Public Administration
Economics & Governance
Faculty of Governance and Global Affairs
Leiden University

The Fundamental Determinants of Economic Performance

An Analyses of the Main Economic Development Hypotheses

Master Thesis
08/06/2017

Author: Tijmen Lennart de Kluijver
Student number: 1287036
Supervisor: Dr. Brendan Carroll
Second reader: Dr. Alexandre Afonso

Abstract

This research analyses the fundamental causes of economic performance. Economic development theory identifies the institution hypothesis, policy hypothesis, geographic hypothesis, and cultural hypothesis as elementary determiners of economic performance. Through analysing a newly merged dataset this thesis provides a platform for objective integrated regression, including all four different schools of thought into one model. The quality of institutions, openness to international trade, prevalence of disease, and human capital are found to be significantly related to economic performance. Those results are robust to the inclusion of additional controls as being war-torn and geographical region dummies.

Table of Contents

Chapter 1: Introduction

The world economy has gone through rapid changes in past centuries. The total world population, per capita income and world GDP increased exponentially. World GDP nearly rose 300-fold (Maddison, 2001, p. 17). This intensity of economic and social-democratic growth begins around the year 1820, starting of a period of exponential growth. Before this time the world economy was of incomparable size in proportion with current economies (Sachs, 2005, p. 26). All parts of the world had roughly comparable economies in the year 1820, and only during the period of modern economic growth inequality started to grow rapidly (Sachs, 2005, pp. 28-29). The process of economic growth was unevenly divided through space as well as time (Maddison, 2001, p. 17). Some countries profited from high economic growth, where others failed to develop their economy. This resulted in a high unequal distribution of wealth. At the present-day, the world is highly unequal and the gap between the richest and poorest countries is still growing. Some poor countries even grow poorer, relatively but also absolutely (Landes, 1998, p. xx). Countries in some cases are currently poorer than they were in the past (Collier & Gunning, 1999, p. 3). Western European countries have always been wealthy compared to other geographical areas. African countries are for example commonly associated with extreme poverty. This creates a clear-cut difference in winners and losers of economic development (Landes, 1998, p. xx). Individuals in rich countries enjoy high health standards, better education, the freedom to act as an individual free of state intervention and possess a general higher standard of wealth, poor countries mostly lack those characteristics (Acemoglu & Robinson, 2012, pp. 40-41). But there are also countries that are currently catching-up with the rich countries and are highly increasing their economic performance. The massive difference in world inequality has both consequences for the developing as for the developed countries. For this reason, it is in the interest of both rich and poor countries to help the poor become healthier and wealthier (Landes, 1998, p. xx).

1.1 Problem definition and research question

Understanding the underlying causes of unequal wealth distribution is crucial and provides the first step towards improve the standards of living throughout hole of the world (Acemoglu & Robinson, 2012, p. 41). To understand the reason for highly unequal economic performance around the world, attention should be focussed on what causes economic performance. Centuries of scientific research have shown us that accumulation of capital and differing rates of technological progress provide a good first step in determining economic performance. But those results cannot explain the massive differences in capital accumulation rates and technologic process in cross-country economic performance (Ahlfeld, Hemmer, & Lorenz, 2005, p. 1). There is in some way the need to involve cultural, geographical, political or policy elements in the analysis to provide a clear image.

Science is making good steps in analysing the cross-country differences in economic performance and report about the differences in accumulation. However, determining the causes of economic performance remains a problem (Collier & Gunning, 1999, p. 4). Although there has been significant research to the fundamental causes of economic performance and growth, no consensus has been reached. Scholars of economic development focussed on distinctive fields and emphasized different explanations in the search for the ultimate drivers of economic growth (Ahlfeld, Hemmer, & Lorenz, 2005). This lead to the development of different school of thought. In my view scholars supporting different schools of thought have not yet been able to fully understand the mindset of the other schools and fail to include all essential elements into one comprehensive model.

This thesis critically assessing a combination of different hypotheses on fundamental causes of economic performance and testing them to a newly gathered dataset. In this way hoping to contribute to the debate over the fundamental causes of cross-country economic performance. Guiding both future research as international development assistance with a focus of determinants of cross-country economic performance. The research will be conducted around the following key research question:

What are the fundamental determinants of cross-country differences in economic performance?

1.2 Academic & practical relevance

This thesis provides a platform for the four main theoretical camps of economic development to be analysed next to each other. Existing scientific research on the different schools did in some way include control elements of other schools in their analyses, but lack the use the correct methods and measurements of this school as conducted by to their leading scholars. Structural integrating all those theories will provide an overview which relationships between variables are strong enough to explain differences in cross-country economic performance. Where scholars are focused on their own viewpoint this thesis provides a purely objective view on what causes differences in economic performance between countries worldwide. Providing a structural integrated regression provides a clear overview which variables critically effect economic performance, while controlling for other variables. Findings or this inquiry can guide future research in the right track for determining the main causes of economic performances and provides understanding of the sources of world inequality.

The practical relevance of economic development studies already became clear in the first part of this chapter. Massive inequality in cross-country economic performance harms both poor and rich countries. This thesis contributes to the search of fundamental causes of economic performance. Extending the knowledge of the causal mechanisms behind economic performance helps improve the strategies used in international development assistance (Ahlfeld, Hemmer, & Lorenz, 2005, p. 14). The

accessibility of knowledge helps to assist countries in increasing their economic performance. Especially by guiding the least developed countries in their development world inequality can party be countered.

1.3 Reader's guide

This thesis consists of a series of chapters to guide the reader through my conducted research. Chapter 2 describes the relevant literature of the field of economic development relevant for this research. Chapter 3 contains of the theoretical framework. The research design and its limitations are discussed in chapter 4. Chapter 5 consist of the regression outcomes. This chapter discusses the findings of analysis. This thesis concludes with the discussion and conclusion, described in chapter 6.

Chapter 2: Literature Review

Economic development is the general field of study regarding the research question. It is a wide theme with a large history of scientific study. It is a field of study that tries to point out the fundamental elements that determine the sources of differences in cross-country economic performance (Acemoglu D. , 2009, p. XV). There are two main approaches in assessing the reasons for differences in cross-country economic development, the formal approach and the informal approach (Ahlfeld, Hemmer, & Lorenz, 2005, p. 3). This chapter will provide an overview of the main theories of economic development. Starting with examining formal regression of economic development. The used methods and the fit with answering the research question will be clarified. Afterward the literature of informal regression will be discussed.

2.1 Formal regression

Formal regression analysis is based on determining theoretical foundations of economic growth beforehand (Ahlfeld, Hemmer, & Lorenz, 2005, p. 3). Classical macro-economic scholars that are interested in explaining the causal relationship between relevant elements and economic growth in this way are for example Adam Smith, Thomas Malthus, John Maynard Keynes and some of the founding fathers of modern growth theory Robert Solow and Trevor Swan (Baumol, Litan, & Schramm, 2007). The Solow Model (and the extended Mankiw, Romer and Weil Model) is widely used in construction those a priori assumptions for economic growth (Ahlfeld, Hemmer, & Lorenz, 2005, p. 3) In those analysis capital, human capital, and technological progress are the largest determines of economic growth (Ahlfeld, Hemmer, & Lorenz, 2005, p. 3). Those models provide valid empirical evidence for the determining factors of growth. However, those models do not provide a definitive answer to the questions why some countries are poor, and others rich (Ahlfeld, Hemmer, & Lorenz, 2005, p. 3). Because this approach does not fit the research question of this thesis only the abstract version will be discussed. Theoretical growth models from the basics of formal regression analysis. Even before the introduction of Solow's Growth Model other auteurs discussed theories of economic growth (Solow, 1994) . Just like the Solow Model, those models focus on the same general economic assumptions for requiring economic development (Solow, 1994, p. 46). The Solow Growth Model is a standard neoclassic model of economic growth and is an extension of the Harrod-Domar Model[1] (Solow, 1994). Another scholar that developed a similar growth models is Trevor Swan (Solow, 2007, p. 3). Both created comparable models and for this reason the model is sometime also revered to as the Solow-Swan neoclassic growth model (Lee, Pesaran, & Smith, 1997).

[1] Based on the original articles by Harrod (1939) and Domar (1946)

The model captures the three main causes of economic growth:

- Technology
- Capital accumulation
- Labour force

The model predicts that increasing capital accumulation and labour force will increase the economic growth rate (Solow, 1994). This growth will however only last temporarily because of diminishing returns. Diminishing returns create a situation where the economy will grow at a steady rate. This growth model is based on purely theoretical economic assumptions that are not easily translated into practical elements. Despite its age, the model continues to be of main importance in the literature (Lee, Pesaran, & Smith, 1997). Technological progress also creates economic growth through innovation. The model originally consists of highly complicated mathematical equations. Those equations will however not be discussed because of the lack of value they add to this literature review.

One of the key assumptions of this model is that less developed countries will catch-up with more developed countries due to the law of diminishing returns to reproducible capital (Sarkar, 1998, p. 2). Poor countries with a low stock of capital per worker can profit from more marginal productivity of capital and from a higher rate of return of capital (Sarkar, 1998, p. 2: Barro, 1996, p. 4). This leads to more economic growth relative to rich countries till the steady-state level of output is reached (Sarkar, 1998, p. 2). This assumes convergence in cross-country economic performance over time (Lee, Pesaran, & Smith, 1997, p. 358). The assumption of the catch-up effect contributed to the debate about convergence or divergence of less developed countries. This debate is conducted by a create number of specialised scholars over the last decades (Solow, 2007). The debate is formed by two opposing sides that form the discussion: the catch-up/ convergence hypothesis and the divergence hypothesis (Sarkar, 1998). The divergence hypothesis is based on the fundamental idea of uneven development. It argues that advanced countries profit from an initial higher productivity than the less advanced countries, no matter the conditions (Sarkar, 1998, p. 1). This leads to an ever-increasing gap between poor and rich countries (Sarkar, 1998, p. 1). This debate creates its own point of study and will not be elaborated upon to extensively. It is however important to recognize the existence of this debate before conducting more in-depth literature review. Certain theories part of the formal regression of economic development argue that convergence is one of the fundamental elements that helps to explain economic performance.

The ‘new’ formal growth theories created by Romer, Lucas, Rebelo and other scholars include the importance of human capital and criticize and extent the neoclassic model (Sarkar, 1998: Ahlfeld, Hemmer, & Lorenz, 2005: Mankiw, Romer, & Weil, 1992: Barro, 1996). Other even more recent studies

point out that the rate of convergence has decreased. Some countries even seem to diverge, indicating that the differences of cross-country economic performance are increasing (Economist, 2014). This contradict the assumption of convergence set up in the Solow growth model. Those new analyses of economic convergence let us doubt if the major growth models are still useful in analysing economic growth in recent times.

The formal regression literature has certain gaps in its literature what creates doubt about the usefulness of this model. One of the well-known 'new' growth models is developed by Mankiw, Romer and Weil (MRW) (Mankiw, Romer, & Weil, 1992). This model has a high degree of explanatory power (Ahlfeld, Hemmer, & Lorenz, 2005, p. 3). It indicates that economic growth is determined by the level of accumulation of capital, human capital (both high) and level of fertility rates (low) in the past few decades (Ahlfeld, Hemmer, & Lorenz, 2005). But this model fails to explain why some countries are capable to accumulate more capital or have more technological process than others (Ahlfeld, Hemmer, & Lorenz, 2005, p. 3). The different focus of formal regression on economic development combined with the doubts of its main theoretical models let me look further in economic development theory in search of a useful theoretical framework. Formal regression fails to provide answers to a basic question of economic development concerning this thesis: 'What can explain the major differences in economic performances between countries?'.

