied across individuals. Patron-client networks dominate political and economic activity and few organizations can be sustained, even within the state. The average lifespan of most organizations is short and linked to the identities of their creators. Even minimal shocks can create recurrent outbreaks of violent conflict over the distribution of economic resources.

*Basic natural states* are more stable, have more developed public law institutions and some social expectations constraining the behavior of elites. Processes such as leadership succession, changes in the coalition, and changes in economic interests need not necessarily lead to violence. Public institutions (courts, administrative bodies, legislative bodies) exist to structure the behavior of elites and to provide organizational forms for elite competition. However, law does not apply equally to all groups of elites but is still differentiated according to political power.

*Mature natural states* can support elite organizations outside of the state and the violence potential is less fractionalized or even close to being monopolized. In their most developed versions, natural states apply public law equally across all elites regardless of their power, which allows for a longer lifetime for elite organizations. They may be formally democratic, but do not have a political system that truly engenders competitive politics and protects political losers. Similarly, access to economic and political opportunities remains restricted for broad segments of the population.

Natural states develop up to the so-called *doorstep conditions* before they can make a transition from natural states to open access orders, which we discuss further below. However, there are no strict dividing lines in this classification. On the one hand, much of the distinction between basic and mature natural states depends on whether or not legal personhood is assigned to organizations and whether elites are allowed to organize outside of the state, although these characteristics do not develop equally across all types of organizations (e.g. corporations, political organizations, and civic associations). On the other hand, all limited access orders share some common features. For example, all natural states restrict access to organizational forms, control trade, and define their legal system to defend elite privileges.<sup>64</sup>

---

<sup>64</sup> The last point is more general and extends far beyond natural states. In fact, North et al. (2009) propose that the origin of *all* legal systems can be traced to elites agreeing on their privileges. This follows from the assumption that all open access orders were natural states at some point and elites first created rights for themselves which were only extended to the general population later on.

From the 19<sup>th</sup> century onwards, North et al. (2009) identify the first transitions from limited access towards institutionalized *open access* orders. Britain, France and the United States were the first countries to make the transition. Contrary to natural states, open access orders have universal rule of law for all individuals independent of their political power. The state, corporations and many other civic or political organizations are *perpetually lived organizations*. These modern types of organizations are separated from the identity of their creators, exist beyond the creators' physical lifetime, and are endowed with the legal rights of a physical person. The possibility to form organizations must be impersonal and open, not only by legal statute but also by belief and custom. As a result, cities can sue states, corporations can sue nations, and any legally associated interest group can and often does legally defend its interests against other organizations or groups. Further, open access orders are democracies based on competitive politics<sup>65</sup> and their armed forces are under consolidated civilian (i.e. parliamentary) control.

In a companion paper, North, Wallis, Webb and Weingast (2007) provide contemporary examples for each type of social order. Haiti, Iraq, Afghanistan, Somalia and other conflict ridden countries with dysfunctional governments are fragile natural states. Burma, Cuba, North Korea and many countries in the Arab world, Sub-Saharan Africa and former Soviet countries are basic natural states, while most of Latin America, South Africa, and India are mature natural states. Many countries in the European Union and all of the Western democracies are open access orders – the smallest and richest group. Although they do not explicitly categorize China, it is most likely either a mature natural state or on the verge of the transition to open access, as it is increasingly opening its markets but still limiting access to political organizations. North et al. (2007) also provide estimates of the likely ranges of GDP per capita for limited access orders (\$400 to less than \$8,000), transition countries (\$8,000 to less than \$20,000) and open access orders (\$20,000 and above).

#### *Underdevelopment and natural states*

It is inherent in the logic of the natural state that it cannot exceed certain levels of per capita income as long as it remains a natural state.<sup>66</sup> The size of the dominant coalition is limited and so is the organizational capacity of natural states. Adding an additional member to the ruling coalition decreases potential rents per member of the coalition and dissipates the gains for elites created by limiting the access of non-elites to resources and organizations. Monopoly rents per member naturally decline as the number of rent recipients increases. Hence, shifts in military technology, trade shocks, deaths of leaders or other factors affecting the distribution of political or economic power threaten the stability of the social order. People outside of the coalition, or non-elites, will not find many of their agreements honored, particularly because they have no means of legal recourse towards elites. This provides disincentives for broad-based economic participation and result in a mismatch between social and private returns, which acts as an obstacle to economic development. Coalitions in natural states purposefully control markets, as opposed to creating a complementary framework for

---

<sup>65</sup>Specifically, “formally organized parties do not necessarily imply competitive politics, nor do formal definitions of corporations imply open access to economic organizations” (North et al., 2009, p. 248).

<sup>66</sup> As we noted earlier, an exception are the major oil producing countries, who distort this relationship.

their functioning. In a highly personalized social order, markets and organizations are primarily a means of political control.

Accordingly, the prime reason that natural states appear to be corrupt to outside observers is that the most “important basic natural state organizations are closely associated with the (private) individual identities of the elites who inhabit them” (North et al., 2009, p. 73). Only as natural states mature and relationships gradually become more impersonal does a clear distinction between public and private organizations emerge. However, as external and internal shocks can easily upset the balance of political and economic power at any time, most societies have progressed forwards and backwards through the varying types of natural states. North et al. (2009) stress that natural states are not static and not progressing linearly towards more mature and open social orders. They are subject to tremendous change. The only aspect that remains constant is the mechanism of limiting access to economic and political organizations and opportunities in order to create rents and stability.

#### *Development and open access orders*

In the long-run, the development potential of open access orders is much higher than that of natural states. By design of their rent-power balance, open access orders handle change and increasing complexity better than natural states. They are more flexible and adaptive. Hence, they are substantially more productive than natural states and have larger governments which are able to support complex markets. North et al. (2009) trace this ability to Schumpeterian creative destruction and open competition in both politics and economics. In its original sense, creative destruction refers to innovating entrepreneurs entering the market-place or creating new markets and thereby destroying the rents of the monopolistic or quasi-monopolistic competition (Schumpeter, 2003[1942]). They extend this logic to the political system. Under conditions of open entry, liberal capitalism and democracy engender a simultaneous process of constant rent creation and destruction. Consolidated control of violence, rule of law for all, impersonal relationships and open access to organizations together guarantee that a) no single group can appropriate economic rents without facing counter-organization of competing interests and b) the process of constant change of economic interests is internalized in the political system because it is reflected in the (potential) political organization of new or changing interests. In an earlier work, North calls this capacity *adaptive efficiency*, which he defines as the ability of an institutional structure to “adapt to the shocks, disturbances, and ubiquitous uncertainty that characterize every society over time” (North 2005, p. 78). In open access orders additional organizations and changing interests have a positive and multiplicative social return, in contrast to imposing costs on social stability in natural states.

#### *The problem of transition*

If the underlying logic of open access orders is different from that of natural states, the most crucial question about long run economic development still remains to be answered. How do societies move from natural states to open access orders? North et al. (2009) provide a clear answer to the conditions required before the transition can occur, but they remain vague on how the actual transition process takes places in practice and which particular incentives

motivate elites to completely change the structure of a society's political and economic institutions.<sup>67</sup>

Before societies can make a transition to open access they need to fulfill three doorstep conditions: (1) establishment of impartial rule of law for elites, (2) the emergence of public and private perpetually lived organizations, and (3) consolidated control of the military. North et al. (2009) suggest that, historically, these characteristics have evolved in this precise sequence, although this need not hold true for future transitions. After these conditions are in place, the 'transition proper' can take place. It is a two-step process in which ruling elites first convert their personal privileges into impersonal rights for all elites and then extend these rights to broader segments of the population. However, they emphasize that reaching the doorstep conditions does not imply an automatic transition to open access.

The rule of law for elites emerges out of relative political stability and routine interactions among elites. In natural states, public law is devised to structure repeated transactions but initially remains strongly biased in favor of the elite factions currently in power. As these states mature, this bias is slowly eradicated *within* elite relationships and in some dimensions elites emerge as one homogenous legal category. North et al. (2009, pp. 77-109) illustrate this process using English land law. The right of inheritance was established in 1100, but was based on a distinction of different types of land owners and different types of ownership. By 1600, only one category (the freeholder) and one type of ownership (free and common socage<sup>68</sup>) remained. In land law, impersonal rule of law for elites was established.

Creating organizations with perpetual life is the second step in depersonalizing society and applying law more equally. According to North et al. (2009), throughout most of history kings or emperors were above the law. As a result, any organization's existence – including the state itself – was subject to the will of an individual. The king could expropriate assets, ban and dishonor his subjects at will. In Europe, separating the individual and corporate entities of kings and popes was part of a long historical struggle throughout the 14<sup>th</sup>, 15<sup>th</sup> and 16<sup>th</sup> centuries. Once this separation was achieved, the office of the king and the *fiscus* became the first perpetual organizations.<sup>69</sup> As the state is essentially an organization of organizations, immortality of some of its parts created perpetual life of the whole. This endowed the state

---

<sup>67</sup> Interestingly, they neglect to fully treat industrialization or the rise of capitalist classes as a source of changing interests of the elite. In their framework, elites at some point realize that their rights are best protected by impersonalizing them and then by extending them as a means of securing them indefinitely. However, North et al. (2009) cannot provide a parsimonious mechanism that compares to elites facing the threat of revolution as in Acemoglu and Robinson (2006).

<sup>68</sup> Free and common socage is the highest form of tenure, which allows land to be freely bought and sold.

<sup>69</sup> Relying on Kantorowicz's (1997[1957]) notion of the king's two bodies, North et al. (2009) show how, in the history of Europe, the personal identity of the king was separated from his corporate body by forming two separate legal persons. Simultaneously, the great Schism in the 14<sup>th</sup> and 15<sup>th</sup> century led to a differentiation between the leader of the church and the corporate body administering the economic assets of the church (known as the *christus* or the *fiscus*). The *fiscus* then led credence to the commitments of the corporate office of the king, and thereby the king was put under the law (a king could no longer alienate his corporate rights, as the *fiscus* would defend the corporate identity) and two perpetual organizations had emerged.

with a greater ability to honor its commitments and created the conditions for the emergence of powerful absolutist states in 16<sup>th</sup> and 17<sup>th</sup> century Europe (North et al., 2009, p. 166).<sup>70</sup>

According to North et al. (2009, p. 169), consolidating the power of the military under political control of the state is the most difficult condition to achieve and to understand. It involves the disarmament of the population and the creation of a standing army and thus represents a fundamental shift in the nature of the natural state power balance. In fact, consolidation of the military is a dangerous process. Only if elites believe that they are able to constrain the military with non-military means, can they create an organization which otherwise can exercise unrivalled violence. Consolidation requires the previous two doorstep conditions. Only when the leader (king) cannot act at will but is constrained by a corporate body with distinct responsibilities, does it become possible to separate military organization (how to fight) from military control (when to fight). Rule of law for elites enables increasing specialization of elites and perpetual organizations support an increasing scale of specialized activities. Rather than acting as an independent driver of the formation of modern states (e.g. Tilly, 1985; Bates, 2001), a powerful consolidated military is the product of increased organizational capacity and specialization of elites.

North et al. (2009) trace how the ‘transition proper’ took place in Britain, France and the United States, who first reached the doorstep conditions and transformed into open access orders. The transition will occur when the doorstep conditions are met and the institutional incentives are still consistent with the logic of the natural state, while at the same time incentive structures are in place which motivate members of the elites to move towards opening up access. They define this process largely in historical narratives describing how even in the 18<sup>th</sup> century factions, parties and corporations were seen as a threat rather than the key to economic development. Chartering corporations was still an economic privilege originating from political power. Further, the U.S. constitution, for example, was a “Constitution against parties” (Hofstadter 1969, p.50 as cited in North et. al, 2009, p. 230) and not sufficient to institute open access. In all three countries, during the 19<sup>th</sup> century the ruling elites realized that their privileges were best protected if they converted them into impersonal rights.

Driven by changing political and economic landscapes, new interests and elites emerged which pressed to have their rights recognized and challenged the position of established privileges. The ruling coalitions of these natural states were bound to realign, but instead of a continuation of limited access, the historical innovation in these countries was to no longer exclude political losers. During the 1830s to the 1850s, Britain, France and the U.S. began to open access to the chartering of joint stock companies, created modern parties and extended the franchise to large parts of the population. In Britain, for example, the Reform Act in 1832 gave rise to modern electoral parties as an unintended side-effect of a new intra-elite bargain and the 1844 corporation act increased the number of corporations from hundreds to thousands within two decades. The specific time-lines and institutional forms in these countries differ, but all of them institutionalized competitive politics and opened the

---

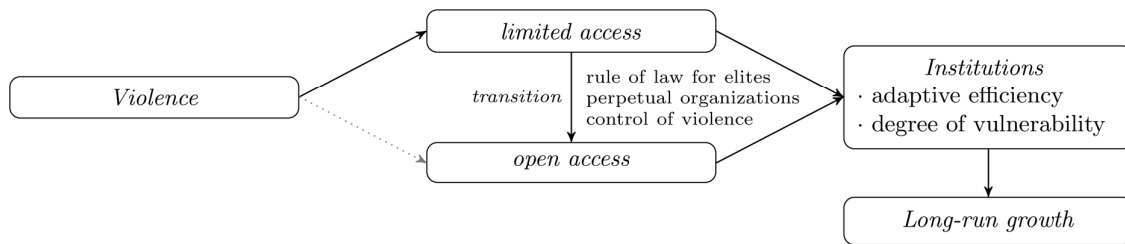
<sup>70</sup> A similar argument can be found in Elias (1969). However, Elias argues that this separation takes place only after the centralization of power in absolutist states in the 16<sup>th</sup> and 17<sup>th</sup> centuries.

chartering of organizations as separate legal entities with limited liability. Only at this point in the 1850s did growth rates of per capita income accelerate to modern rates of 1-1.5% per annum (North et al., 2009, p. 247).

### *The model*

Extracting a causal model from a framework of which the authors warn that it does not generate “explicit empirical tests or deterministic predictions about social change” (North et al., 2009, p. xii) might seem paradoxical. However, in spite of this initial warning, the core of their argument clearly suggests a *causal* relationship between the type of social order and income per capita. In our view, stylizing the relationships in their framework helps to reveal its strengths and weaknesses, even though this might involve a crude oversimplification of their reasoning.

FIGURE 10: VIOLENCE AND LONG-RUN GROWTH



Source: own illustration.

Notes: In the framework of North et al. (2009) countries first move to limited access and can later on transition to open access. However, violence influences both rent-sharing mechanisms and motivates a system where political power and economic rents are balanced even though there is a historical progression.

In its most parsimonious form the model consists only of five elements. First, all societies face the problem of coping with violence. Second, they initially begin to contain violence by limiting access to political and economic participation, forming elite coalitions and distributing rents in accordance with the elites’ violence potential. Third, some societies make a transition to contain violence through opening access, depersonalizing rights and rules and compelling everyone into political and economic competition. Fourth, both limited access and open access orders achieve a balance of political power and the distribution of economic rents, but the structure of this balance differs. Limited access orders explicitly balance power and rents by distributing rents among elites, while open access orders implicitly align political power and economic returns by creating equal opportunities to seek rents and to limit rent-seeking by others. Open access orders abolish permanent monopoly rents. The two alternative ways of achieving a balance between political power and rents determine the degree of stability of the political equilibrium and how easily it can adapt to changes arising from socio-economic dynamics or external influences without deteriorating into violence. In other words, the nature of the balance determines the adaptive efficiency and vulnerability to shocks of a society. Open access orders are more adaptively efficient and less vulnerable. Finally, the flexibility of institutions and the vulnerability to shocks together determine the average long-run growth potential. Limited access order can grow at rapid rates for short periods, but the pattern of growth is volatile and often negative. In the long run this lowers

the average growth rate. Open access orders grow more slowly, but steadily and with less severe contractions, leading to a steady long-run rise in living standards.

Historically, there was no choice between limiting or opening access at the onset. The problem of violence resulted in the emergence of natural states, which only later and under certain conditions progressed towards open access orders. Forager orders failed to effectively control violence between and within groups. Natural states addressed this deficiency by explicitly matching the distribution of rents with the distribution of power. The different types of natural states represent different degrees of professional organization of inter-elite and intra-elite relationships. Only by moving more and more towards rule of law for elites, perpetually lived organizations and consolidated violence potential (the doorstep conditions), as well as changing elite interests (the transition proper), the first modern liberal capitalist orders could emerge. However, for transitions today these pathways may be different.

The theory has many strengths and weaknesses. Reframing the problem of development in terms of control of violence reveals two key insights that have otherwise escaped most of the literature. First, ostensibly inefficient and corrupt institutions can exist for the deeper purpose of ensuring social stability and restraining violence among competing factions, but these institutions are inherently fragile and can easily descend into violence. Second, the degree of impersonalization and formal rule-based nature of this rent-sharing agreement define the organizational and adaptive capacity of society, which together are the key to modern economic growth. This gives rise to an interesting policy prescription. Development policy should be careful in trying to transplant open access institutions into limited access orders – as they would not function in the same way. Instead it should pursue incremental institutional reforms which are aimed at strengthening a country's organizational and adaptive capacity to move them closer to the doorstep conditions for a transition to an open access order.

Conversely, there are also problems with their approach. North et al. (2009) repeatedly emphasize that this is not a stage theory of development and that no teleological force pushes countries along the axis towards open access orders. Nevertheless, the theory remains reminiscent of earlier works in modernization theory (e.g. Rostow, 1960), with its simplified dichotomy between traditional and modern societies.<sup>71</sup> In different ways, these works take modern developed societies as the benchmark and then trace the component of a process of necessary social evolution from traditional societies to this benchmark. Similarly, the absence of an explicit causal mechanism for the transition makes it difficult to compare the key factors in North et al. (2009) to the theories reviewed previously. They do not provide a theory of socio-economic change which can be applied to understand why some regions forge ahead and others fall behind.

In contrast to Acemoglu and Robinson (2006), North et al. (2009) emphasize that their approach breaks with the zero-sum logic of homogeneous elites responding to challenges by non-elites through institutional reform. They argue that elites do not act as a homogeneous

---

<sup>71</sup> Thus mature natural states lump together extremely centralized social orders such as Maoist China and far less centralized orders such as Latin American states in the twentieth century where the ability of the state to impose taxes and burdens on the population was extremely limited.



group, but that they consist of competing elite factions. Further, their transition to open access follows the logic of positive-sum game in which elite factions recognize the gains from reform towards rule of law and impersonalization of privilege, leading to large personal and social benefits. This approach turns much of the collective action literature, which emphasizes coordination problems, interest-group bias and free-riding, on its head without providing systematic evidence or micro-based theory in favor of such a conclusion. The role of inequality also remains ambiguous. The distribution of political power seemingly defines the distribution of rents, but there is no distinct place for initial inequality or the consequences of colonialism which have been emphasized elsewhere. Similarly, geography is not treated as a significant independent factor. In addition, their characterization of open access orders follows a very classically liberal notion, neglecting criticisms relating to inefficiencies of modern liberal societies (Bates, 2010). We address some of these criticisms further in the following and a later subsection in which we directly compare their framework to the contribution of Khan (2010).

### *Criticism and additional evidence*

An important problem is that relying on two broad ‘social orders’ to characterize nearly all of human history obscures the unit of analysis. In feudal societies centralized power is weak. The key unit of analysis is the manor and its surroundings rather than wider society, but North et al. (2009) characterize feudal societies as a type of natural state. The governmental and social structure of modern natural states such as India or China has little in common with natural states in feudal Europe. Similarly, foraging orders have continued to exist to the present day in acephalous societies, but the unit of analysis here is the clan and the lineage rather than the state. North et al. (2009) also neglect that population size and density matter for state formation. The problem of maintaining social order in sparsely populated societies is very different from that in densely populated societies, as the size of the economic surplus is too small in sparse populations to create a centralized authority (e.g. Boserup, 1981).<sup>72</sup> In addition, all open access orders underwent a demographic transition towards significantly lower fertility rates. Demography is thus essential for the understanding of transitions and is intertwined with the emergence of different social orders. The same holds for technological advance, which North et al. (2009) only discuss in passing. Implicitly, when they refer to Schumpeterian creative destruction as an engine of innovation, they treat technological advance as endogenous to institutions. However, their theory does not explicitly address how technological advance in the past and present relates to the institutional structure and how both together affect economic development. This is especially apparent during the transition from limited access to open access for which establishing control over violence is a doorstep condition, but the role of weapons technology is not systematically analyzed.

### Military mobilization and open access

In a contribution that is also discussed in North et al. (2009), Bates (2001) argues that violence is associated with significant social costs and that the older European states emerged out of the need to raise funds domestically in a quest to mobilize for war. These states created

---

<sup>72</sup> Demographic trends play a key role in the earlier path-breaking study of North and Thomas (1973).

political alliances with elites, including the creation of parliamentary control of military ventures, in an effort to lower their borrowing costs and increase the credibility of debt repayment. In short, the liberal or open access political institutions were a by-product of the need to mobilize against external threats (Bates, 2001, p. 83). Modern developing states, however, are not subject to the same incentives due to the availability of foreign aid, international capital markets, foreign direct investment and a changed international system. Based on this premise, Bates (2010) argues that North et al. (2009) do not account for the full breath of the military mobilization thesis by Tilly (1985) and Bates (2001). North et al. (2009) themselves criticize both authors for relying on the ‘single-actor’ state model which assumes that politics and economics are distinct at the outset and fails to fully incorporate the dynamics of natural states. However, the international system and inter-state competition plays little role in their own argument. According to Bates (2010), threats from without challenged the survival of political incumbents leading them to pursue growth-promoting policies. By increasing political accountability and enforcing elections internal threats can be created that provide similar incentives. In fact, he argues that there is evidence from Africa that frequent elections reduce investors’ fears of opportunistic behavior (Humphreys and Bates, 2005) and positively influence domestic growth policy (Bates, 2008)<sup>73</sup>. By emphasizing elections and accountability over other incremental reforms aimed at intra-elite relationships, this view leads to very different policy prescriptions.