2.2 Informal regression

Informal regression analysis provides a framework for the search for the fundamental drivers of economic growth, and in this way, could provide explanations for cross-country inequality (Ahlfeld, Hemmer, & Lorenz, 2005, p. 3). Next to accumulation of capital, human capital and technological progress additional influential variables on growth need to be considered (Ahlfeld, Hemmer, & Lorenz, 2005, p. 3). Those variables can be divided into different classes: the geographic hypothesis; the institutions hypothesis; and the policy hypothesis (Ahlfeld, Hemmer, & Lorenz, 2005, p. 4).

Maddison (2001) seeks to explain the differences in world economic performances in the long run for the period from 1000 A.D. till around the end of the twentieth century. He refers to geographical factors, international trade and capital movements and technological and institutional innovation as leading interactive processes determining economic performances (Maddison, 2001, p. 18). He also recognizes the importance of human capital as a factor for economic growth (Maddison, 2001, p. 23). Other influential scholars come up with comparable classes. Acemoglu and Robinson (2012) try to explain the huge differences in incomes and living standards between rich and poor countries in their book 'Why nations fail' (Acemoglu & Robinson, 2012). Although this book strongly support the explanatory power of institutions on economic performances, it recognizes the determining factors

other academics use. Those academics use determinant factors like geography, cultural attributes and failing political policy for explaining differences in economic performances (Acemoglu, Robinson, 2012, p. 3). Analysing the existing literature of development economics made clear that literature concerning this point of study can be divided in four schools of thought:

- Institutions hypothesis
- Policy hypothesis
- Geographic hypothesis
- Cultural hypothesis

Those schools of thought provide platforms of understanding economic development and each consist of their own theoretical framework. Upcoming pages will provide a summary of the most important literature on those four hypotheses and their interconnectors. Every hypothesis is reviewed through the same structure. First the general idea of the hypothesis is described and this section is followed up with the general development of the hypothesis and its relationship with the other hypotheses. The most important implementations and relevance of every school of thought will be extensively addressed in the theory. Theories emphasized by the hypotheses will provide direction in the search for fundamental causes of economic performance.

2.1.1 Institution hypothesis

The decisive role of institutions in economic performances has become an influential perspective in explaining cross-country economic differences. The institution hypothesis argues that differences in the quality of institutions causes differences in economic performance across countries (Ahlfeld, Hemmer, & Lorenz, 2005, p. 8). Countries with better institutions represent more secure property rights and less distortionary policies of the government. Increasing the level of income through more investments in physical and human capital (Acemoglu, Johnson, & Robinson, 2001, p. 1369). North and Thomas (1973) were among the first scholars who published about the link between property rights and economic growth. They argue that economic growth will occur when property rights are protected. This makes it beneficial for civilians to undertake socially productive activity, what in its own way establishes economic growth (North and Thomas, 1973, p. 8). Establishing property rights brings along costs, but will stay beneficial if they top transaction costs in a situation where property rights are not established. Governments can authorize those rights in a more efficient way than private organizations (North and Thomas, 1973, p. 8). In this way institutions cover the task of protection property rights. North and Thomas review the economic history of the western world through those assumptions. North and Thomas set a new way of looking at the causes of economic development which would be elaborated upon by many other influential scholars to come.

Development hypothesis

Acemoglu and Robinson (2012) highly contradict other hypothesis of explaining economic performances in their book (Acemoglu and Robinson, 2012, pp. 45-69). Per those auteurs geographic, cultural and policy factors only affect economic development through institutions Acemoglu and Robinson, 2012, p. 3). Acemoglu and Robinson (2012) link politics to economics to explain world inequality (Acemoglu and Robinson, 2012, pp. 68-69). Economics are needed to link the effect of different types of policies to economic incentives and behaviour. Politics are needed to determine how decisions are made (Acemoglu and Robinson, 2012, pp. 68-69).

One of the most important reports arguing for the importance of institutions is conducted by those same auteurs. Acemoglu et al. (2001) use current institutions and their connection to the instrumental variables: (potential) settler mortality, settlements, and early institutions to describe differences in current cross-country economic performances (Acemoglu, Johnson, & Robinson, 2001). They argue that European colonialism pursued different types of colonization policies which created different sets of institutions. Some colonies where constructed as permanent residences, like North America, and other colonies where mainly aimed on extracting as much of the natural recourse available from the colony, mainly in Africa and Latin America (Acemoglu, Johnson, & Robinson, 2001, pp. 1374-1375). This creates a distinction between extractive and inclusive institutions**.** Assuming that the colonization strategy was influenced by the feasibility of settlements and that colonial state and institutions persisted after independence they link the instruments settler mortality to amount of settlements, early institutions and even to current institutions (Acemoglu, Johnson, & Robinson, 2001). They then argue that good institutions emerge in settler colonies because of the copying of European institutions. Creating a situation where institutions enable incentives for economic activity. Colonies used for their natural resources lack the development of those kind of institutions. According to Acemoglu et al. (2001) this difference in development of institutions causes the big difference in current economic performances between countries.

Those auteurs used estimates of settler's mortality rates faced by soldiers, bishops and sailors during the seventeenth, eighteenth and nineteenth centuries (Acemoglu, Johnson, & Robinson, 2001, p. 1375). This settler mortality data received critique from other scholars. David Albouy contributed to this critique, claiming that 36 of the 64 countries of Acemoglu et al. (2001) database are assigned with wrong mortality rates (Albouy, 2008). Reassigning correct mortality rates would case the relationship between institutions and economic performance to become insignificant. The different viewpoint of the effect of institutions on economic performance and the relationship with mortality rates created

a long-lasting debate between scholars. This debate also let us doubt about the effect of institutions on economic performance.

Relationship with other hypotheses

Supports of the institution hypothesis recognize the importance of the other hypotheses but argue that those hypotheses only indirectly effect economic performance through institutions. Acemoglu rejects the direct effects of the other hypotheses in his book and describes why those theories fail to explain differences in cross-country economic performance (Acemoglu & Robinson, 2012). Easterly and Levine also test for this relationship and come to the same conclusion (Easterly & Levine, 2002). Following their argumentation only the difference in institutional quality majorly accounts for economic performance differences.

2.1.2 Policy hypothesis

Some scholars believe that the regardless of the factors described in the other hypotheses domestic macro-economic policies determine the country's economic performance (Ahlfeld, Hemmer, & Lorenz, 2005: Easterly & Levine, 2002). Ahlfeld et al. (2005) describe this school of thought as the policy-hypothesis (Ahlfeld, Hemmer, & Lorenz, 2005, p. 12). This hypothesis seeks to explain economic performances through emphasizing different approaches of governmental economic policy. The creation of a stable macroeconomic framework through effectively using international trade could account for the increasing welfare (Ahlfeld, Hemmer, & Lorenz, 2005, p. 12). Acemoglu and Robinson (2012) use the term the 'ignorance hypothesis' to describe this school of thought (Acemoglu & Robinson, 2012, p. 63). This hypothesis emphasizes the importance of (failing) macro-economic policy of the government. The hypothesis is closely linked to the First Welfare Theorem, which identifies the ideal situation of a market economy. This situation is favoured in macro- economic theory and provides the best climate for economic development. Market economy describes the situation where individuals and firms are free to purchase economic activities (Acemoglu and Robinson, 2012, p. 64).

Development hypothesis

Neoclassical growth models describe technological change as exogenous in economic growth models. Arguing that a country's economic performance is unaffected by the degree of openness of the economy to world trade (Harrison, 1996, p. 419). Modern growth theories do recognize the impact trade policy could have on long-run growth (Harrison, 1996, p. 419). Harrison (1996) describes the correlation between openness of the economy and economic growth. Trade policy is one of the major outcomes of a country's macro-economic policy that could determine development. Frankel and

Romer (1999) also discuss the effects of trade on income. This relationship is logical highly subject to reversed causality. By instrumenting for trade, they conclude that trade has a statistical and practical significant effect on income (Easterly & Levine, 2002, p. 12). In their view developing countries lack the ability to efficiently interact with the major trading markets. This holds back their economic development (Frankel & Romer, 1999). Frankel and Romer do however conclude that the correlation between trade and income cannot determine the direction of causality (Frankel & Romer, 1999, p. 379). The two variables are integrated in such a way that determining the relationship and strength of the causal relationship is complicated (Harrison, 1996, p. 443).

Sachs and Warner (1997) come to the same conclusion through analysing the sources of growth in Sub-Saharan Africa. They emphasize the role of (poor) economic policies, with a special focus on the lack of openness to international markets, as cause for the slow economic growth of Africa (Sachs & Warner, 1997, p. 361). Africa lacks behind in the process of trade liberalization (Sachs & Warner, 1997, p. 362). Per Sachs and Warner (1997) Africa's economy could have grown over 4 percent per year per capita if the governments would have used appropriate policies (Sachs & Warner, 1997, p. 362). The slow growth of Africa's economy can be explained using the same variables used to describe the economic performance of other countries. Meaning that there is no need for a unique 'Africa theory' of economic development (Sachs & Warner, 1997, p. 337).

Relationship with other hypotheses

The ignorance hypothesis differs from the other hypothesis in the way that it comes up with suggestions how to solve the underlying problem. For this reason, multilateral organisations emphasize with this approach and imbedded the main point in their policy recommendations (Easterly & Levine, 2002, p. 11). Following this hypothesis enlightenment of the economic situation and know how to solve contemporary problems provide the opportunity to counter market failures and fight poverty (Acemoglu & Robinson, 2012, p. 67).

Economic institutions and macro-economic policy are closely connected and the distinction between the two variables can be blurred (Sachs & Warner, 1997, pp. 337-338). Conceptualization and operationalization as will be discussed in chapter 4 of this thesis will provide a useful framework to analyses the different effects of the variables on economic performance. This combined with the output of regressions will point out if, and in which way the variables are causes of economic performance, and if they are internally linked to each other.

Frank and Romer describe the relationship between geographic factors, government policy, and economic development (Easterly & Levine, 2002, p. 12). Arguing that geographic factors affect

economic performance through governmental policy, with the focus on trade openness (Frankel & Romer, 1999). Geographical factors however would only influence economic performance through its effect in trade volume (Ahlfeld, Hemmer, & Lorenz, 2005, p. 12).

2.1.3 Geography hypothesis

General theories that acknowledge the importance of geographical factors as causes of world inequality can be summarized as the 'geography hypothesis' (Acemoglu & Robinson, 2012, p. 48: Easterly & Levine, 2002, p. 3). Geography hypothesis can be divided into two broad sub categories that explain the effect on economic performance in different ways: geographical location and climate conditions (Ahlfeld, Hemmer, & Lorenz, 2005). Those sub categories are collectively recognized by scholars active on this focus of economic development and can again be divided in different trends. More in depth information about those sub categories and the underlying causal mechanisms will be provided in the theory.

Geographic factors are easy to understand and can be measured through many different indicators. Geographical factors can for example be distinguished in sub-categories like: a country's climate, natural resources, transport costs, disease burden, population density, agricultural productivity, and many others. All those factors are determined by the geographical characteristics of the country and could directly affect economic performances. This creates a situation where scholars use a great number of different indicators for measuring the effect of geographic factors.