Bates (2010) agrees with North, Wallis and Weingast on the dynamics of natural states but criticizes their depiction of open access orders, which he argues relies too heavily on a combination of classic liberal economic thought (e.g. John S. Mill) and pluralist politics (e.g. Robert Dahl) without incorporating the criticisms that have been leveled against these traditions. Open access politics and open competition can result in situations where large organizations such as trade unions do not necessarily reflect the interests of their members (e.g. the *Iron Law of Oligarchy*, Michels, 1911) and interest groups with encompassing interests have difficulty organizing against distributional and special interest groups (Olson, 1971). This can result in the increasing prominence of rent-seeking interest groups in democracies which adversely affects economic growth (Olson, 1982). A similar point was made by Acemoglu and Robinson (2008) who, as discussed earlier in Section 3.3, derive a state of captured democracy in which elite interests dominate the economic sector. Further, Bates (2010) highlights, that when openness and entry is the prevailing mechanism of social organization, threatening exit is a possible strategy to wield political influence (Hirschman, 1970). With increasing global trade integration and capital mobility, exit of economic actors is only gaining in importance. However, changes in the international economic order and international institutions receive no attention in North et al. (2009). In sum, the automatic alignment of rents and power as well as the prevention of permanent rents in open access orders may work considerably less well than North et al. (2009) suggest.

---

<sup>73</sup> Although Bates (2008) also finds that increased political accountability has initially coincided with more public spending, exchange rate undervaluation and increased money supply, which he attributes to incumbents aiming to escape political defeat.

### Reinterpreting North and Thomas as an example of transition

It is interesting to note that North et al. (2009) make no reference to the early and influential work of North and Thomas (1973), which in part reignited the study of institutions. This work focused on the breakthrough of the West, contrasting the experiences of the Iberian Peninsula and North Western Europe. The key insight of these authors was that whether or not private incentives were aligned with collective welfare was decisive in the process of transition to modern economic growth.

In *The Rise of the Western World*, North and Thomas (1973) focus on the emergence of efficient institutions. Efficient institutions are defined as institutions which motivate self-interested individuals to act in ways which contribute to collective welfare. Among the efficient institutions discussed by North and Thomas are well-defined property rights, which guarantee that individuals will profit from the fruits of their own exertions. Only under such conditions will individuals be willing to make risky investments in future productive capacity. The rise of the joint-stock company allowed individuals to cooperate in raising capital and limited the risks by isolating personal wealth from the specific assets invested in the company. It limits the risk of large-scale investments for individuals and their families. Protection of intellectual property (patent rights) is one of the conditions for a continuous stream of innovations. It provides a spur to invention, innovation and technological change, by allowing the owners of intellectual property rights to charge fees for the use of their patented innovations. Land reforms that create well defined individual rights to land motivate farmers to invest in increased land productivity. Bookkeeping, financial accounting and interest were key institutional ingredients of the pursuit of sustainable profit through productive activity, with systematic reinvestment of part of the current profits. The notion of rentability contrasted with other types of search for profit such as piracy, rent-seeking or maximization of short-term speculative profits at the expense of long-run profitability. Medieval Catholicism (and traditional and modern Islam) prohibited the charging of interest on loans as an immoral activity (Lewis and Algoud, 2001). This worked as a barrier to the rise of modern capital markets, capital accumulation and a rational attitude towards investment. Protestantism had no such prohibitions.

Efficient institutions do not emerge automatically but are driven by factor price movements. The rise of efficient institutions depends on the costs and benefits involved in the creation and maintenance of such institutions for different individuals and groups. When population density increased in Europe at the end of the middle ages, this reduced transaction costs and facilitated the development of interregional trade. Production for the market and the money economy become viable alternatives to the traditionally determined exchange relationships of feudal economic systems. Previously these had been more efficient, because the *transaction costs* of market exchanges were too high. Governments were able to guarantee property rights at lower costs per person than private groups, because the costs could be distributed over larger numbers of people. Also, government intervention avoided the *free rider* problem. The development of individual property rights went hand in hand with the increasing importance of the state.

Not all governments, however, promoted more efficient institutions. Sometimes, government policy was determined by social classes, whose interests lay in the preservation of inefficient

institutions as in the Iberian Peninsula. The influence of governments on the development of institutions in its turn was influenced by the power relationships between different classes in society. In North Western Europe and in particular the Dutch republic the decentralized nature of the political system and the influence of cities and entrepreneurial classes resulted in efficient economic institutions which led to economic breakthroughs. Thus North and Thomas succeed in an ingenious fashion in combining the neoclassical economic analysis of institutional changes in terms of costs and benefits with the historical study of the power relationships between classes and interest groups. It provides a theory of transition to modern economic development which, as we have criticized, is less developed in North et al. (2009).

North and Thomas (1973) have since been criticized for placing too much emphasis on the importance of private property rights and intellectual property. For example, Nuvolari (2004) has made a case for the importance of collective invention and knowledge sharing in the development of steam engines in the eighteenth century. But what remains important about this early work is that it links institutional analysis to a specific juncture in economic history, explaining the breakthrough in North Western Europe and the relative stagnation in the Iberian Peninsula through a comparative analysis of institutional characteristics.

In a later contribution, North (1990) warns us that efficient institutions do not automatically supplant less efficient institutions and distances his theory from factor prices as the core determinant of institutions. When a society has embarked on a certain institutional path, later developments depend on choices made earlier on in the development process. Such *path dependence* is one of the explanations for the increasing divergence of richer and poor societies in the world economy. However, a careful reading of the earlier work shows that it does not argue that more efficient institutions automatically supplant less efficient institutions.<sup>74</sup>

### 3.5.2 POLITICAL SETTLEMENTS

Khan (2010) develops an alternative theory of how formal institutions and informal institutions interact in affecting economic growth and stability simultaneously. Based on this theory, he derives policy implications for growth-enhancing institutional reforms in developing countries and highlights the importance of incremental institutional changes. In the following, we briefly review his contribution and compare it to the limited/open access orders framework.

Like North et al. (2009), Khan (2010) places the distribution of power in a society at the center of the analysis of a society's political and economic structure, but Khan's definition of power is broader. He defines power as *holding power*, which refers to an organization's potential to inflict costs and violence on others, but also its ability to absorb the costs and violence inflicted upon it, as well as its capacity to mobilize members and non-members in favor of a cause. In other words, a powerful group has a heightened ability to hold out in conflict and a correspondingly high probability to win a conflict with other parties over time,

---

<sup>74</sup> As is evident in the statement that "therefore we have no guarantee that productive arrangements will emerge" (North and Thomas, 1973, p. 8). "However, Marx failed to recognize that there is nothing inevitable about growth and Smith did not tell us how to ensure an efficient government that will devise and maintain a set of property rights that will ensure economic growth" (North and Thomas, 1973, pp. 157-158).

or at least the capacity to impose serious costs on others. Such powerful groups and organizations have incentives to refrain from the actual use of violence and other means of threatening social stability if their degree of holding power is proportionally matched by the economic rents distributed to them through formal and/or informal institutions. Stated succinctly, power relationships shape institutions, institutions shape the distribution of rents, and rents buy the acquiescence of powerful elites and organizations.

In Khan (2010), the balance between economic distribution and power distribution is a *political settlement* – a concept that parallels what North et al. (2009) call the double balance of a social order. In his terms, a political settlement is “an interdependent combination of a structure of power and institutions at the level of a society that is mutually ‘compatible’ and also ‘sustainable’ in terms of economic and political viability” (Khan, 2010, p. 20). Although defined as sustainable, for Khan (2010) a settlement is never static, as economic, political and demographic dynamics constantly alter the underlying distribution of holding power. While these basic components differ only minimally from North et al. (2009), the features that describe a political settlement and the theoretical interaction of formal and informal institutions clearly distinguish his contribution from other studies.

In a criticism of traditional New Institutional Economics as pioneered by North and Thomas (1973), Williamson (1985) and North (1990), Khan (2010) argues that property rights-based theories of formal institutions provide limited insights for the analysis of developing countries. These theories argue that strengthening of private property rights creates an incentive structure conducive to economic development. Well-defined private property rights lower transaction costs, which otherwise impose a wedge between the private and social returns to economic activity. Well-enforced property rights are therefore market-augmenting. According to Khan (2010), this is only a partial and idealized reflection of reality and misses important elements that are necessary to achieve stable growth in the medium term. Most importantly, the introduction of property rights not only lowers transaction costs but also skews the distribution of economic benefits towards the formal sector. This modification of the distribution of rents gives rise to resistance from those with informal power, which in some cases may entirely offset or negate the efficiency effects arising from reform. This is likely to happen in all developing countries due to two basic facts. First, the formal sectors in all developing countries are relatively small. Second, even these small formal sectors are embedded in a complex structure of formal and informal institutions.

Strengthening formal rights is not a universal solution, not even for addressing seemingly simple market failures in land acquisition (Khan, 2009a) or for increasing technology absorption (Khan, 2009b). The feasibility of property rights institutions depends on the degree of formalization as reflected in the political settlement, and thus the distribution of formal versus informal institutions and the corresponding distribution of holding power. Khan (2010) argues that precisely in this area the developed and the developing world differ fundamentally. Developed countries have a distribution of benefits and rights based on formal sectors with significant holding power and, correspondingly, formal institutions govern the distribution of benefits. Developing countries, in contrast, are largely based on informal or non-capitalist production, which implies that formal institutions cannot support

an adequate distribution of benefits and rights alone. Plenty of informal institutions exist to 'correct' this inability of formal institutions by distributing rents to informal groups with significant holding power. In consequence, the institutional structure in developing countries is different and formal institutions perform in a different way than in developed economies. This performance difference is not necessarily inefficient as it serves the deeper purpose of providing stability through incorporating powerful groups left out of formal institutions.

Khan (2010) characterizes all developing countries as transition economies, although they are not necessarily moving towards formality. Hence, transition is a state and not a process. Through contact with more developed countries, their original pre-capitalist political settlements collapsed. Pre-capitalist societies predominantly distributed rents arising from land-based holding power through formal institutions, while in transition societies the new capitalist sectors are not able to provide a stable allocation of benefits through formal institutions. Colonialism exacerbated this problem by reallocating power to groups who did not have formal rights but had enough organizational capacity to locally administer the empire of a foreign power. This strategy was especially pursued in colonies where Europeans did not settle in large numbers or where the local population was large. In this process, the powers formed new 'intermediate classes', whose holding power is largely informal and based on patron-client networks. Khan argues that anti-colonial struggles were often supported and headed by the very same groups, which further reinforced their organizational holding power and drove an increasing wedge between the formal and informal distribution of power (Khan, 2010, p. 27).

Based on these relationships between power, stability and institutions, Khan (2010) derives a typology of four types of political settlements: pre-capitalist, capitalist, and clientelist political settlements, as well as political settlements in crisis. The *pre-capitalist settlement* is largely a historical category describing a situation in which formal rights existed but were based on feudalism and/or military power. Historically, it is the point of departure for all societies. The formal institutions in pre-capitalist societies were not growth-promoting but primarily aimed at providing stability. Technological progress was slow and economic production was based on agriculture. The rise of artisans and trading already began to upset this order in the West and in the less developed nations, European expansion and military defeat by other powers led to the total collapse of this political settlement. The advances of some countries and the resulting international order hence eradicated the remaining pre-capitalist societies. This characterization is a major departure from North et al. (2009), as their categories of *natural states* hold throughout time and do not differ after the rise of the advanced industrial nations and the beginning of European expansion. By comparison, the pre-capitalist settlement can at best be likened to basic natural states before the 18<sup>th</sup> century, or even earlier depending on when and if a country was colonized or engaged in war with a foreign power.

Next, in the *capitalist political settlement* the formal sector is large and the incomes generated by it are the dominant source of holding power. The institutional set up consists of formal institutions aligned with the formal sector's holding power, which are growth-supporting and contain redistribute conflicts. Standard institutional economics describes the implications of this settlement very well, as the efficiency of institutions is linked mainly to

transaction and enforcement costs of formal institutions. These are *potentially* Weberian states with formal property rights and impersonal rule-based law, but most importantly they are defined by a dominant formal sector. They possess some, but not all, of the characteristics of open access orders.

On the contrary, a *clientelist political settlement* is characterized by a mismatch of informal holding power and formal institutions. There are property rights and other formal productive rights, but they are not well enforced as powerful groups can influence the production process regardless of their formal rights. Khan (2010) subsumes developmental states and democratic states in this category, but also military dictatorships and states on the brink of crisis. What they all share is that personalized power dominates, patron-client networks are widespread and formal institutions do not work in the same way as in capitalist settlements because the formal productive sectors they rely on are relatively small and just emerging. It is the primary form of political settlement for countries in transition from collapsed pre-capitalist political settlements. In comparison to North et al. (2009), this category appears to encompass the whole range from basic natural states to countries in transition to open access.

There are four forms of *clientilism* depending on the power of excluded factions and the concentration of power at higher or lower level factions in the patron-client network of the ruling coalition. If excluded factions are weak and power is concentrated at the top of the patron-client network, then a *potential developmental coalition* emerges that has high enforcement capabilities and is not threatened from without. If the excluded interests are powerful but power is still centralized at the top of the patronage structure, then a (*vulnerable*) *authoritarian coalition* exists that has strong enforcement abilities but is subject to severe threats and risk of defeat from groups outside of the coalition. If the lower level factions are powerful but the excluded factions are not, the predominant clientelist variant is a (*weak*) *dominant party* which incorporates broad interests but is unable to enforce these successfully. Last, if both the lower factions and the excluded factions are powerful, society is in a state of *competitive clientilism* in which powerful groups compete and stability only arises if power can be shared and rotates among them. Evidently, the growth potential under a developmental coalition is highest and lowest in competitive clientilism, while the other two represent intermediate cases.

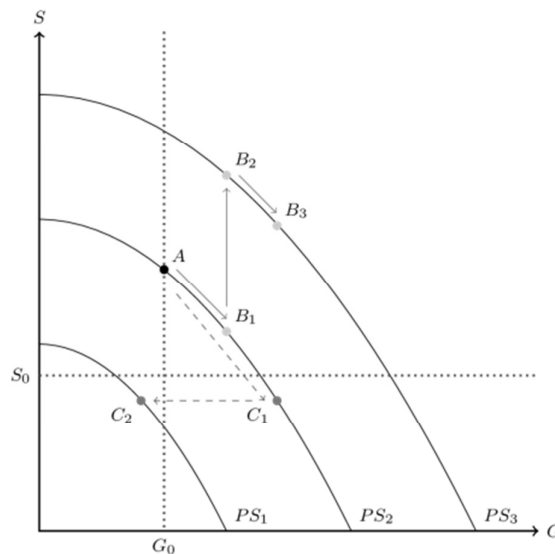
Last, Khan (2010) defines a small category of *political settlements in crisis*. It is marked by the total collapse of formal institutions. Hence, formal rights are not enforced at all. Violence specialists define the nature and scale of all political and productive activities or engage in predation, which further deteriorates the economic base. In this type of settlement, society has fallen below its minimum level of stability and/or minimum level of sustainable growth leading to widespread violence. As the formal sector collapses, most production is – by definition – based on illegal or semi-legal activities including smuggling, organized crime, drugs, diamonds, foreign funding of warlords and organized racketeering. This settlement is very similar to fragile natural states, as in North et al. (2009), or the concept of state failure often used in the political science literature.

The role of informal institutions becomes readily apparent when analyzing the effects of reforming a particular institution and taking the initial distribution of power as exogenous.

Powerful informal groups have many ways to resist the redistribution of rents arising from, for example, an attempt to strengthen private property rights. Property rights institutions can be partially enforced when powerful groups exert their informal power to appropriate benefits, ultimately leading to a reduction or eradication of the intended efficiency gains. Similarly, those formal groups that benefit from stronger property rights could aim for better enforcement by voluntarily transferring rents to informal groups with holding power. Again, this form of partial enforcement leads to lower efficiency effects although the institution seems better enforced to the outside observer. In general, when the distribution of holding power is skewed towards informal sectors, enforcement costs are higher, so that an analysis of transaction costs must include the *transition* costs arising from partial enforcement, ensuing political unrest and distributional conflict. On this basis, Khan (2010) argues that growth-enhancing institutional reforms necessarily have to be incremental, acknowledge the usefulness of ‘inefficient’ rents in providing stability and allow for institutional configurations which do not resemble the market-augmenting institutions of developed countries. For Khan (2010) such an approach to reform may in fact prove more useful in achieving successful development than a focus on market-augmenting governance structures.

Khan (2010) generalizes this argument by formulating a ‘growth-stability trade-off’. If a particular institutional reform is undertaken that is not in line with the distribution of holding power, then powerful groups will mobilize in opposition. As any reform is essentially a change in the distribution of rents, the degree of resistance is decisive in determining success reforms or adverse outcomes. Powerful groups can hinder reform through partial enforcement, re-alignment of informal rents or outright opposition and conflict. A developing country’s type of political settlement or form of clientilism defines the severity of the trade-off reformers face, but its initial location is exogenous to any specific reform attempt. Khan (2010) derives many implications for institutional reform based on this trade-off in different political settlements, which we summarize in Figure 11 below.

FIGURE 11: GROWTH-STABILITY TRADE-OFF AND INSTITUTIONAL REFORM



Source: reproduced and modified from Khan (2010, p. 43).



The essence of Figure 11 is that all societies are broadly characterized by three main parameters: a minimum level of stability ( $S_0$ ), a minimum level of growth ( $G_0$ ) and a location/slope of the growth-stability trade-off inherent in the current political settlement ( $PS_{1-3}$ ). The stability axis ( $S$ ) is an ordinal ranking of stability as perceived by the ruling coalition. The growth axis ( $G$ ) measures the achieved growth rates under each type of settlement and institutional configuration. A downward move along a political settlement curve ( $PS_{1-3}$ ) has associated transition costs arising from threats to stability ( $\Delta S$ ) and positive growth effects ( $\Delta G$ ). The trade-off inherent in any political settlement curve ( $PS_{1-3}$ ) is concave, as bigger institutional changes impose higher transition costs. The political settlement itself determines the location and slope of the curve. The more powerful the ruling coalition is, the steeper the trade-off and vice versa. The figure shows three political settlements.  $PS_2$  is the initial political settlement with a corresponding initial institutional configuration at point  $A$ .  $PS_3$  represents a more favorable political settlement and  $PS_1$  an adverse settlement, or settlement in crisis, below the minimum growth and stability threshold. In general, institutional reforms which do not risk a deterioration of the political settlement must ensure growth rates higher than the minimum growth threshold ( $G_0$ ) and above the minimum stability threshold ( $S_0$ ), otherwise society moves to an adverse settlement. Each settlement is characterized by a set of initial conditions encompassing endowments, technological abilities, formal and informal institutions and a distribution of holding power. The political settlement is endogenous to the performance of institutions – if the settlement changes the growth-stability effect of institutions shifts in the same direction.

Based on this simple illustration, Khan (2010) elaborates on the effects of different reform strategies. The first strategy is implementation through confrontation, which is synonymous to a downward move along the curve of one political settlement. Consider a move from the initial situation  $A$  to  $B_1$ , powerful groups resist and thereby reduce stability, but in the end this approach leads to a higher rate of growth. If resistance subsides, society as a whole may be able to move a better settlement ( $B_2$ ). As a second strategy we can imagine a situation in which powerful informal groups are given concessions directly from the beginning, such as a direct share in the benefits arising from reform. The trade-off curve would initially be flatter, implying that higher growth rates are readily achievable with less significant costs to stability, but then become much steeper once the concessions are withdrawn. Khan (2010) likens such a situation to industrial policy, where many groups initially benefit from subsidies but over time the focus then shifts to conditions imposed by governments and banks on the recipients of subsidies.