The idea that geographical factors could cause differences in economic performance is reasonable when we compare the richest countries with the poorest countries. Most rich countries cluster around Western Europe and countries with related characteristics, the poorest countries cluster around (Sub-Saharan) Africa and countries with related continental characteristics (Acemoglu & Robinson, 2012, p. 45: Gallup, Sachs, & Mellinger, 1999, p. 183). Sub-Saharan Africa is the lowest income region in the world and seem to suffer from a chronical failure in acquiring economic growth (Collier & Gunning, 1999, p. 4). When analyse those regions through different periods of time the overall picture would be remarkably consistent as the current situation (Acemoglu & Robinson, 2012, pp. 45-46) However, currently some leading developing countries with other negative associated geographical factors do catch up with the developed countries (Acemoglu & Robinson, 2012, p. 48). Arguing for the insufficient explanation offered by the geographic hypothesis.

Development hypothesis

Montesquieu was one of the first scholars to publish about the relationship between climate conditions and economic development (Ahlfeld, Hemmer, & Lorenz, 2005, pp. 4-5: Easterly & Levine, 2002, p. 5). He uses different explanations to argue that warm climates create a situation where people are rewarded to rest instead of work. Warm climates in this way suppress economic activity (Montesequieu, [1748] (1989)). Another influential scholar who was one of the first to acknowledge the importance of physical geography was Adam Smith. Smith stressed the importance of transport costs and their ability to negatively influence economic development (Sachs, 2005, p. 58).

Modern theories of economic development mostly neglected the effect of climate conditions for a long period (Ahlfeld, Hemmer, & Lorenz, 2005, p. 5). Landes (1998) argues that this is partly due to Ellsworth Huntington who over included to many different element in the geographical hypothesis. The hypothesis lost its credibility and acceptability due to its classifying of groups of people related to racial thoughts (Landes, 1998, p. 4). Only recently this school of thought is once again part of the economic development debate, using climate's importance for agriculture and labour productivity and accumulation as channels for measuring the effect (Ahlfeld, Hemmer, & Lorenz, 2005, p. 5).

Sachs et al. (2001) provide a useful way of capturing the effect of geography on economic performance in three major ways: transport costs; prevalence of disease; and agriculture productivity. Easterly and Levine (2003) also provide a useful way of capturing geographic factors in three main elements: tropics, germs and crops. Those elements represent the following factors in order: climate zone, disease environment conditions, and agriculture productivity (Easterly & Levine, 2002). In their research, Easterly and Levine however notice that geographical factors only directs economic development through their effect on human diseases and on natural resources (Easterly & Levine, 2002, p. 24). Those three main elements are commonly used by other scholars for analysing the geography hypothesis, although sometime in a slightly different phrasing (Gallup, Sachs, & Mellinger, 2001: Sachs, 2003).

Relationship with other hypotheses

Many of the modern influential auteurs of the geography school recognize the importance of institutions for economic development (Easterly & Levine, 2002: Gallup, Sachs, & Mellinger, 1999). Easterly and Levine find evidence that tropics, germs and crops as indicators of geographical factors only affect economic development through institutions (Easterly and Levine, 2003). Gallup, Sachs and Mellinger believe that both geographical factors as institutions directly affect economic development (Gallup, Sachs, & Mellinger, 1999, p. 182). They recognize the possible complementary function of physical geography and agglomeration economies (Gallup, Sachs, & Mellinger, 1999, p. 184).

Scholars that support the institutional hypothesis argue that geography only indirectly affects income through institutions (Acemoglu & Robinson, 2012, pp. 48-49). Other authors however keep publishing articles where they defend the direct affect geographic factors have on economic performance after controlling for institutions (Sachs, 2003). He argues that concluding that geography has no direct effect on productivity doesn't fit the historical evidence (Sachs, 2003, pp. 2-3). Many of the institutional choices in the past are directly affected by geographical factors (Sachs, 2003, pp. 2-3). This discussion about the direct or indirect effect of geography on economic performance provides enough reason to include geographic factors in the search for the fundamental causes of economic performance.

2.1.4 Cultural hypothesis

The culture hypothesis stresses the importance of culture in explaining different levels of economic performance. This does not only include religion, but also include values and ethics (Acemoglu & Robinson, 2012, p. 57). The cultural hypothesis differs from the other hypotheses in a way that not all scholars active on the topic of economic development recognize this hypothesis as an independent school of thought. Some scholars separately analyse social and human capital as control when analysing the causes of economic performance, but do not recognize the cultural hypothesis as crucial factor.

Development hypothesis

The first publications of the cultural hypothesis related to the prosperity of culture. Max Weber recognized religion as one of the fundamental causes of economic performance (Acemoglu, Johnson, & Robinson, 2001, p. 1388). Cultural of religious norms could hinder economic development when they postpone certain parts of the population. Undermining part of the population from economic activities decreases the potential economic development (Sachs, 2005, p. 60). Modern scholars like Knack and Keefer (1997) strongly support the culture hypothesis. Landes (1998) describes the importance of culture as cause of economic performance in his book 'The Wealth and Poverty of Nations: Why some are so rich and some so poor'.

The cultural hypothesis does not only include religious factors but stretches its definition and includes beliefs, values and ethics to its domain (Acemoglu & Robinson, 2012, p. 57). Modern scholars changed the focus of the hypothesis to mainly social and human capital. Those two concepts will be used as representatives for this hypothesis in this thesis. Social capital is not homogenous, it can take on many different shapes (Putman, 2000, p. 1). This leaves substantial room for different indicators. The change of focus in the field of social capital brought some measurement problems along, especially in measuring long-term trends (Putman, 2000, p. 3). A change of focus in literature requires new methods

of collecting data. This new data gathering is only possible for contemporary research but lacks comparable data for historic analyses.

The concept of social capital has historically known marginal different definitions. However, there has been convergence in identifying the concept of social capital (Putman, 2000, p. 1). Van Schaik (2002) provides a bread, but clear definition of the concept of social capital:

> "Social capital thus refers to qualities in social relationships which enhance the capacity of the participants to achieve their interests and which at a more general level constitute a resource for social development. Those important qualities are interpersonal trust, mutual supportiveness, shared norms and understanding." (Van Schaik, 2002, p. 6)

This definition captures the main idea of the concept and directly links it to (social) development. The causal mechanisms underneath will be discussed in chapter 3. Fukuyama has highly contributed to social capital theory by describing the effect of trust. He argues that trust forms a basic for other social capital to emerge (Van Schaik, 2002, p. 6). Social capital can only arise if social networks are maintained (Van Schaik, 2002, p. 7). Knack and Keefer (1997) react on the book published by Putman (1993) 'Making democracy work'. They argue that Putman's measurement of social capital, through the measurement of memberships in formal groups, does not accurate measures social capital. In their opinion trust and civic norms do, and they find significant associations between those variables and economic performance (Knack & Keefer, 1997, p. 1251).

Human capital as measured through education is found to be highly correlated with economic performance. Many scholars found positive correlations between schooling and growth rates of economic performance across countries (Bils & Klenow, 2000, p. 1160: Glaeser, 1993). On more year of school enrolment is found to be associated with an additional 30 percent higher annual growth (Bils & Klenow, 2000, p. 1160). Bils and Klenow do however argue for nuancing the effect education has on economic performance. There is a possibility that the strong relationship could be due to omitted factors related to the level of education and high economic performance (Bils & Klenow, 2000, pp. 1160-1161). Additionally, the relationship could be highly subject to reversed causality (Bils & Klenow, 2000, p. 1161). In this example reversed causality would mean that economic development would cause an increase in human capital. Found correlations do not provide specific mechanisms how schooling creates economic growth (Glaeser, 1993, p. 333). Because of those problem causal relationships are hard to indicate.

Relationship with other hypotheses

As mentioned, social capital is often neglected as possible fundamental cause of economic performance by other hypotheses. Human capital is however highly included as control in other hypotheses. Human capital is closely linked to institutions. There is an intense discussion about how to view the effect of human capital on development. Glaeser et al. (2004) defend the position that human capital directly affects economic development on its own (Glaeser, Porta, Lopez-de-Silanes, & Shleifer, 2004). They argue that human capital is the driving force behind economic performance and even for the creation of institutions. A high degree of human and social capital creates a situation where institutions can flourish (Glaeser, Porta, Lopez-de-Silanes, & Shleifer, 2004, pp. 295-298). Acemoglu, et al. (2014) however, are convinced that human capital and institutions are interrelated in another way (Acemoglu, Gallego, & Robinson, 2014, p. 875). In their view institutions create an environment where human capital can be created (Acemoglu, Johnson, & Robinson, 2001, p. 1369). The discussion about the relationship between institutions, human and social capital, and economic performances provides enough ground for this thesis to include all elements as separate variables.

2.3 Summary

Economic development is a wide theme with a big history of scientific study. Formal regression is good in explaining the causes of economic growth, but fails to describe what determines the differences in economic performance between countries. Informal regression does provide the possibility to clarify the underlying causes of economic performance. It separates theories linked to the fundamental determinants into four hypotheses: the institution hypothesis, policy hypothesis, geography hypothesis and cultural hypothesis.

Chapter 3: Theory

In this chapter I will describe the most important theoretical element described in economic development literature. Upcoming four components are directly linked to the hypotheses described in the literature review. I start by describing theory about the effect of institutions on economic performance. Additionally, the effects of macro-economic policy, geographical factors, and social and human capital on economic performance will be discussed. Every section is constructed in the same way. First the general concept of the theory is described. Additionally, the elemental causal mechanisms are specified. Lastly, the general theoretical expectations are summarized.

3.1 Institutions

North (1991) describes institutions as the following: "Institutions are the humanly devised constraints that structure political, economic and social interaction." (North, 1991, pp. 97). Institutions include both formal rules (property rights, constitutions and laws) and informal constraints (North, 1991, p. 97). Informal constraints provide unofficial guidelines for what is socially acceptable (Ahlfeld, Hemmer, & Lorenz, 2005, pp. 8-9). Institutions provide incentives that constrain human activity and form the structure of the economy (Acemoglu & Robinson, 2012, p. 73: North, 1991, p. 97). Secure private property and an unbiased system of law are crucial for forming well established institutions (Acemoglu & Robinson, 2012, pp. 73-74). Institutions can take the form of both economic or political institutions. Both kind of institutions are crucial for the effective institutional matrix (North, 1991, p. 98).

Political institutions form formal economic constraints through enforcing property rights, constitutions and laws. They provide the underlying framework which creates the possibility for economic development. Economic institutions produce a set of economic rules of the game that generate the opportunity for sustainable economic growth (North, 1991, p. 98). Acemoglu and Robinson describe the relationship between the two forms of institutions as following:

> "… while economic institutions are critical for determining whether a country is poor or prosperous, it is politics and political institutions that determine what economic institutions a country has." (Acemoglu & Robinson, 2012, p. 43)

Both kind of institutions interact and are highly path dependent (Acemoglu and Robinson, 2012, pp. 43-44). Especially political institutions are highly change resistance, and they are the institutions that determine the rules of society (Acemoglu & Robinson, 2012, p. 44). Both political and economic institutions determine economic development.