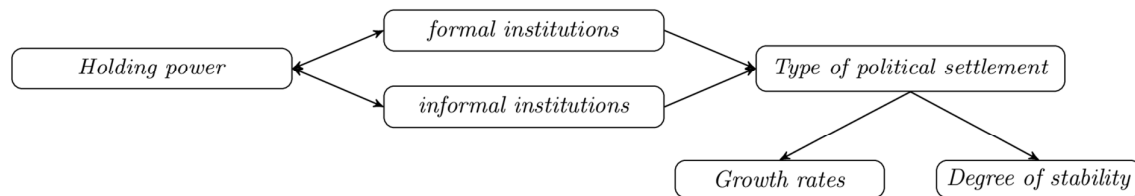
Relying on implementation through confrontation, we can follow Khan (2010) to illustrate two possible reform paths with severely different economic and political implications. The first path is an ideal type of incremental reform. A move from  $A$  to  $B_1$ , implies that with a limited loss of stability, higher growth was achieved. Based on this reform and subsiding opposition, society moves to an advantageous settlement ( $PS_3$ ) and point  $B_2$  in the subsequent period. After this shift, incremental institutional reform is continued, as represented by a move to  $B_3$ . In all respects,  $B_3$  is a significant improvement over  $A$ . However, the outcome may be very different in the case of *discontinuous* or *big-bang reforms*. Consider an outright move from point  $A$  to  $C_1$ , the growth-equivalent of  $B_3$ . The

costs of reform are high enough to push society below the minimum stability threshold ( $S_0$ ) necessary to sustain the current political settlement ( $PS_2$ ). The outcome is a collapse of the current settlement and a decline into a political settlement in crisis ( $PS_1$ ), whereby shifting institutional performance from  $C_1$  to  $C_2$  in the next period. Point  $C_2$  on this settlement lies below both the minimum growth and stability threshold. Therefore, the political-economic order disintegrates and can potentially move even further towards the origin until a new form of stable clientelist settlement emerges.<sup>75</sup>

### *The model*

Khan (2010) himself graphically illustrates the causal links of his framework. We modified and combined his illustrations in order to show all the major relationships at once (see Figure 12 below). As opposed to only concentrating on violence potential, Khan (2010) begins his broader notion of the distribution of holding power. Holding power drives the creation of formal institutions, such as property rights, and informal institutions, such as patron-client networks. The ratio of formal institutions to informal institutions depends on whether holding power is primarily located in the formal productive sectors or informal non-capitalist sectors. Both types of institutions and the distribution of holding power are in simultaneous causal relationship, where power shapes institutions which create and maintain a certain distribution of rents, while these rents in turn shape and sustain the distribution of holding power. For this reason, any attempt at institutional reform immediately affects the underlying distribution of rents and power in a society.

FIGURE 12: HOLDING POWER, GROWTH AND STABILITY



Source: own illustration.

Institutions and the distribution of power interact until a sustainable allocation of rents and privileges is achieved. Sustainability requires the distribution of rents to be roughly in line with the distribution of power and also that this distributive arrangement is economically viable. The result of this matching process is what Khan (2010) calls a political settlement. The type of institutional structure inherent in a political settlement has two major implications. First, it has an efficiency effect arising from creating a set of transaction costs through formal institutions and associated enforcement/contestation costs through formal and informal institutions which together determine growth outcomes. Second, it has a distributive

<sup>75</sup> Many more pathways are possible and are discussed in more detail in Khan (2010, p. 36-48), as well as in case studies of Thailand, Maharashtra (India), Bangladesh and Tanzania. Khan (2010) also elaborates on the structure of patron-client networks to illustrate the dynamic effects between the distributions of power of lower level versus excluded factions and analyses the implications of differences in holding power of investors versus differences in technological-entrepreneurial capabilities. The resulting framework is more diversified than depicted here and concentrates more on the structures of ruling coalitions and parties than North et al. (2009).

effect arising from the formal and informal distribution of benefits and rents. This effect determines stability in a society. Khan (2010) defines a minimum threshold for both growth and stability. If either threshold is undercut by institutional reforms or other shocks to the political settlement, the prevailing arrangement collapses and the matching process begins anew.

It is important to note, that his framework is more a model of how to promote growth in the medium run than a model of growth in the very long run comparable to those reviewed previously. Khan (2010) does not attempt to explain the rise of the formal sector in the advanced economies. Instead, he concentrates on the effects of contemporary institutions on growth and stability. While using comparable elements, his framework differs in many ways from how we depicted North et al. (2009). First, he relies on a broader definition of power but that difference is more qualitative than causal. Second, power explicitly creates formal and informal institutions, which in turn sustain the distribution of power. While this process is comparable to limiting or opening access, it works differently. Informal institutions only cease to play a significant role once the mainstay of holding power originates from the formal sector. Third, the framework explicitly accommodates the rise of colonialism and allows for pre-existing degrees of initial at the point of collapse of the pre-capitalist settlements. Such circumstances represent external modifications to the distribution of holding power, or in the case of historical inequality represent the initial shape of the power distribution. However, characterizing the allocation of rents as a function of power clearly subsumes the dynamics of economic inequality to the dynamics of political power.

#### *Comparing social orders and political settlements*

We do not criticize the work of Khan (2010) directly but instead focus on summarizing the major differences and commonalities between Khan (2010) and North et al. (2009). In many ways, these authors argue along similar lines with comparable terminology. Both contributions emphasize that the institutional arrangements of developing countries are based on power and serve to provide stability. Further, both highlight the positive aspects of redistributing rents to achieve an elite bargain and, on this basis, argue for incremental reforms consistent with the prevailing social arrangement. However, Khan (2010) goes further and identifies other types of productive rents, such as rents used to increase technological capabilities and create productive assets (Khan 2000a, 2000b). He describes developing countries as facing a problem of rent management and rent governance, where the management of informal institutions poses the most important and challenging task to reformers. The treatment of informal institutions is more implicit in North et al. (2009), who do not directly refer to informal institutions as correcting the inadequate distribution of rents. However, their general approach and views on corruption clearly illustrate that they matter.

The biggest differences relate to the taxonomies these authors derive and which elements they associate with certain settlements or social orders. The capitalist political settlement in Khan (2010) shares many of the features of open access orders, such as elements of a Weberian state and impersonal rule of law. Nevertheless, he emphasizes different factors, such as the vertical structure of power in patron-client networks and the power of excluded factions that lead to the different types of political settlements. Most crucially, for becoming

an open access order, or achieving a capitalist settlement, formal production matters. In his framework a capitalist settlement cannot appear without a sufficient base of economic activity originating from the formal productive sector. This stands in direct contrast with the doorstep conditions identified by North et al. (2009), which for Khan (2010) are consequences of a capitalist settlement rather than independent consequences. In doing so, Khan (2010) allows for more of a reverse causal effect running from production over power to institutions coexisting with the causal direction from institutions to production. He also distinguishes between pre-capitalist societies and new contemporary forms of political settlements, while the categories of natural states in North et al. (2009) are historical constants since the Neolithic revolution. Last, Khan (2010) embeds his framework much more in the developing world since the 1950s, aiming at a direct analysis of contemporary reform policy more than historical development. This also entails that Khan (2010) only crudely sketches how the pre-capitalist societies collapsed, or how capitalist/open access societies function and even barely touches on how societies can make a successful transition to a capitalist political settlement.

The key lesson from both contributions is that developing countries today are faced with the task of maintaining social stability and promoting economic growth at the same time. Interactions between these two factors ultimately determine success in development. If reformers err on the side of growth-enhancing reforms, social stability will decline as splintering factions compete over resources. Conversely, if they err on the side of stability, the economy may grow at rates far away from resembling catch-up growth. Reforms need to be consistent with the logic of natural states or growth-stability trade-off inherent in the prevailing political settlement. The strength of Khan (2010) lies in his extensive treatment of the nature and role of informal institutions, but he considers growth and stability as two separate outcomes of a political settlement. North et al. (2009) approach the issue differently and argue that adaptive efficiency and vulnerability to shock are directly derived from the type of rent-distribution structure a society adopts. Both provide a political economy perspective of reforms which is lacking in the growth diagnostics framework of Hausmann et al. (2005b). The binding constraints framework focuses on identifying the reforms with the strongest economic impacts, but neglects to tackle the political economy of reform.<sup>76</sup>

Especially the approach of North et al. (2009) allows us to reconceive long-run economic development as a sequence of growth episodes. At each stage the adaptive efficiency of an institutional structure defines vulnerability of the political order and the degree of vulnerability directly affects growth volatility. The fundamental problem is not that developing countries do not grow or cannot experience substantial growth accelerations (Hausmann et al., 2005a). In fact, they often grow more rapidly than their developed counterparts. Development is a problem of growth volatility, the frequency of negative growth periods and the severity of growth collapses (Rodrik, 1999). In this respect, North et al. (2009) and Khan (2010) provide new and challenging ways of thinking about economic growth. The next steps of this type of institutional analysis should be to empirically

---

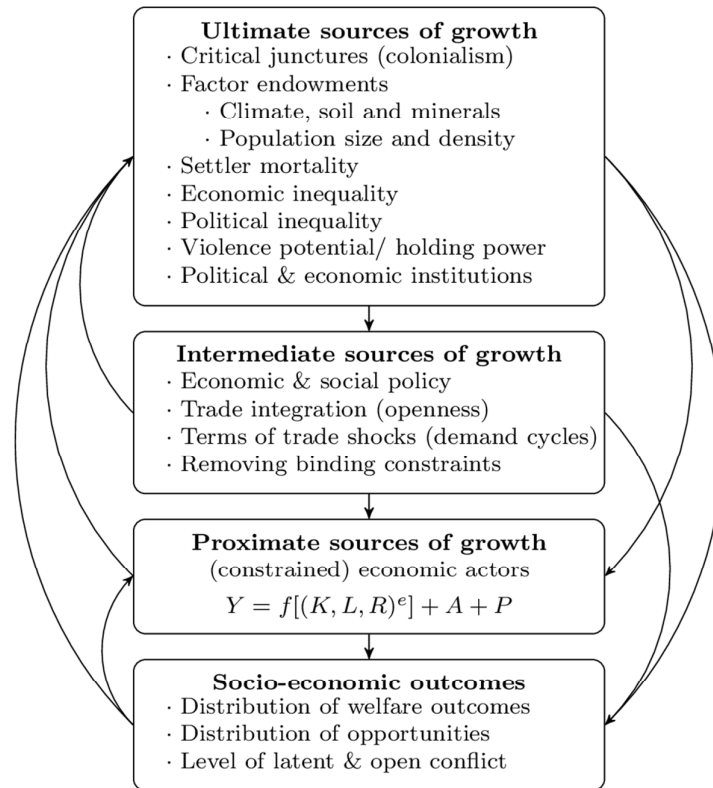
<sup>76</sup> At least in its theoretical exposition, the many case studies using growth diagnostics often rely on ad-hoc political economy explanations with little guidance from the decision tree.

operationalize and test these frameworks, and to incorporate missing links, such as population dynamics and technology, more formally into the theory.

#### 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 examined theories and variables. Figure 13 below presents a modified version of the framework including only the factors examined in this review of the literature.

FIGURE 13: MODIFIED SOURCES-OF-GROWTH FRAMEWORK



Source: own illustration.

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.<sup>77</sup> At best they have been partially incorporated into the incentive structures provided by institutions.

<sup>77</sup> 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]

#### 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 population.<sup>78</sup> 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 less 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.

---

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). Other exceptions are Harrison (1985) and Harrison and Huntington (2000).

<sup>78</sup> 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.

(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 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.

The analysis of Rodrik et al. (2004) does not offer a unique theory of long-run growth of its own, but 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 argument 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.<sup>79</sup> They provide substantial credibility to the link between institutional characteristics and later industrialization and the emergence of modern economic growth. However, it is 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).

North et al. (2009) and Khan (2010) reintroduce violence and political power as an *ultimate* source of growth. The fact that power is distributed unequally within societies creates political instability and a constant threat of violence. Both theories argue that violence and conflict are contained by balancing the distribution of economic rents with the distribution of power. They conceptualize differences in the institutional structure resulting from different variants of these rent-sharing agreements as natural states/ open access orders or as political settlements. In short, the distribution of power shapes political and economic institutions. These approaches are much broader than studies focusing on ex-colonies, but can incorporate ‘critical junctures’ as externally imposed realignments of the distribution of violence potential and holding power. Nevertheless, they are not well integrated with other factors stressed in the colonial institutions literature, such as factor endowments or geography.

In North et al. (2009) elites create rents by limiting access to economic opportunities in order to control factional violence. In consequence, they institutionalize political and economic inequality in line with their violence potential. A transition to a new indirect type of rent allocation becomes possible only after the doorstep conditions of rule of law for all elites, perpetually-lived organizations and consolidated control of the military are fulfilled, and elites realize the gains from turning elite privileges into impersonal rights. The result of such a transition is an institutional structure of open access to political and economic organization, which creates equal opportunities to seek rents and to limit rent-seeking of others. Hence,

---

<sup>79</sup> The theories seem to suggest an almost linear relationship between institutionalized inequality and long run growth performance. This is somewhat 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, inequalities in incomes and wealth have been increasing sharply in the most advanced economies. The relationship between institutionalized inequality and long-run growth performance only holds in the very long run for cross-country comparisons. 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.



rents are distributed according to the ability of organizations to create and maintain a competitive edge against existing and newly forming interests. The explicit nature of the rent-sharing agreement in natural states (developing countries) makes them less adaptively efficient and more vulnerable to shocks. On the contrary, realignments of political and economic power are internalized in the institutions of highly developed open access orders. In Khan (2010) this process is different. Like violence potential, holding power drives the creation of formal and informal institutions. However, whether the political settlement is conducive to productive activities and can maintain social stability at the same time depends on the shape of the underlying distribution of holding power. In this theory, the key difference between developing and developed economies is that in the former a large share of holding power originates from the informal sector, while in the latter the formal sector is more powerful and can accommodate a rent distribution relying exclusively on formal institutions.

## 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 in part 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 inequalitarian 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 in general. 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 but 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 only weakly affects growth independently (Rodrik et al., 2004; Rigobon and Rodrik, 2005).<sup>80</sup> 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

---

<sup>80</sup> We rely on Rodrik et al. (2004) to have effectively overturned the initial results of Frankel and Romer (1999) indicating a large and independent effect of trade integration.

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, 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.

North et al. (2009) and Khan (2010) add a political economy perspective to the analysis of growth-promoting institutional reforms. Institutions affect efficiency (transaction costs) and the distribution of economic rents simultaneously. Their theories challenge the paradigm that corruption is unequivocally inefficient by arguing that limiting access and distributing rents to powerful groups is essential for ensuring social stability. In developing countries, too disruptive reform attempts modeled on developed country institutions can lead to the collapse of the current institutional structure or will be ‘corrected’ to reflect the underlying distribution of power through corruption and other informal mechanisms. On this basis, these authors argue that economic policy targeting institutional reforms should be incremental and not too far removed from the prevailing institutional structures.

#### 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 formerly economically successful 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.<sup>81</sup> In

---

<sup>81</sup> 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

addition, inegalitarian colonial institutions were associated with continued colonial plunder of economic surpluses, which 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 long-run outcomes in initially relatively homogeneous societies, while inequality remain unchanged or became 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.<sup>82</sup> 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.

We can also interpret the concept of ‘binding constraints’ not only as a reform strategy but as constraints placed on the economic actors in general, 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 innovation, 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. Or they can express adverse socioeconomic outcomes, such as low human capital which 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. In other words, they resemble a mixture of ultimate, intermediate, and proximate sources.

#### 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

---

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).

<sup>82</sup> 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.

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 a country's population. 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 in comparison to Engerman and Sokoloff (1997, 2002) they place less of an emphasis on the interaction between ultimate causes and social outcomes.

Rodrik (1999) further connects socio-economic outcomes to growth performance in the short and medium term. His theory and evidence shows how trade shocks interact with latent social conflict and harm growth especially in countries with underdeveloped conflict management institutions.<sup>83</sup> 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).

In an approach which we see as complementary to Rodrik's work, North et al. (2009) and Khan (2010) argue that the level of latent or open conflict is a manifestation of a mismatch between the current distribution of power and current rent distribution. They characterize conflict as an expression of internally or externally triggered realignments of power to which institutions have difficulty to adjust. Hence, ultimate sources of growth interact with broader social and economic dynamics in having a direct impact on socio-economic outcomes. In the worst cases, this leads to growth collapses and conflicts over resources as in fragile natural states or political settlements in crisis. Khan (2010) illustrates this in the form of a growth-stability trade-off, where a certain minimum growth rate and minimum level of stability must be maintained to not trigger distributive conflict. If either of these thresholds is surpassed, then the current institutional arrangement collapses. North et al. (2009) subsume how

---

<sup>83</sup> 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).

societies handle this dynamically reinforcing relationship under the concept of adaptive efficiency. Long-run growth in many developing countries is low, because their institutions are less adaptive to change, resulting in a volatile and often negative short-run growth patterns. On the contrary, developed countries resolve distributive conflicts peacefully within formal institutions, allowing them to grow along a moderate and steady growth path.

## 5. CONCLUSION

In the introduction to this paper we identified three themes: (1) Are differences in institutional arrangements a fundamental determinant of differences in long-run economic growth since 1500? (2) Does political and economic inequality affect growth by influencing the nature of institutions? (3) To what extent do past institutions determine current institutions?

### 5.1 *Institutions and Growth*

In the surveyed literature, we find strong support for the proposition that institutional arrangements are indeed among the ultimate sources of long-run growth since 1500. Several studies explicitly argue that more extractive or exploitative colonial institutions resulted in slower long-run growth and more inclusive institutions in more rapid growth. Their findings indicate that institutional differences explain the *largest* part of cross-country differences in GDP per capita. A diverse body of political economy contributions focusing on the degree of access of large segments of the population to economic, political and social organizations is in line with these results. With regards to the kind of institutional arrangements that are conducive to growth, some studies focus specifically only on property rights and the rule of law, while others emphasize the overall degree of inclusiveness of institutions.

Whether institutions are the *most* fundamental source of economic growth or to what extent they are themselves endogenously determined remains highly debated. The key studies surveyed in this paper have been criticized both with regard to their methods and their underlying theoretical frameworks. For example, Glaeser et al. (2004) have shown that settler mortality – the innovative identification device for the causal effect of institutions in the ‘critical junctures’ literature – is negatively correlated with human capital accumulation and thus not a reliable instrumental variable. Further, some authors such as Przeworski (2004a,b) reject this search for principal causality in institutional theory and econometric analysis altogether on the grounds that institutions and development are mutually endogenous.

At the ultimate level of causality there is an interesting debate about the respective roles of geography and institutional characteristics. In the ‘critical junctures’ literature geography only plays a role via the local disease environment, which determined European settlements and subsequently the kind of institutions that emerged in former colonies. Another strand of the long-run growth literature gives more weight to both the direct and indirect effects of geography on growth (Diamond, 1997; Bloom and Sachs, 1998; McArthur and Sachs, 2001; Sachs, 2001). Rodrik et al. (2004) show that geography only affects long-run growth indirectly through institutions, but these indirect effects are very large. Engerman and

Sokoloff (1997, 2002) place an even greater emphasis on geographical factors or ‘factor endowments’ as historical determinants of institutions in the New World.

Most of the literature included in this paper neglects to discuss the role of knowledge accumulation and technology. To the extent that these receive any attention, they are subsumed under the incentives provided by institutions. Moreover, much of the modern literature on institutions is detached from the rich classical literature on the dynamics of class relationships (e.g. Moore, 1966) and their interactions with institutional change and technological advance. In the reviewed works, the existence of classes is either implied, or modeled at the macro-level as a dualistic antagonism between elites and the rest of society.

### *5.2 The role of political and economic inequality*

The effects of economic and political inequality on long-run growth performance are receiving increasing attention in the recent literature. However, it is premature to propose firm conclusions at this stage, as their respective effects are only beginning to be theoretically and empirically distinguished.

Engerman and Sokoloff (1997, 2002) argue that inequality of economic outcomes is determined by factor proportions. They regard economic inequality in Latin America as a major cause of adverse institutions and present evidence supporting this argument. Easterly (2007) tests this theory and finds large effects of structural inequality on current institutions, the accumulation of human capital and present levels of GDP per capita. Nunn (2008a) has presented evidence that runs counter to this hypothesis. In addition, Easterly’s (2007) application of a theory restricted to New World colonies to the whole world is questionable. Finally, Acemoglu and coauthors have argued that political inequality as captured by inclusive institutions rather than economic inequality is the key determinant of long-run growth. However, economic inequality features centrally in their theories of democratization.

Political inequality also plays an important role in the work of North, Wallis and Weingast and Khan. Focusing on the control of violence in human societies, North et al. (2009) argue that for much of world history social stability arises from compromises between elites who decide to limit access to political and economic organizations and distribute the resulting rents amongst themselves in so-called *natural states*. Hence, inequality provides social stability in natural states, but also limits their development potential. Only the transition to more inclusive and egalitarian social orders (*open access orders*) and the transformation of elite privileges into impersonal rights creates the conditions for sustained long-run growth and development. Nevertheless, how such transitions occur still remains rather elusive.

Khan (2010) also emphasizes the distribution of rents among elites and other powerful groups as an instrument to bring about social stability. The power of elites depends on their holding power and stability hinges on whether holding power and economic returns are sufficiently aligned. In post-colonial societies where formal capitalist and informal sectors coexist, stable political settlements have to take account of the informal sector. There is a trade-off between growth-enhancing institutional reforms and declining social stability as powerful groups

resist reform attempts that alter their share of rents. Disregard of the informal sector is one of the key reasons why institutional reforms so often fall short of expectations.

### *5.3 Path dependence*

The third theme of this review is the issue of time-persistent institutions. The reviewed literature generally argues that institutions are strongly path-dependent and to some extent persist over centuries. All the empirical studies discussed in this paper generally conclude that early institutions still affect institutions and development outcomes today.

This raises the important question of the very feasibility of institutional reforms. The more institutions are path-dependent, the less scope remains for radical institutional reforms. Nevertheless, the surveyed literature also indicates that institutions can change, both incrementally and under the influence of large external shocks (e.g. especially colonization, but also war, revolution and economic crisis). Hence, an institution-centered view on development highlights the importance of historical path dependence of institutions on the one hand, but also points to potentially large long-run payoffs in cases of successful institutional transformations on the other hand.