Acemoglu and Robinson (2012) distinguish extractive and inclusive economic institutions. Inclusive economic institutions are based on secure private property, objective law systems, and provision of public services. Those characteristics provide equal chances for individuals and firms for economic activities (Acemoglu & Robinson, 2012, pp. 74-75). Extractive economic and political institutions lack the incentives for economic activity and are aimed only on enriching the group the benefits from the extraction (Acemoglu & Robinson, 2012, p. 372). There is no possibility for efficient economic activities and this leads to failure of economic development and could even lead to massive social tensions (Acemoglu & Robinson, 2012, pp. 372-373).

Causal mechanism

The general underlying economic theory of institutions is that the provision of good institutions creates a situation which allows the economy to flourish. Institutions create stability and certainty by determining a framework for transaction and production costs (Acemoglu, Johnson, & Robinson, 2001, p. 1369: North, 1991, p. 97). North argues that transaction costs are critical in assessing economic performance and that both transaction as production costs can be lower through using effective institutions (North, 1991, p. 98). The provision of incentives through institutions allows and encourage participation of individuals and firms in economic activities (Acemoglu & Robinson, 2012, p. 74). A good provision of political and economic institutions creates an economic environment that induces increasing productivity, followed by a high degree of economic performance (North, 1991, p. 98). Resources are allocated more efficiently and cause a greater level of economic development (Acemoglu, Johnson, & Robinson 2001, p. 1369). Absence of such institutions discourage investment and specialization what leads to lower economic performances (Acemoglu & Robinson, 2012, p. 75: Knack and Keefer, 1995, p. 207).

Theory of the effect of institutions on economic performance are supported by significant results in many regression of data sample. This could however also be due to reversed causality. Countries with high economic performance are expected to also possess high levels of good quality institutions (Acemoglu, Johnson, & Robinson, 2001, p. 1369). Growing economies could cause the rise of good institutions (Glaeser, Porta, Lopez-de-Silanes, & Shleifer, 2004).

3.2 Macro-economic policy

The literature review provides a good direction for the search of theories linking macro-economic policy to economic performance. The main argumentation of this hypothesis assumes that governmental policies affect the domestic economic situation. The causal mechanisms underneath is related to market openness and market failures. Additionally, policy failures will inevitably also affect

the domestic market. Economic theory however focuses on the first two factors because they provide a useful framework for cross-country comparable study. Country specific policy failures are useful for describing case specific economic development, but do not contribute to finding a valid answer on what causes economic performance on world scale. For this reason, this hypothesis focusses on general macro-economic policy.

A problem of the policy view is that this instrument, in contrast to the other hypotheses, is mainly focussed on short-term (Ahlfeld, Hemmer, & Lorenz, 2005, p. 13). A country's macro-economic policy is formed by its government. Most countries, especially countries with democratic regimes, are subject to continuously changing governments. A change in government could cause a change in macro-economic policy. Additionally, governments in power tend to be more focussed on short-term effects (Ahlfeld, Hemmer, & Lorenz, 2005, p. 13). Other hypotheses tend to affect economic performance through more stable factors, making the search for fundamental causes of economic performance a bit easier.

Causal mechanisms

The assumption that trade openness leads to increasing economic performances is a classical economic assumption. A high degree of openness to international trade is associated with high economic development (Easterly & Levine, 2002, p. 11). Good allocation of the production factors and making use of competitive advantages of different countries leads to economic growth. Liberal trade regimes arrange for this efficient allocation and are desirable per this hypothesis (Ahlfeld, Hemmer, & Lorenz, 2005, p. 12). Ahlfeld et al. describe the importance of liberal trade as following:

> "International trade facilitates the realisation of economies of scale, intensifies competition in domestic markets and supports the creation, diffusion and absorption of foreign technologies." (Ahlfeld, Hemmer, & Lorenz, 2005, p. 12)

This sentence provides framework for several general macro-economic theories. The main idea of those economic theories is that trade positively effects economic performance. A high degree of openness to international trade is associated with liberal markets. Openness is an instrument that provides the opportunity for trade to flourish and economies to grow. The other way around, anti-export policy decreases the exports and reduces the economic performance of a country (Collier & Gunning, 1999, p. 14). Governmental policies could in this way harm the openness of the economy through interfering with the free market. Those policies could take form as import tariffs or strict import rules. Making it less profitable for other countries to export to the home country.

Openness of an economy and economic performance are subject to the issue of possible reversed causality. Literature suggests that causality runs in both directions (Harrison, 1996, p. 443). Openness causes higher economic performance, but well preforming economies also tend to trade regimes. This makes it hard to determine the direction of the relationship.

Market failure occurs when the circumstances for an ideal market economy are not met. The policy hypothesis argues that the fight against market failures, and effectiveness of reaching a market economy determines if a country is wealthy of not. Ignorance also includes the non-willingness of governments to improve the economic situation of a county. This could be the case when governments sustain the current, less efficient, economic situation to maintain their political position (Acemoglu & Robinson, 2012, p. 65). Countries with well-organized governmental policies are better capable of providing policies to counter market failures. This is translated into more economic growth. Countries characterized by their governmental policies do not possess the ability to create the situation of market economy (Acemoglu & Robinson, 2012, p. 64). Governments can consciously or unconsciously create policy that harm the accessibility of their economy to world trade. Examples of such policies are import tariffs and strict import regulation.

3.3 Geographical factors

The three major factors of the geographic hypothesis (geographic location, prevalence of disease, and agricultural productivity) described in the literature review are recognized by different auteurs. Those factors overlap and help to form the main theoretical sub elements of the geographic hypothesis. All three factors fit the already described sub categories of the geographic hypothesis. Geography has its own subcategory and disease burden and agricultural productivity are part of the climate conditions category.

Causal mechanisms

The geographic location affects economic performance mainly through transportation costs (Ahlfeld, Hemmer, & Lorenz, 2005, p. 6). Transport costs arise when the production regions and major markets are separated. Big distances, rough terrain of other trade barriers increases transport costs (Collier & Gunning, 1999, p. 13). Ahlfeld et al. (2005) emphasizes the access to waterways and the proximity to global centres of commerce as most important elements. Transport costs increase the market price and lower the competitiveness of the producing region, which in turn lowers economic growth. Trade through sea transport is less costly than land- or air-based trade (Ahlfeld, Hemmer, & Lorenz, 2005: Gallup, Sachs, & Mellinger, 2001). Countries with access to sea of other water transport have an advantage on pure landlocked countries.

A country is landlocked when it is surrounded by neighbour countries and lacks direct access to the sea. This could lead to increased transportation costs and high barriers for trade (Collier & Gunning, 1999, p. 13: Sachs & Warner, 1997, p. 5: Sachs, 2005, pp. 57-58). Cross-border trade is subject to higher costs and other difficulties compared to internal trade (Gallup, Sachs, & Mellinger, 1999, p. 184). Organization of infrastructural development and labour migration across border are tougher to accomplish. Additionally, landlocked countries are subject to economic incentives of coastal countries and could increase costs or trade ensue those incentives (Gallup, Sachs, & Mellinger, 1999, p. 184). Being landlocked is however not a determining for low economic performance (Sachs, 2005, pp. 58-59). There are highly developed countries that are landlocked and still highly export orientated. Characteristics of the neighbour countries effect economic performance and transport costs through their degree of accessible and attractive markets (Collier & Gunning, 1999, p. 15).

Access to sea of other water ways could contributes to economic development by lowering transport costs of trade. This effect is the opposite of being landlocked and has a positive effect on economic performance. Both the possession of land close to waterways as the population living in this area matter for reducing or increasing transport costs. The possession of land nearby transportable waterways lowers transport costs just like a high population density in those areas lowers transport costs (Collier & Gunning, 1999, p. 8: Gallup, Sachs, & Mellinger, 1999, p. 184). Both factors contribute to the reduction of transport costs (Ahlfeld, Hemmer, & Lorenz, 2005, p. 7). This works the same way around, lacking those factors will increase transport costs and creates a situation with poor market integration (Collier & Gunning, 1999, pp. 8, 12-13).

The geographic location of a country also matters for its accessibility to core areas of the world economy. Lust like being landlocked, the possession of land in coastal areas of areas connected to the sea by rivers and the population density in those areas, the accessibility to areas of world trade also effects transport costs and economic performance. The accessibility to those core economic regions also helps to determine the transport costs (Gallup, Sachs, & Mellinger, 1999, p. 195).

When analysing world inequality through looking at the world map wealthy and poor countries seem to cluster to each other. Neighbour countries political and economic situation can lead to positive or negative spill overs in the home country (Ahlfeld, Hemmer, & Lorenz, 2005, p. 7). Economic growth and political stability could affect the neighbour countries, where political instability and economic recession will cause negative spill overs (Ahlfeld, Hemmer, & Lorenz, 2005, p. 7). Ahlfeld, et al (2005) do however provide a valid argument why this “neighbourhood-thesis” should only form an extent on the geography thesis, not a replacement:

"From this point of view, the location of a country relative to a regional pole of growth could indeed provide an explanation for its development progress but would not necessarily provide the fundamental reasons of growth which the geography-hypothesis tries to reclaim." (Ahlfeld, Hemmer, & Lorenz, 2005, p. 7)

As stated in the literature review climate conditions can directly affect economic performances through different factors. Prevalence of disease is for example highly bound to climate zones. Certain parts of the world are exposed to climate specific diseases. A high number of infectious diseases are endemic to tropical and subtropical climate zones (Gallup, Sachs, & Mellinger, 2001). Increasing the risk of getting infected with a certain deadly disease. Conditions in other climate zones do not fit their environmental needs and for this reason contain less risks of getting infected. Those diseases thrive through the, for them, beneficial climate. Mostly diseases in which the pathogen spends large periods of time outside the human host. External carriers transport the diseases from host to host. Examples are mosquitoes (carriers of malaria) and parasitic worms (carriers of helminthic infections) (Gallup, Sachs, & Mellinger, 2001). High prevalence of disease could alter the age structure of a country's population. This could negatively affect economic development (Gallup, Sachs, & Mellinger, 2001: Collier & Gunning, 1999, p. 8).

Agricultural productivity highly depends on climate conditions (Ahlfeld, Hemmer, & Lorenz, 2005). Some climate zones are generally more productive than others. Certain climate conditions even provide hostel conditions for livestock and agriculture (Collier & Gunning, 1999, p. 8). Higher rates of agricultural productivity provide better changes for economies to develop. Climate zones where crops tend to naturally flourish create better opportunities and boost per capita income (Gallup, Sachs, & Mellinger, 2001). Especially tropical climate zones are subject to many climate factors that negatively affect agricultural productivity (Ahlfeld, Hemmer, & Lorenz, 2005). Agricultural productivity in tropical climate zones can be 30 to 50 percent lower relative to temperate-zones (Gallup, Sachs, & Mellinger, 1999, p. 197). The proportion of agricultural activity in percentage of the total GDP is typically higher for low income countries (Collier & Gunning, 1999, p. 8). Combining this high dependence on agriculture with low agricultural productivity creates a very fragile situation. This situation could highly hinder economic development (Collier & Gunning, 1999, p. 8).

Theory describes the importance of both the prevalence of disease and agricultural productivity for economic performance. Both factors link climate to their ability of being productive. Tropical climates are subordinated to lower levels of economic productivity. This leads to the consideration of a direct effect of tropical regions to economic performance. This effect is however only because of prevalence of disease and agricultural productivity effect economic performance, and those variables are closely

linked to tropical regions (Gallup, Sachs, & Mellinger, 1999, p. 200: Ahlfeld, Hemmer, & Lorenz, 2005, p. 3).