Unfortunately, most of the literature does not attempt or fully succeed to bridge the gap between the finding that ‘institutions matter’ and policy relevant questions addressing institutional reforms in the short run. Several contributions in this review argue in favor of incremental institutional reforms that are consistent with the nature of the prevailing power structure, while simultaneously providing growth-promoting incentives. However, specific policies targeted at strengthening institutions while facilitating growth cannot easily be derived from this literature. The analysis of growth accelerations even leads to the rather unexpected finding that neither wholesale institutional reforms nor other standard variables are a precondition for igniting short-run growth. Consequently, we must conclude that few unambiguous policy implications towards institutional reform can be found in this literature.

To achieve lasting and stable long-run growth, societies require more than temporary short-run growth accelerations. However, why growth accelerations occur frequently but are at the same time hard to transform into sustained growth remains largely unexplained. This is a promising area for future research. The literature surveyed in this review provides some pointers. It suggests that adverse institutions make a country more vulnerable to growth collapses, less able to sustain growth spurts and less flexible and adaptable in the face of external shocks or internal challenges. North et al. (2009) and Khan (2010) provide first attempts to relate the characteristics of political institutions to short-run volatility and long-run growth paths. Conceptualizing long-run development as a sequence of growth episodes – sluggish growth, volatile growth, growth accelerations and growth collapses – which are in varying ways influenced by institutional structures is a promising research agenda.

## 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." *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 reexamination 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. 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. "Reevaluating the Modernization Hypothesis." *unpublished*, Harvard University.
- Acemoglu, D., S. Johnson, J.A. Robinson, and P. Yared. 2008. "Income and democracy." *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." *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?" *Quarterly Journal of Economics*, 106 (4): 979–1014.



- Aron, J. 2000. "Growth and Institutions: A Review of the Evidence." *The World Bank Research Observer*, 15 (1): 99–135.
- Austin, G. 2008. "The 'reversal of fortune' thesis and the compression of history: perspectives from African and comparative economic history." *Journal of international development*, 20 (8): 996–1027.
- Balassa, B. 1982. *Development strategies in semi-industrial economies*, Johns Hopkins University Press.
- Barro, R.J. 1999. "Determinants of democracy." *Journal of Political Economy*, 107 (6): 158–183.
- Bates, R. 1981. "States and markets in tropical Africa: The political basis of agricultural policy." Berkeley: University of California Press.
- Bates, R. 2010. "A Review of Douglass C. North, John Joseph Wallis, and Barry R. Weingast's *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*." *Journal of Economic Literature*, 48 (3): 752–756.
- Bates, R.H. 2001. *Prosperity and violence*, WW Norton & Co.
- Bates, R.H. 2008. "Political Reform." in B.J. Ndulu, S.A. O'Connell, R.H. Bates, P. Collier, and C.C. Soludo, eds., *The political economy of economic growth in Africa, 1960-2000*, Cambridge University Press, pp. 348–390.
- 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.
- Boserup, E. 1981. *Population and technical change*, University of Chicago Press.
- Chang, H.J. 2002. *Kicking away the ladder: development strategy in historical perspective*, Anthem Press.
- Clark, G. 2007. *Farewell to Alms: A Short Economic History of the World*, Princeton: Princeton University Press.
- Coates, D., J.C. Heckelman, and B. Wilson. 2007. "Determinants of interest group formation." *Public Choice*, 133 (3): 377–391.
- 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." *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.
- Denison, E.F. 1967. *Why growth rates differ: postwar experience in nine western countries*, Washington, D.C.: The Brookings Institution. Assisted by J.P. Poulhier.
- Diamond, J. 1997. *Guns, germs and steel: the fates of human societies*, W. W. Norton & Company.
- Dixit, A. K. 2004. *Lawlessness and economics: alternative modes of governance*, Princeton University Press.
- Dixit, A. K. 2007. "Evaluating recipes for development success." *The World Bank Research Observer*, 22 (2): 131–157.
- 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.
- Dosi, G., M. Cimoli, and J.E. Stiglitz. 2009. *Industrial Policy and Development: The Political Economy of Capabilities Accumulation*, Oxford University Press.
- 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." *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.
- Easterly, W. and R. Levine. 2003. "Tropics, germs, and crops: how endowments influence economic development." *Journal of monetary economics*, 50 (1): 3–39.
- Edwards, S. 1989. "Real exchange rates, devaluation, and adjustment: exchange rate policy in developing countries."
- Elias, N. 1969[1939]. *Über den Prozeß der Zivilisation*, 2nd ed., Vol. I & II, Bern & München: Francke Verlag.
- 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." *Journal of Economic History*, 65 (04): 891–921.
- Fieldhouse, D.K. 1982. *The colonial empires: a comparative survey from the eighteenth century*, Macmillan.
- Frankel, J.A. and D. Romer. 1999. "Does trade cause growth?" *American Economic Review*, 89 (3): 379–399.
- 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.
- Greenwald, B.C. and J.E. Stiglitz. 1986. "Externalities in economies with imperfect information and incomplete markets." *Quarterly Journal of Economics*, 101 (2): 229.
- Greif, A. 2006. *Institutions and the path to the modern economy: Lessons from medieval trade*, Cambridge University Press.
- 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. "La Mortalite des Eveques Latino-Americains aux XVIIe et XVIII Siecles." in "Annales de Demographie Historique" 1986. pp. 29–39.
- Hall, R.E. and C.I. Jones. 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *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." 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." *Quarterly Journal of Economics*, 123 (2): 441–487.
- Heston, A., R. Summers, and B. Aten. 2009. "Penn world table version 6.3." Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.
- Hirschman, A.O. 1958. *The strategy of economic development*, New Haven, CT: Yale University Press.
- Hirschman, A.O. 1970. *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*, Harvard 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.
- Hofstadter, R. 1969. *The idea of a party system*, University of California Press.
- Humphreys, M. and R. Bates. 2005. "Political institutions and economic policies: Lessons from Africa." *British Journal of Political Science*, 35 (03): 403–428.
- Johnson, S., W. Larson, C. Papageorgiou, and A. Subramanian. "Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates." Working Paper 15455, National Bureau of Economic Research 2009.
- Jong-A-Pin, R. and J. De Haan. 2011. "Political regime change, economic liberalization and growth accelerations." *Public Choice*, 146, 93–115.
- Kant, I. 1903[1795]. *Perpetual peace: a philosophical essay*, Allen & Unwin.
- Kantorowicz, E.H. 1997[1957]. *The Kings Two Bodies: A Study in Mediaeval Political Theology*, Princeton: Princeton University Press.
- Kapp, K.W. 1950. *The social costs of private enterprise*, Harvard University Press.
- Kaufman, D., A. Kraay, and P. Zoido-Lobaton. 1999. "Governance matters." *World Bank Policy Research Working Paper*, 2196.
- Kaufmann, D., A. Kraay, and P. Zoido-Lobaton. 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.
- Khan, M.H. 2009. "Governance Capabilities and the Property Rights Transition in Developing Countries." *Project Report*. DFID.
- Khan, M.H. 2009. "Learning, Technology Acquisition and Governance Challenges in Developing Countries." *Project Report*. DFID.
- Khan, M.H. 2000. "Rent-seeking as Process." in M.H. Khan and K.S. Jomo, eds., *Rents, rent-seeking and economic development: theory and evidence in Asia*, Cambridge: Cambridge University Press, pp. 70–144.
- Khan, M.H. 2000. "Rents, Efficiency and Growth." in M.H. Khan and K.S. Jomo, eds., *Rents, rent-seeking and economic development: theory and evidence in Asia*, Cambridge: Cambridge University Press, pp. 21–69.
- Khan, M.H. 2010. "Political Settlements and the Governance of Growth-Enhancing Institutions." *Unpublished Draft Paper*.
- Knack, S. and P. Keefer. 1995. "Institutions and economic performance: cross-country tests using alternative institutional measures." *Economics & Politics*, 7 (3): 207–227.
- Krueger, A.O. 1974. "The political economy of the rent-seeking society." *American Economic Review*, 64 (3): 291–303.
- Krueger, A.O. 1997. "Trade policy and economic development: how we learn." *American Economic Review*, 87 (1): 1–22.
- Krueger, A.O. 2002. *Economic policy reform: the second stage*, University of Chicago Press.
- Krugman, P. 1994. "The myth of Asia's miracle." *Foreign affairs*, 73 (6): 62–78.
- Krugman, P. 2008. "Inequality and Redistribution." in N. Serra and J.E. Stiglitz, eds., *The Washington Consensus Reconsidered*, Oxford University Press, USA, chapter 4, pp. 31–40.

- Kuczynski, P.P. and J. Williamson. 2003. *After the Washington consensus: Restarting growth and reform in Latin America*, Washington, D.C.: Institute for International Economics.
- Kuznets, S. 1955. "Economic growth and income inequality." *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, M.K. and L.M. Algoud. 2001. *Islamic Banking*, Edward Elgar: Cheltenham.
- 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." *American Political Science Review*, 53 (1): 69–105.
- Lucas, R.E. et al. 1988. "On the mechanics of economic development." *Journal of Monetary Economics*, 22 (1): 3–42.
- 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.
- Maddison, A. 2010. "Statistics on world population, GDP and per capita GDP, 1-2008 AD." Groningen Growth and Development Centre, Groningen University.
- 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.
- Marshall, M.G. and K. Jaggers. 2009. "Polity IV Dataset (Computer file; version p4v2009)." Center for International Development and Conflict Management, College Park (MD), University of Maryland.
- Marshall, M.G., K. Jaggers, and T.R. Gurr. 2010. "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2008." Integrated Network for Societal Conflict Research Program. University of Maryland, Center for International Development and Conflict Management, College Park, MD.
- Mauro, P. 1995. "Corruption and growth." *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*.
- Meisel, N. and J. Ould Aoudia. "Is 'Good Governance' a Good Development Strategy?" Working Paper 58, Agence Française de Développement 2008
- Menshikov, S. 1993. "The 'socialist experiment' and transformation towards the market." in A. Szirmai, B. Van Ark, and D. Pilat, eds., *Explaining economic growth: essays in honour of Angus Maddison*, North Holland.
- Michels, R. 1962[1911]. *Political Parties*, New York: Collier.
- Mitchell, W.C. 1910. "The rationality of economic activity." *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.
- Moore, B. 1966. *Social origins of democracy and dictatorship*, Boston: Beacon Press.
- Murphy, K.S., A. Shleifer, and R.W. Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy*, 97 (5): 1003–26.
- Murrell, P. 1995. "The Transition According to Cambridge, Mass." *Journal of Economic Literature*, 33 (1): 164–178.

- Myrdal, G. 1968. *Asian Drama: An Inquiry into the Poverty of Nation*, Vol. 1-3, New York: Pantheon.
- Naim, M. 2000. "Washington Consensus or Washington Confusion?" *Foreign Policy*, (118): 87–103.
- North, D.C. 1990. *Institutions, institutional change, and economic performance*, Cambridge: Cambridge University Press.
- North, D.C. 2005. *Understanding the process of economic change*, Princeton University Press.
- North, D.C. and R.P. Thomas. 1973. *The Rise of the Western World, a New Economic History*, Cambridge University Press.
- North, D.C., J. Wallis, and B. Weingast. 2009. *Violence and Social Orders A Conceptual Framework for Interpreting Recorded Human History*, Cambridge University Press.
- North, D.C., J.J. Wallis, S.B. Webb, and B.R. Weingast. 2007. "Limited access orders in the developing world: A new approach to the problems of development." *World Bank Policy Research Working Paper Series*, 4359.
- Nunn, N. 2008. "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. 2008. "The Long-Term Effects of Africa's Slave Trades." *Quarterly Journal of Economics*, 123 (1): 139–176.
- Nuvolari, A. 2004. "Collective invention during the British Industrial Revolution: the case of the Cornish pumping engine." *Cambridge Journal of Economics*, 28 (3): 347.
- Olson, M. 1971[1965]. *Logic of Collective Action: Public Goods and the Theory of Groups*, Harvard University Press.
- Olson, M. 1982. *The rise and decline of nations*, New Haven, CT: Yale University Press.
- Ortiz, G. 2003. "Overcoming Reform Fatigue." *Finance & Development*, 40 (3): 14.
- Persson, T. and G.E. Tabellini. 2005. *The economic effects of constitutions*, The MIT Press.
- Pomeranz, K. 2001. *The great divergence: China, Europe, and the making of the modern world economy*, Princeton University Press.
- Pritchett, L. 1998. "Patterns of Economic Growth: Hills, Plateaus, Mountains, and Plains." *World Bank Policy Research Working Paper No. 1947*.
- Przeworski, A. 2004a. "Institutions Matter?" *Government and Opposition*, 39 (4): 527–540.
- Przeworski, A. 2004b. "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. 1994. "King Kong Meets Godzilla: The World Bank and The East Asian Miracle." in A. Fishlow, S. Haggard, C. Gwin, and D. Rodrik, eds., *Miracle or design? Lessons from the East Asian experience*, Overseas Development Council, chapter 1, pp. 13–54.
- Rodrik, D. 1996. "Understanding economic policy reform." *Journal of Economic Literature*, 34 (1): 9–41.
- Rodrik, D. 1998. "Where Did All The Growth Go? External Shocks, Social Conflict, and Growth Collapses." *NBER Working Paper 6350*.
- Rodrik, D. 1998. "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. 2006. "Goodbye Washington consensus, hello Washington confusion." *Journal of Economic Literature*, 44 (4): 973–987.
- Rodrik, D. 2008. "Second-best institutions." *American Economic Review*, 98 (2): 100–104.
- Rodrik, D. 2008. "The real exchange rate and economic growth." *Brookings Papers on Economic Activity*, 2008 (2): 365–412.
- Rodrik, D. 2010. "Diagnostics before prescription." *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.
- Romer, P.M. 1986. "Increasing returns and long-run growth." *Journal of Political Economy*, 94 (5): 1002–1037.
- Rosenstein-Rodan, P.N. 1943. "Problems of industrialisation of eastern and south-eastern Europe." *Economic Journal*, 53 (210): 202–211.
- Rostow, W.W. 1960. "The stages of economic growth." Cambridge University Press.
- Sachs, J.D. 1985. "External debt and macroeconomic performance in Latin America and East Asia." *Brookings Papers on Economic Activity*, 1985 (2): 523–573.
- Sachs, J.D. 1991. "Crossing the valley of tears in East European reform." *Challenge*, 34 (5): 26–34.
- Sachs, J.D. 1992. "Privatization in Russia: some lessons from Eastern Europe." *American Economic Review*, 82 (2): 43–48.
- 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.
- Santiso, C. 2004. "The Contentious Washington Consensus: Reforming the Reforms in Emergin Markets." *Review of International Political Economy*, 11 (4): 828–844.
- Schumpeter, J.A. 2006[1943]. *Capitalism, socialism and democracy*, Routledge.
- Shapiro, C. and J.E. Stiglitz. 1984. "Equilibrium unemployment as a worker discipline device." *The American Economic Review*, 74 (3): 433–444.
- 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." *Journal of Economic Perspectives*, 14 (3): 217–232.
- Solow, R.M. 1956. "A contribution to the theory of economic growth." *Quarterly Journal of Economics*, 70 (1): 65–94.
- Solow, R.M. 1957. "Technical change and the aggregate production function." *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." *Quarterly Journal of Economics*, 124 (2): 469–529.
- Stiglitz, J.E. 2002. *Globalization and its Discontents*, WW Norton & Company.
- Stiglitz, J.E. 2008. "Is there a Post-Washington Consensus Consensus?" in N. Serra and J.E. Stiglitz, eds., *The Washington Consensus Reconsidered*, Oxford University Press, USA, chapter 5, pp. 41–56.
- Szirmai, A. 2005. *Dynamics of socio-economic development: an introduction*, 2nd ed., Cambridge University Press.
- Szirmai, A. 2008. "Explaining success and failure in development." *UNU-MERIT Working Paper #2008-013*.

- Szirmai, A. 2012a. "Proximate, Intermediate and Ultimate Causality: Theories and Experiences of Growth and Development" *UNU-MERIT Working Paper*, forthcoming.
- Szirmai, A. 2012b. Dynamics of socio-economic development: an introduction, 3rd 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.
- Tilly, C. 1985. "War making and state making as organized crime." in P.B. Evans, D. Rueschemeyer, and T. Skocpol, eds., *Bringing the State Back In*, Cambridge University Press, pp. 169–191.
- Tilly, C. 1992. Coercion, capital, and European states, AD 990-1992, Blackwell.
- Uzawa, H. 1965. "Optimum technical change in an aggregative model of economic growth." *International Economic Review*, 6 (1): 18–31.
- Veblen, T. 1899. "The theory of the leisure class: An economic theory of institutions." Macmillan.
- Williamson, J. 1990. What Washington Means by Policy Reform, Washington, D.C.: Institute for International Economics.
- Williamson, J. "Did the Washington Consensus Fail?" Outline of speech at the Center for Strategic & International Studies, Washington, DC 2002.
- Williamson, J. 2008. "A Short History of the Washington Consensus." in N. Serra and J.E. Stiglitz, eds., *The Washington Consensus Reconsidered*, Oxford University Press, USA, chapter 1, pp. 14–30.
- 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." *Quarterly Journal of Economics*, 110 (3): 641–680.





## The UNU-MERIT WORKING Paper Series

- 2012-01 *Maastricht reflections on innovation* by Luc Soete
- 2012-02 *A methodological survey of dynamic microsimulation models* by Jinjing Li and Cathal O'Donoghue
- 2012-03 *Evaluating binary alignment methods in microsimulation models* by Jinjing Li and Cathal O'Donoghue
- 2012-04 *Estimates of the value of patent rights in China* by Can Huang
- 2012-05 *The impact of malnutrition and post traumatic stress disorder on the performance of working memory in children* by Elise de Neubourg and Chris de Neubourg
- 2012-06 *Cross-national trends in permanent earnings inequality and earnings instability in Europe 1994-2001* by Denisa Maria Sologon and Cathal O'Donoghue
- 2012-07 *Foreign aid transaction costs* by Frieda Vandeninden
- 2012-08 *A simulation of social pensions in Europe* by Frieda Vandeninden
- 2012-09 *The informal ICT sector and innovation processes in Senegal* by Almamy Konté and Mariama Ndong
- 2012-10 *The monkey on your back?! Hierarchical positions and their influence on participants' behaviour within communities of learning* by Martin Rehm, Wim Gijsselaers and Mien Segers
- 2012-11 *Do Ak models really lack transitional dynamics?* by Yoseph Yilma Getachew
- 2012-12 *The co-evolution of organizational performance and emotional contagion* by R. Cowan, N. Jonard, and R. Weehuizen
- 2012-13 *"Surfeiting, the appetite may sicken": Entrepreneurship and the happiness of nations* by Wim Naudé, José Ernesto Amorós and Oscar Cristi
- 2012-14 *Social interactions and complex networks* by Daniel C. Opolot
- 2012-15 *New firm creation and failure: A matching approach* by Thomas Gries, Stefan Jungblut and Wim Naudé
- 2012-16 *Gains from child-centred Early Childhood Education: Evidence from a Dutch pilot programme* by Robert Bauchmüller
- 2012-17 *Highly skilled temporary return, technological change and Innovation: The Case of the TRQN Project in Afghanistan* by Melissa Siegel and Katie Kuschminder
- 2012-18 *New Technologies in remittances sending: Opportunities for mobile remittances in Africa* Melissa Siegel and Sonja Fransen
- 2012-19 *Implementation of cross-country migration surveys in conflict-affected settings: Lessons from the IS Academy survey in Burundi and Ethiopia* by Sonja Fransen, Katie Kuschminder and Melissa Siegel
- 2012-20 *International entrepreneurship and technological capabilities in the Middle East and North Africa* by Juliane Brach and Wim Naudé
- 2012-21 *Entrepreneurship, stages of development, and industrialization* by Zoltan J. Ács and Wim Naudé
- 2012-22 *Innovation strategies and employment in Latin American firms* by Gustavo Crespi and Pluvia Zuniga
- 2012-23 *An exploration of agricultural grassroots innovation in South Africa and implications for innovation indicator development* by Brigid Letty, Zanele Shezi and Maxwell Mudhara
- 2012-24 *Employment effect of innovation: microdata evidence from Bangladesh and Pakistan* by Abdul Waheed

- 2012-25 *Open innovation, contracts, and intellectual property rights: an exploratory empirical study* by John Hagedoorn and Ann-Kristin Ridder
- 2012-26 *Remittances provide resilience against disasters in Africa* by Wim Naudé and Henri Bezuidenhout
- 2012-27 *Entrepreneurship and economic development: Theory, evidence and policy* by Wim Naudé
- 2012-28 *Whom to target - an obvious choice?* by Esther Schüring and Franziska Gassmann
- 2012-29 *Sunk costs, extensive R&D subsidies and permanent inducement effects* by Pere Arqué-Castells and Pierre Mohnen
- 2012-30 *Assessing contingent liabilities in public-private partnerships (PPPs)* by Emmanouil Sfakianakis and Mindel van de Laar
- 2012-31 *Informal knowledge exchanges under complex social relations: A network study of handloom clusters in Kerala, India* by Robin Cowan and Anant Kamath
- 2012-32 *Proximate, intermediate and ultimate causality: Theories and experiences of growth and development* by Adam Szirmai
- 2012-33 *Institutions and long-run growth performance: An analytic literature review of the institutional determinants of economic growth* by Richard Bluhm and Adam Szirmai