3.4 Social & Human capital

Social and human capital, represented by the cultural hypothesis, focus on the main important of personal and community researches. Those factors are formed through maintaining social networks and increasing individual skills and value.

Causal mechanism

Social capital can be described as the framework of social relationships, creating resources for individuals (Knack & Keefer, 1997, p. 1251). Interpersonal trust and norms of civic cooperation are some of the indicators that could be used to measure social capital (Knack & Keefer, 1997, p. 1251). Those resources contribute to economic development through creating a platform with high incentives for economic activity (Knack & Keefer, 1997). Social capital can affect economic performance through the need for social interaction. A high degree of trust in other individuals of firms leads to lower transaction costs (Knack & Keefer, 1997, p. 1252: Van Schaik, 2002, p. 6). Every economic transaction is in some way related to trust. Lower transaction costs could in turn lead to a higher productivity (Knack & Keefer, 1997, p. 1253). The system also works the other way around. Low level of trust could discourage investments, innovation and productivity (Knack & Keefer, 1997, p. 1252). Economic transactions need to be better monitored to achieve the same level of certainty that is needed. Monitoring brings along costs. Money spend on monitoring cannot be spend on more economic efficient activities.

Social capital theory in a way uses the same argumentation as the institutions hypothesis. The difference is that social capital emphasizes the crucial role of trust in social relations, where the institutional hypothesis emphasizes the importance of institutions to create this degree of certainty. Societies that contain high levels of trust are less dependent on formal institutions for this certainty (Knack & Keefer, 1997, p. 1253).

Where social capital focusses on the relationship between actors, human capital focusses on the sum of the individual resources of the actors. Social and human capital are interconnected (Knack & Keefer, 1997, p. 1253). Human capital is the measurement of level of education, skill and wealth of individuals (Acemoglu, Gallego, & Robinson, 2014, p. 876). As mentioned in the literature review, education and economic performance are highly correlated. Previous research found significant positive relationships but lack the description of specific causal mechanisms of how schooling affect economic growth

(Glaeser, 1993). Theory describes the possible causal mechanisms underneath this relationship. Theoretically skilled workers possess more expertise what translates in a higher level of productivity. Barro and Lee (2012) also emphasize the importance of human capital, especially education, on economic progress. A high number of well-educated workers leads to a higher-level of labour productivity (Barro & Lee, 2012, p. 184). Education as human capital contributes to the absorption of technology (Barro, 2001). This again contributes to higher economic performance. An additional year of schooling provides more ability to build on human capital (Bils & Klenow, 2000). Reversed causality as described in the literature review could possibly affect the relationship between schooling and economic performance. Careful regression analysis emphasizing theory and well-considered choices in research design need to create a podium for analysing the effect human capital has on economic performance.

3.5 Summary theoretical assumptions

The theory described above and the literature review of the hypotheses creates the possibility to set expectations related to economic performance. Theory describes the positive relationship between the existence of political and inclusive economic institutions. To be more specific, it is described as the main cause of economic performance. Following this argumentation, I expect the economic performance of a county to increase as an effect of the existence of good institutions. It represents a positive relationship. Meaning that performance will increase for every increase of quality or number of institutions.

Macro-economic policy directly links governmental policy, and the ability to counter market failures and facilitate open markets, to economic performances. As reported by this hypothesis, macro-economic policy could affect economic performance by using openness of their domestic economic market as instrument. A higher degree of openness increases trade and economic performance. Following theory, I expect macro-economic policy linked to countering market failures and creating trade openness to be positively related to economic performances.

Theory related to elements of geographic location mainly focus on transport costs and their effect on economic performance. Access to waterways would lower transport costs and contribute to trade and economic performance while being landlocked increases transport costs. Following this argumentation, I expect a negative relationship between (factors that increase) transaction costs and economic performance. Those factors are determined by the geographical location of the country.

Theory related to climate conditions focus on both the prevalence of disease as agricultural productivity. Both element can be transformed into usable hypothesis. A high prevalence of disease

will be negatively related to economic performance and high agricultural productivity tends to positively affect economic performance. This assumption seems to be heavily applicable to tropical regions. Theory assumes that those tropical regions contain climates where diseases thrive and agricultural productivity is low. Those factors negatively relate to economic development.

Social and human capital theory stresses the positive effect of both elements on economic development. I expect indicators of those two factors to be positively correlated to economic performance in my sample. High degrees of social and human capital create additional value and tend to incentive for economic activity. Both elements are expected to cause a higher extend of economic performance.

Chapter 4: Research Design

Economic analyses have their deficiencies when preforming research. The field of study does not provide the possibilities of running controlled experiments because of the scale, time lag problem and social economic impact (Baumol, Litan, & Schramm, 2007, p. 42). Historic statistical data is used to point out relationships between independent and dependent variables. In this way, economic analyses are always historically aimed in order to provide theories for future use. This thesis form no exemption. The practical use of those theories relays on the assumption that economies will continue to perform in the same way as they did in the past (Baumol, Litan, & Schramm, 2007, pp. 42-43). Explicit, this means that theories formed through analysing historic economic data are used to interpret current economic performances. It makes scholars able to make a prediction about future events. The explanatory power of the models depends on their quality. By analysing quantitative historical data this thesis tries to point out the fundamental causes of economic performance on a cross-country level.

Data from different existing dataset is merged based on theoretical assumption provided by the literature. This merge of different datasets makes us able to analyse the four main hypotheses of economic development in one model. Creating a fair levelled playing field for testing the four different hypotheses. This research is based on existing hypotheses determining causes of economic performances, representing deductive research. Theoretical expectations are tested by analysing actual observations. Regression analyses will determine which of the hypotheses provides the most explanatory power for defining economic performance in our data sample, while controlling for all other factors. The size and capacity of this paper will inevitably limit the scope of the theories as originally elaborated by the auteurs. Nevertheless, the use of a newly established database provides the opportunity to objectively test the existing hypotheses.

This chapter elaborates on how the research is conducted. Starting off with describing the research approach. This includes the unit of analysis, the associated method of analysis, and the limitations of the picked research design. Afterwards I describe the dependent variable, the main independent variables and the additional control variables. Those parts describe the operationalisation of the key variables used in this study. The independent variables are described and classified through the elements described in theory. A summary of the variables can be found in Appendix Table B.

4.1 Research approach

This thesis uses quantitative analysis as research type. All statements made in theory about relationships between the independent variables and dependent variable can only be acknowledged when a statistical significant correlation is found in the sample (Babbie, 2012, p. 93). A correct time order between the cause and effect and nonspuriousness between the two are the other main criteria

for nomothetic causal relationships in social research (Babbie, 2012, p. 93). Based on those assumptions and theory, the upcoming part of this chapters sets the frame for nomothetic causality to be tested.

4.1.1 Unit of analysis

This thesis uses countries as the unit of analysis. The different datasets that are merged contain slightly different sample countries. All data provided in the datasets are aggregated to countries. Most countries match in all four datasets but some countries lack observations for some dataset specific variables. To be able to analyse the relationship between the independent variables and our key dependent variable, every observation needs to contain data about economic performance. Without this data, there is no possibility in linking the independent variables as plausible causes of economic performance. Geodata, the main dataset used, provides data on economic performance for 150 countries. Those countries were selected based on containing a population exceeding 1 million (Gallup, Sachs, & Mellinger, 1999, p. 198). Countries with a population below 1 million represent such a small percentage of the world population, and additionally could disproportional affect the outcome of the regressions that those countries are not included. Countries used in the database represent all continents and form a good representation of the total world population. Appendix Table A provides an overview of the specific countries part of the population sample. The same databases used in this merged dataset are widely used by other scholars in their own research (Acemoglu, Johnson, & Robinson, 2001: Easterly & Levine, 2002: Glaeser, Porta, Lopez-de-Silanes, & Shleifer, 2004).

There are two major reasons to support countries as measurement level. Countries are the most commonly used unit of analysis for global research. Other possible measurement levels could be local or regional scale. Urban cities could for example be compared with remote hinterlands. In this way, another geographic indicator would serve as global comparable measurement. However, most of the available data of economic performance is based on countries as measurement level. The second reason is that two thirds of the income difference of individuals on global scale can be explained through the variation in per capita GDP between countries (Ahlfeld, Hemmer, & Lorenz, 2005, p. 2). Within-country differences, like differences between individuals based on age, class or geography, only account for the approximately one third of the income inequality. Another practical argument related to the understanding that cross-country differences are crucial in explaining global world inequality is that policy to counter inequality, and especially extreme poverty, should be focused on increasing wealth at national level (Ahlfeld, Hemmer, & Lorenz, 2005).

4.1.2 Method of analysis

This research conducts a statistical cross-sectional analysis. Officially cross-sectional study involves observations at one point in time (Babbie, 2012, p. 105). Differences between the units of analysis are central when analysing the data. The variables contain measurement of different time periods, but are analysed by comparing those variables on one level. Based on theory the independent variables are expected to affect the main dependent variable. The relationships between the variables can all be analysed at the same time through conducting multiple regression analysis. Some variables vary in the year of observation and another represent the average of the total period they represent. The variable that measures openness for example contains the average of the degree of openness from the period 1965 till 1990. Here a longer time series is needed for the variable to reflex the intended operationalization. The specific year of observation for each variable will be elaborated upon in the upcoming pages. Country specific observations are conducted in the period from 1960 till 1995. The observation data of some of the variables vary because of the availability of data.

Another possible method of analysis is longitudinal study. Longitudinal studies vary in time and analyse the same phenomenon over a certain period (Babbie, 2012, p. 106). This study can be viewed as a cross-section study because the variation of time in the measurement of the variable is not the focus of regression. The variation in observations of the different countries is. As described this inquiry does not fit the term cross-section study perfectly. This creates a situation where the differences in time periods of the observations of the variable need to be considered. Keeping this in mind will prevent me from drawing wrong conclusions. The data gathering of our main independent variables is conducted before observing our dependent variable. The progress of time between the observations provides direction to a possible correlation. However, this does not mean that the progress of time does reflex a causal relationship between the variables.

4.1.3 Limitations

The lack of panel data for analyses excludes the possibility for checking reversed causality. There is reason to believe that some independent variables and the dependent variable are subject to complex causality. Possible correlations could work both ways. Those concepts need to be individually assessed by theory and the represented variables need to be individually assessed to control for possible reversed causality problems. Chapter 5 consist of the analyses of the variables described in this chapter and will elaborate more on the outcome of the regressions.

It is important to realize that none of the main schools will provide a hundred percent accurate explanation of the difference in cross-country economic performances. This is partly due to the massive number of unobserved variables in cross-country regression analyses that determine a

country's economic development. This thesis will therefore not be able to provide an all including answer to what causes cross-country economic performance. The economy as a hole, and differences in economic performances specifically, are affected by a vast number of element. It is practically impossible to account for all the factors influencing economic performance.

The variables representing the different hypotheses could be internally linked. Theory already mentioned the relationships between the different hypotheses. There is also a possibility that the variables measure the same concepts, but segregate the underlying causal element underneath into different parts. The focus of what causes economic performance varies but the measurements could indicate the same thing. A to high degree of interaction between the independent variables could cause multicollinearity. This could affect the explanatory power of the main variables.

4.2 Dependent variable: Economic performance

Gross Domestic Product (GDP) is one of the most general accepted measurements of the domestic market economy (Stockhammer, Hochreiter, Obermayr, & Steiner, 1997). It measures the monetary value of all final goods and services produced in the year of interest in a country. The US-dollar (USD) is internationally the most common used currency to express GDP. Easily put, it measures the added value of total production in a specified period. It provides an absolute number of economic performance indicated by a certain currency.

There are however additional indicators of economic performance. The Index of Sustainable Economic Welfare (ISEW) is one of those other indicators. This measurement counters the biggest critique of GDP as measurement of economic activity. The main point of critique is that GDP just measures the economic production of the total market, but fails to measure wealth (Costanza, 2009, p. 4). ISEW also takes unpaid household labour, social costs, environmental damage and income distribution into account (Stockhammer, Hochreiter, Obermayr, & Steiner, 1997). By including those variables, this index extents the measurement of economic performances and converts it into economic welfare (Stockhammer, Hochreiter, Obermayr, & Steiner, 1997). This changes the focus of observation to a measurement of welfare. This indicator is very useful in searching for the welfare development of countries. This different purpose of measurement need additional indicators. Proponents of using the ISEW as indicator argue that growth of countries beyond a certain size leads to additional costs of growth that exceed the benefits (Lawn, 2003, p. 105). Those additional costs will only be integrated into the model when the ISEW is used instead of GDP.

However, this thesis searches for the determiners of economic performances, not economic welfare. GDP is a measurement of economic activity, not economic well-being (Costanza, Hart, Talberth, &

Posner, 2009, p. 4). For this reason, the GDP measurement better reflex the factor I want to measure. GDP is certainly not an ideal measurement of wealth. It measured income, but it ignores equality, social norms and environment. Especially that characteristics of GDP in relationship with ISEW makes this indicator suitable as indicator for our dependent variable. It only measures the economic performance.

The indicator GDP as measurement of economic performance can be modified to make it more workable. The level of GDP per capita for example provides a better way of comparing cross-country levels of economic performance compared to the real level of GDP. It divides the total GDP through the total population. In this way, GDP is relativized to one single citizen. The additional including of the purchasing power parity (PPP) relates the GDP per capita to the costs of living in each country. This creates a cross-country comparable level of economic performance.

The level of GDP per capita (PPP) is naturally a product of past economic growth (Knack & Keefer, 1995, p. 215). Using log GDP per capita (PPP) provides a solution for a few of the problems faced when using the normal level of GDP per capita. Using (natural) logarithm as a mathematical function rescales the indicator to make it easier comparable across countries. The same data is used, but just described in a different way. This way of describing data is much more meaningful and robust way of describing the data. Most scholars recognize the superiority of this measurement and use it in their analysis to compare the cross-country economic performances (Acemoglu, Johnson, & Robinson, 2001: Sachs & Warner, 1997: Easterly & Levine, 2003). The natural logarithm of GDP per capita makes the data easier accessible. This level of measurement is the best indicator of economic performances combined in one number that existing economic literature offers.

The specific data I use as indicator of economic performance originates out of the geodata dataset provided by Gallup, Sachs and Mellinger (1999). The GDP per capita (PPP) data is originally assembled by the World Bank (1998). I take the logarithm of the initial GDP per capita (PPP) to rescale the variable.

4.3 Independent variables

This thesis provides the framework for testing four main hypotheses for explaining the difference in global economic performances. Those four hypotheses all represent a large school of thought in the literature of economic development. In this section the indicators that capture the main ideas of the different schools are described and related to this thesis specific research question and dataset. A careful process of conceptualization, through translating the abstract terms to systematized concepts, and the process of operationalization, creating clear measurable indicators for those systematized concepts, helps clarifying this inquiry. It sets the research direction, structure and boundaries of this thesis.

Some of the independent variables that affect economic performance will take time before their effect is translated into growth of the GDP per capita. This creates an inevitable time lag problem. The laps of time will affect the relationship between the causes and effects, this will bias my estimates of the 'true effect' between the variables. But it does not directly imply a correlation between the variables. Only through analysing the data with the use of theory, possible interconnectors can be defined. Additionally, any possible correlation will inevitably lead to biased results in measuring the causes of economic growth. To partly counter this effect, the mean independent variables will be analysed in such a way that the observations are conducted before the observation of the dependent variable. In this way correlation is possible. Reasons to assume correlations are mentioned in theory.

Upcoming pages will start off with discussing the methodology for testing the institution hypothesis. It describes the process and my considerations for choosing different indicators as measurement for reflecting the hypothesis. Extending on the research design of the geographical hypothesis, policy hypothesis and cultural hypothesis will be the follow up of this section. The total number of variables is of course only a limited representation of the massive number of variables proposed in economic development literature as determinants of cross-country economic performance. The selection of the variables will be discussed below. A summary of the data descriptions and sources of the variables can be found in Appendix Table B.

4.3.1 Institutions

As briefly described in the literature review and theory, the characteristics of institutions makes the variable hard to measure. The variable 'institutions' is a complex concept and lacks direct indicators. Through the lack of direct data of institutions authors use different measurements as indicators (Knack & Keefer, 1995, pp. 207-208). Scientist have failed to transform North's description of institutions into indictors of institutions (Knack & Keefer, 1995, p. 207).

Sachs and Warner (1997) provide a general measure of institutional quality. Sachs and Warner follow the same measurement indicators as use by Knack and Keefer (1995). The institutional quality index consists of the average of 5 sub-indexes of the International Country Risk Guide (ICRG) and is developed by the Political Risk Services Group (PRS Group) (Sachs & Warner, 1997, p. 342) (Knack & Keefer, 1997) (PRSGroup, 2017):

- *The rule of law index*: "reflects the degree to which the citizens of a country are willing to accept the established institutions to make and implement laws and adjudicate disputes"

- *The bureaucratic quality index*: measures "the autonomy from political pressure", and "strength and expertise to govern without drastic changes in policy or interruptions in government services."
- *The corruption in government index*: measures whether "illegal payments are generally expected throughout lower levels government", in the form of "bribes connected with import and export licenses, exchange controls, tax assessments, police protection, or loans."
- *The risk of expropriation index*: measures risk of "outright confiscation" or "forced nationalization."
- *The government repudiation of contracts index*: measures the "risk of a modification in a contract taking the form of a repudiation, postponement or scaling down."[2]

Those elements are all deformed into a 10-point scale. Then they take the average score. High scores indicate high level of institutional quality. See the PRS Group's web page for extended methodology (PRSGroup, 2017).

The PRS Group is a private organization that measures quantitative risks around the globe. This thesis is limited to the freely accessible data of the PRS Group data published as part of the database provided by scholars. This unfortunately cuts of the opportunity of using different years of institutional quality index measurement conducted by the PRS. However, this does not from a critical barrier for this inquiry. Theory describes political and economic institutions as the structural framework that shape the incentive structure, and mentions that they are highly path dependent and change resistant. Those characteristics makes it reasonable to expect institutions to roughly stay remain on the same level in the related period. Following this argumentation, the institutions index measured for one year will be highly like the institution index in the nearby years.

The ICRG uses different sub-indexes to construct one clear cross-country comparable indicator. Combining the different sub-indexes both the quality of political and economic institutions, as described in the theory, are measured. In the field of economic institutions, the focus lies on extractive economic institutions. The index measures those economic institutions that provide direct incentives for economic activities and in this way, contribute to economic development. This index fits the criteria set in theory.

Knack and Keefer (1995) provide a more precise description of the composition of the institutional quality index (Knack & Keefer, 1995, pp. 225-226). Next to the ICRG they also add the Business Environmental Risk Intelligence (BERI) database, which consist of 4 sub-indexes: contract

[2] All quotes originate from the description of Knack & Keefer (Knack & Keefer, 1995, pp. 225-226).

enforceability, infrastructure quality, nationalization potential, and bureaucratic delays (Knack & Keefer, 1995, pp. 211-212). Sadly, this database is nowhere to be found. This makes this data non-usable for regression analyses.

Both indexes contain the average of the underlying sub-indexes because of the strong correlation between those individual indicators (Knack & Keefer, 1995, p. 212). Knack and Keefer argue that the consequent risk of multicollinearity and avoidance of omitting indicators from the equation makes aggregation of the separate indicators needed (Knack & Keefer, 1995, p. 212). In their research, they use the first available data for each variable for each country (Knack & Keefer, 1995, p. 210). In this dataset, most ICRG data is from the year 1982 and smaller parts from the year 1984 and 1985.

Those scholars also provide another alternative for measuring institutions through analysing political violence with the Gastil indices (Knack & Keefer, 1995, p. 208). Those indicators focus on measuring political violence through analysing revolutions and coups, and assassinations (Knack & Keefer, 1995, p. 208). Sachs and Warner notice that those indicators however found to have a lower impact on economic development than the measurement of institutional quality as discussed earlier (Knack & Keefer, 1995, pp. 207, 217). Additionally, this indicator does not fit theory as good as the ICRG.

Some other choose to only use the risk of expropriation by government of the ICRG as measurement for institutions (Acemoglu, Johnson, & Robinson, 2001). Glaeser et al. (2004) however highly criticise the risk of expropriation by government, as well as governmental effectiveness, and constraints on executive as useful measures of institutions (Glaeser, Porta, Lopez-de-Silanes, & Shleifer, 2004, p. 273). They argue that the first two indicators do not measure institutions, but only reflex the past governmental policy on expropriation and governmental quality. Theory describes institutions as structural systems that constrain political, economic and social interactions. Glaeser et al. (2004) argue that this aspect is not covered in the described indicators. The third level of institution measurement they reject is constraint of the executive (Glaeser, Porta, Lopez-de-Silanes, & Shleifer, 2004, p. 274). In their view this measurement is not an indicator of institutions, but as a measurement reflexing the outcome of most recent elections. They argue that the ICRG dataset include subjective assessments and measures other dimensions than just institutions. In this way, it measures the events affecting the political domestic situation, and not the 'rules of the game' as indicated by institutions (Glaeser, Porta, Lopez-de-Silanes, & Shleifer, 2004, p. 276). Although most auteurs recognize the subjective side of part of the ICRG data, they also recognize the involvement of local experts and willingness of customers to pay substantial fees for this information as testament for the validity of the data (Barro, 1996, p. 19).

The indicator used by Acemoglu et al. (2001) only measures a part of the influence of institutions. Where those auteurs focus their view of institutions only on this indicator of protection against

expropriation, I choose belief that the total available data of institutional quality as measured by the ICRG provides a more complete measurement of institutions. I recognize the concerns of Glaeser et al. (2004) related to the problems with using the ICRG dataset as measurement for institutions. But I do not agree that the index provides a misrepresented image of institutions. The index combines 5 sub indicators that all represent a part of the measurement of institutions. The total package fits the theoretical identification of institutions in the best way possible. Knack and Keefer (1995) notice that the institutional quality index provides a detailed rating of institutions on different dimensions that better fits the description of property rights and closely related to institutions as emphasized by North (1990) compared to the BERI data and Gastil indices data (Knack & Keefer, 1995, p. 210). The measurement of institutional quality includes different measurement levels of institutions and better fits the framework for measuring institutions as described in the theory. This argumentation combined with the realization that the original dataset of Knack and Keefer (1995) is nowhere to be found leaves me with the dataset of Sachs and Warner (1995).

4.3.2 Macro-economic policy

As mentioned in the literature review and theory trade and income are highly interwoven (Frankel & Romer, 1999, p. 379). Trade cannot be used as a variable for measuring macro-economic policy for two clear reasons. Governments can highly influence trade through their policy, but they do not directly determine the amount of trade. Another argument is that trade and economic performance are correlated in such a way that determining the causation is almost impossible (Frankel & Romer, 1999). This complicated connection and the lack of coverage of trade as indicator for macro-economic policy clarifies that just measuring trade is not sufficient. Measuring trade policy also poses difficult questions (Harrison, 1996, p. 420). The link between macro-economy policy and trade is also opaque. For this reason, I need another indicator of governmental policy.

Sachs and Warner (1995) can up with an indicator suitable for measuring macro-economic policy. As described in the theory, openness is one of the major indicators of macro-economic policy. Openness is highly related to trade but on the same time also highly related to governmental policy. A domestic government can directly affect the openness to international trade of her economy by changing their policy. Sachs and Warner (1995) provide a useful framework in measuring openness of a country's economy. They use the variable openness to measure the degree of openness to international trade of the economy of a country. This variable measures the proportion of years an economy is open to international trade during the period 1965 till 1990. In this way, this variable counters the problem of short-term vision. By taking the mean of openness of this period the short-term vision of all the governments that where in power is translated into a long-term average.

The following 5 criteria must be satisfied for an economy to be classified as open (Sachs & Warner, 1997, p. 339):

1. An average tariff rates that does not exceed 40%
2. An average quota and licensing coverage of import below 40%
3. A black-market exchange rate premium below 20%
4. No extreme controls on exports
5. Not considered as a socialist country per the standard of Kornai (1999)[3]

Combining the 5 criteria and analyse the degree in which countries meet the criteria over a set timeframe (1965-1990) provides a good idea to what extend governments use (efficient) macro-economic policies to intervene in the market. A high degree of openness relates to a low level of governmental intervention. Following up theory, openness is accepted to be positively related to economic performance.

The variable openness accurately represents the policy hypothesis. It includes different measurement levels of macro-economic policy that could affect economic performance. I choose to use only the variable openness to reflect the policy hypothesis because of its main importance in theory and because of the measurement errors of other variables. I did not find any suitable variables related to market failures. Market failures do not provide suitable indicators to form an own variable. Openness can however also be viewed as result of effective policy against market failures. A country's inflation rate and national saving rate could be linked to macro-economic policy (Sachs & Warner, 1997, pp. 3-4). The direct relationship with governmental policy is however hard to determine, and governmental policy is very country specific.

4.3.3 Geographical factors

The literature review of geographical factors already provides a clear image which variables could be used. As described different auteurs who support this hypothesis recognize the same kind of patterns. The time scope and scale of this thesis creates boundaries for the specification of all the hypotheses, and requires me to choose certain general indicators for summarizing the general idea of the hypothesis. As already described in the theory, Gallup et al. (2005) provide a useful view how to measure possible correlations. Fitting the two sub categories of the geographic hypothesis as described in theory.

[3] Detailed description can be found in Sachs and Warner (1995), p. 65-66

Geographic location

I will first describe the variables that represent the geographic location. Gallup et al. (1999) provide useful indicators in their data set for analysing the direct effects of geographic location on economic development. This category has different indicators because the effects of the geographic location cannot be described by using only one of two indicators. This wide range of indicators creates the possibility to analyse the relationship to economic performance with great focus, but it lacks one general indicator that includes all the crucial measurement elements. Theory describes that the geographic location mainly affects economic performance through transport costs. Following this argumentation, the variables used are linked to transport costs.

The variable 'percentage of land within 100 km of the coast or river' will be used to measure the effect of geographic location based on assumptions regarding the effect of transportation costs on economic performance. The variable 'percentage of population within 100 km of the coast or river' is closely linked to this variable. Both measure the degree of to which the country possesses easy accessible waterways. Theory links both the possession of land and the population density in those areas to declining transport costs. The hypothesis stats that those costs negatively affect economic performance.

The variable 'percentage of land within 100 km of the coast or river' is the proportion of land fitting that criteria measured by dividing this area through the total country's territory. This includes coastal areas and ocean-navigable rivers. Coastline in the Artic and sub-Artic region above the winter extent of sea ice and connected rivers are excluded. Those waterways are not usable for trade through water transport and the surrounding area does not profit from lower transportation costs. The variable 'percentage of population within 100 km of the coast or river' is measured in the same way but uses the county population distribution in 1994 as data. Both variables are measured on a scale from 0 till 100. 0 represents the situation where zero percentage of the country's population or land is within 100 km of the coast of river. 100 represent the situation where the total country's territory is within 100 km of the coast or rivers.

Data for those variables originate from the geodata dataset (Gallup, Sachs, & Mellinger, 1999). Gallup et al. (1999) used descriptions of Rand McNally and Co. (1980), Britannica Online (1998), and Encyclopaedia Encarta (1998) for classifying rivers as ocean navigable. Calculations for the coastlines and ocean-navigable rivers are based on digital coastlines and rivers reported by the Environmental Systems Research Institute (ESRI) (1996). The data used for measuring population distribution is based on the year 1994 and originates from detailed geographic information system of the world population data set as described by Tobler et al. (1995). All those descriptions are described in the article of Gallup,

et al. (1999). Specific references and more in-depth information related to the classification of this variable can be found in its original source and the linked descriptions (Gallup, Sachs, & Mellinger, 1999, p. 216).

A variable linked to both the previous variables as transport costs theory is landlocked. This variable will be included in the regression analyses as a dummy to test for the effect of being surrounded by neighbour countries. The country is assigned a 1 when the county has no access to the coast and is assigned a 0 when the country is directly connected to the sea. One problem concerning this variable is that it does not differentiate between different kinds of landlocked countries. As described in theory being landlocked is not a determining for having relative low economic performance. There are landlocked countries, for instance in Western Europe, that highly exceed the average country's economic performance. This could be due to cluster-effect as described by the neighbourhood-thesis. As already argue including this effect does not provide answers to the search for fundamental reasons for differences in economic performance. If I would exclude Western Europe landlocked countries the explanatory power of the variable, and the geographical hypothesis would largely decline.

Acknowledging the importance of transport costs for economic performance, the last variable that is included as indicators of the effects of geographic location is the CIF/ FOB shipping costs ratio. The abbreviation FOB stands for 'free on board' and measures the total costs of the transported good or services till shipment. This does not include insurance and freight and this value is usually declared by the exporter (Gaulier, Mirza, Turban, & Zignago, 2008, p. 2). The abbreviation CIF stands for 'costs of insurance and freight' and includes those costs to the total value. The ratio between those values provides the CIF/ FOB ratio (Gaulier, Mirza, Turban, & Zignago, 2008, p. 2). All data is on information of the year 1995. Both measures are reported by the International Monetary Fund (IMF) (Gallup, Sachs, & Mellinger, 1999, p. 194). This data is provided data in the geodata dataset (Gallup, Sachs, & Mellinger, 1999). Gallup et al used this data as described by Radelet and Sachs (1998), who directly used data from the IMF. Just as the variables percentage of land and population within 100 km of the coast or river additional specifications concerning the data of this variables can be found in its original source (Gallup, Sachs, & Mellinger, 1999).

This variable provides cross-country comparable ratio of additional transport costs. Because the value including costs of insurance and freight is divided by the costs excluding those factors the value of each observation will exceed 1.0. Additional transport costs are not expected to exceed the f.o.b. costs with enormous amounts, for this reason the CIF/ FOB ratio just tops 1.0, but stays close to this number.

This variable has the disadvantage that not every country possess data for CIF and FOB to calculate the ratio. Furthermore, the IMF only makes rough estimates of those two measurements using county

specific import data (Gallup, Sachs, & Mellinger, 1999, p. 195). In this way, the variable lacks standardization in measurement. Nevertheless, the CIF/ FOB ratio is widely acknowledged and used as estimators of transport costs for and international trade (Gaulier, Mirza, Turban, & Zignago, 2008, p. 2). This indicator does mainly measures transport costs, but it does not directly link those additional costs to geographic location factors. This means that this variable is not directly linked to the geographic. It relates to the transport costs and its relationship with geographic location. Its additional value for checking the relationship between transaction costs and economic performance is of such importance for linking geographic location factors to economic performance that it is included in this section.

Climate conditions

Climate conditions are divided in two sub categories: prevalence of disease and agricultural productivity. Theory provides the assumption that a high disease burden and low agricultural productivity cause lower economic development. Both elements are based on the worse climate conditions of tropical regions.

The variable that will be used as measurement of the prevalence of disease is the falciparum malaria index 1994. It measures the percentage of a country's land area subject to falciparum malaria. Falciparum malaria is the deadly version of malaria (Sachs, 2005, p. 196). Falciparum malaria is an infectious disease that transported through mosquitoes who carry malaria parasites. Malaria has been relatively under control since 1964 in areas where the distributing mosquitoes where fragile (Gallup, Sachs, & Mellinger, 1999, p. 196). The habitat of those mosquitoes is now limited to tropical regions. The scale of this measurement is 0-1. Zero means no risk of falciparum malaria and 1 is 100 percent risk of falciparum malaria. Malaria is used as indicator because it is one of the most common tropical diseases and the pattern of malaria risk fits a wide range of other infectious diseases where the factor of transmission depends on the tropical climate (Gallup, Sachs, & Mellinger, 1999: Ahlfeld, Hemmer, & Lorenz, 2005: Acemoglu, Johnson, & Robinson, 2001). Other auteurs like Acemoglu use this same measurement.

The falciparum malaria index 1994 variable originates from the geodata dataset (Gallup, Sachs, & Mellinger, 1999). They constructed this variable using a global map of the intensity of malaria, and falciparum malaria in particular, in 1994. The global malaria map was provided by the World Health Organization (WHO, 1997). The parentage of falciparum malaria originates from 1990 (Gallup, Sachs, & Mellinger, 1999, p. 217).

This variable provides an easy accessible measurement of the prevalence of disease. It is however only related to infectious disease in tropical regions and excludes all other diseases that could affect economic performance through their relation with climate conditions. Like other variables the malaria index variable is subject to reversed causality. Malaria could contribute to poverty, but poverty could also cause higher malaria incidence (Sachs, 2005, p. 197). Careful analyses are needed to control for this possible two-sided effect.

One factors that is directly linked to agricultural productivity is soil quality (Collier & Gunning, 1999, p. 8). The Agric.data dataset, based on literature from Matthews (1983) and data from FAO (1995), provides a useful measurement for the soil suitability (PortlandStateUniversity, 2017). This database includes a variable that measures the percentage of very suitable soil for each of the six rainfed crops. This provide a clear general view of a country's agricultural productivity completely based on geographic climate factors. This dataset provide data for 164 countries. The variable represents percentage of very suitable soil on a scale from 0 till 100. A high outcome number represents a high degree of very suitable soil. The variable is the mean of the estimate of the percentage of very suitable soil for the six rainfed crops and is stated for each country (PortlandStateUniversity, 2017). Consult the original sources for more in-depth information about the classification of the variable. Soil suitability can be viewed as a general indicator of agricultural productivity. Easy put, this indicator provides an indicator of the very suitable soil for agricultural production for each country measured in percentage of the total country's territory. This measurement is exogenous from other variables because it results totally from the geographic factors. This variable has the advantage that, though to its exogenous characters, it is detached from the effects of reversed causality (Easterly & Levine, 2002, p. 15). Assuming that soil quality tends to stay roughly the same over time.

The disadvantage of this variable is that it only measures the possible agricultural production based on soil quality. It does not consider the actual production of agricultural goods. This is however very useful for the variable surfing as indicator of the geographical hypothesis. This variable can measure the effect of agricultural productivity detachment from its cohesion with other factors. Possible cohesion could influence the effect of agricultural productivity on economic performance I am trying to measure. Another disadvantage of this variable is that it represents an estimate of the soil quality for the whole country based on a limited amount of rainfed crops. This is only a narrow indicator of the measurement of soil quality.

Next to soil suitability I will also directly test for the effect of tropical regions. Both the percentage of land in geographical tropics and the percentage of the total population living in geographical tropics are included as variables. The geographic tropics is the area between the Tropic of Cancer (23° 45¢

north latitude) and the Tropic of Capricorn (23° 45¢ south latitude). Covering the latitudes where the sun is directly overhead at least once a year (Gallup, Sachs, & Mellinger, 1999, p. 218). Those variables originate from the geodata dataset and indicate the percentage on a scale from 0 till 1 (Gallup, Sachs, & Mellinger, 1999, p. 216). Low numbers indicate low percentages and vice versa, high number high percentages. The same dataset was used to construct the proportion of a county's land area within the geographic tropics as was used for constructing the percentage of land within 100 km of the coast or river (ESRI, 1996: Gallup, Sachs, & Mellinger, 1999, p. 216). The percentage of total population living in geographical tropics uses the population distribution data of 1994 (Gallup, Sachs, & Mellinger, 1999).

If theory correctly portrays reality the only effect of tropical regions on economic performance would be through the effect prevalence of disease and agricultural productivity have on economic performance. Including and excluding those different variables in regression analyses will help determine the internal relationships between the variables.

4.3.4 Social and human capital

Social capital

There is not one all including reliable measurement for presenting social capital (Putman, 2000, p. 14). As described in theory, Knack and Keefer (1997) provide a good framework for analysing social capital. They use data of the World Value Surveys conducted two time periods: one in the year 1981 and one in 1990-1999. Every survey wave was conducted on thousands of respondents (Knack & Keefer, 1997, p. 1255). They combined the two survey waves and created two classes of indicators (Knack & Keefer, 1997, p. 1255). Knack and Keefer recognize the existence of possible biased survey indicators, formed by translation difficulties, sampling error, and response bias. But they emphasize that their data values produce data consistent with data from independent sources (Knack & Keefer, 1997, p. 1255).

Knack and Keefer use to different indicators to measure social capital: interpersonal trust and norms of civic cooperation. Interpersonal trust is measured as the percentage of respondents that answers 'most people can be trusted' to the following survey question: "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" (Knack & Keefer, 1997, p. 1256). The other possible reply was: 'can't be too careful' (Van Schaik, 2002, p. 9).

The other variable represents the strength of norms of civic cooperation (Knack & Keefer, 1997, p. 1256). Respondents were asked to answer 5 statements about civil norms with a number in the scale 1-10. The number one represents the answer that this action is 'never justifiable', the number ten represents the answer 'always justifiable' (Knack & Keefer, 1997, p. 1256). They then flip the scale so

that large number represent a high degree of cooperation. The outcomes of the different statements are combined and form a 50-point scale.

Human capital

Human capital can be viewed as an extent of the cultural hypothesis. As discussed in the theory, human capital is closely linked to institutions. This variable is however also included as a separate measurement as part of the cultural hypothesis to test for a direct causal relationship with economic performance. Theory stresses the main importance of education as indicator of human capital.

Data used by Glaeser et al. (2004) does not seem to be public. Luckily I found an even better dataset provided by Barro and Lee (2013). Barro and Lee update their original data on educational attainment in the world. Previous data on educational attainment in the world was measured by school enrolment ratios. This data does however not adequality represents the amount of human capital (Barro and Lee, 2013, p. 184). The new dataset uses additional data, a better methodology and in this way, provides a better indicator for the level of human capital (Barro and Lee, 2013, p. 184). This dataset provides a wide range of categories of educational levels for different subclasses. Barro (1996) in his previous work points out the high level of explanatory power of years of schooling for the population of age 25 and older (Barro, 1996, pp. 15). Both male and female are included.

Gallup, et al. (1999) use the same indicator in their research, but their data consist of the old measurement data conducted by Barro and Lee in 1993 (Gallup, Sachs, & Mellinger, 1999, p. 217). To stay close to the indicator used by Gallup et al. (1999) I use the same indicator, but extracted from the newly conducted data by Barro and Lee (2013). This new data provides by Barro and Lee (2013) better reflect the levels of education in the world. For this reason, this new data replaces the data used in the geodata dataset used by Gallup, et al. (1999). Corresponding to those scholars, I will also use the average years of schooling in 1965 for the age 25 years and older as variable for human capital. This variable indicates the initial level of schooling. I choose to include this year, and not more recent years because average years of schooling will inevitably be highly correlated with economic performance. Analysing this variable lets me check for the importance of initial level of schooling on economic performance. By using this year at least the most extreme reversed causality is out of the picture. The level of economic performance of measured in 1995 is not able to affect the average level of education in the year 1965. Following theory, I expect this variable to be positively correlated with economic performance. Gallup et al (1999) find that the initial level of schooling positively and significantly affect the growth of GDP between 1965 and 1990 (Gallup, Sachs, & Mellinger, 1999, pp. 201-202).

4.4 Additional control variables

Geographic regions

Additional control variables will be included using geographic regions. Geographic regions are used for classifying countries to regions. Those regions form blocks of countries sharing the same geographic region, combined with certain additional characteristics. Those regions serve as dummy variables and every county is part of only one region. The specific geographic region is coded 1 when the country is part of that region, and is coded 0 when the country is not. The following geographic regions are subdivided in the master database: Western Europe (eu); Sub-Saharan Africa (safri); South Asia (sasia); Transition countries (Former Soviet Union & Eastern Europe) (transit); Latin America & Caribbean (latam); and East and Southeast-Asia (eseasia).[4]

Geographic regions represent specific countries based on, among other things, geographical factors and cultural and political characteristics. Those variables do however not capture the main geographical factors as pointed out in theory. For this reason, they are included as control variables and not as separate elements of the geography hypothesis. Those variables could be used to check if being part of a certain geographical region directly affect economic performance. Previous studies have pointed out that the Sub-Saharan Africa dummy is negatively and significantly correlated with economic performance (Sachs & Warner, 1997, p. 17). Acknowledging those studies, regression analyses should include the direct effect of this dummy. The same idea goes for the effect of Western Europe. Both elements will be included in the final regressions.

Participation in external war

Regression analyses will also include an external war dummy. This dummy has a value of 1 when the country has participated in at least one external war in the period of 1960 till 1985 (Gallup, Sachs, & Mellinger, 1999, p. 218). This variable originates of the 'Data set for a panel of 138 countries' constructed by Barro and Lee (1994). The justification of this control variable will be described in the analysis section.

[4] See Appendix Table A for specific classification of countries to geographic regions

Chapter 5: Regression Analysis

The first section of this chapter will be dedicated to describing the dataset. The main statistics of the variables are presented and dummies of geographic regions are used to provide an image of the distribution of the variables between the observations. Afterwards simple OLS regressions are described to check if the existing theory fits the data observations. Both methods help me check for outliers, possible linearity assumptions and degree of missingness of the variables. Thirdly the interconnectedness between the independent variables is analysed. Lastly, multiple regression analysis are conducted to test for the influence of multiple independent variables while including additional controls. This method is commonly used by social researchers when multiple independent variables are expected to simultaneously affect the given dependent variable (Babbie, 2012, p. 467). I anticipate that this method will increase the explanatory power of the model and improves its goodness of fit.

Stata (version 13.1) was used to analyse the dataset and test for possible correlations. Ordinary Least Squares (OLS) regressions form the basics four the regression models. Stata serves as main program for analysing the data, but additionally Microsoft Word and Excel are used to translate the results of the analyses into workable tables and figures. All statistical regressions tests are run based on the 95% confidence level. This means results are viewed as significant results when the p-value is smaller than 0.05. Every regression table will provide additional information about the specific degree of statistical significance.

This thesis is conducted to search for the fundamental causes of economic performance. For this reason, both the significance of the results as the explanatory power of the independent variables are of great importance. The statistical explanatory power of the variable in the model is presented by the R-square. Additionally, the adjusted R-square will be presented in multiple regression analyses. The R-square will keep on increasing when adding additional variable to the model. The adjusted R-square however considers what kind of variables you are adding. This makes sure that only the data that truly ads to explaining the outcome variable increases the adjusted R-square. Scholars active in economic development research differ in their use of explanatory power indicators. Some scholars chose to include R-square (Gallup, Sachs, & Mellinger, 1999: Glaeser, Porta, Lopez-de-Silanes, & Shleifer, 2004) and others use adjusted R-square (Knack & Keefer, 1997: Sachs & Warner, 1997). For this reason, both indicators will be included. R-square will serve as mean indicator, but the adjusted R-square will be elaborated upon when the regression present noteworthy results.

5.1 Descriptive analysis

Descriptive statistics are needed to provide a general view of the relationships between the variables and inform about the total amount of observations for each variable. This basic descriptive data is helpful in pointing out which variables should be used for regression analyses. My original research design serves as the red line for the regression analyses. Working on this thesis however created a situation where running regressions provided additional new information about the variables that is used to updating the research design. The current research design naturally covers all relevant aspects for conducting the analyses. Interaction between different parts of this thesis created a situation where every aspect of feedback was integrated in the related parts of the research. Additionally, analysing different element of the dataset provided in-depth information about the relationship between the variables and observations. In this way collecting, observing and analysing the data provided feedback for the original interests, research design and theory (Babbie, 2012, p. 116). This research process contained of a great number of interrelated inquiry cycles before the results can to be. Not every least step of this process will be described in this thesis. All main elements that determined the crucial choices made in the process of conduction this thesis are described in one of its chapters. All descriptions serve the goals of providing reliable, valid and useful information conducting the inquiry.

5.1.1 Description variables

I start with providing the descriptive statistics for the main variables. This provides a first overview of how the variables are related to the different hypotheses. The specific description and scale of the variables are described in the research design[5]. The descriptive data as described in Table 1 provides an overview of the number of observations, mean, standard deviation, minimum and maximum in this order for the following variables: Log GDP per capita (PPP) 1995 (loggdp95); Institutional quality index (icrg82); Openness to international trade (open6590); Percentage of land within 100 km of the coast or river (lnd100cr); Percentage of population within 100 km of the coast or river (pop100cr); Landlock dummy (landlock); CIF/ FOB shipping costs ratio (ciffoc95); Falciparum malaria index 1994 (malfal94); Agricultural soil quality (soilsui1); Percentage of land in geographical tropics (tropicar); Percentage of total population living in geographical tropics (troppop); Interpersonal trust (trust); Strength of norms of civic cooperation (civic); and Average years of schooling (yr_sch). The abbreviations of the variables will additionally be used to refer to the variables for practical reasons.

[5] See Appendix Table B for summary variables: data descriptions and sources

The total number of observation in the database is 150 countries. The original merged database consisted of 181 countries, but for 31 of those observations the gdp95 variable, and for this reason the loggdp95 variable was missing. I dropped those countries from the dataset because they add no value to the database when they do not contain information about the main dependent variable. The dropped observations only contained data of one, or a combination of, the variable(s) soil1, trust, civic, and yr_sch.

As pointed out in Table 1, the variables trust and civic possess the least amount of observations. This is due to the small amount of observations of the data gathered by Knack and Keefer, published in their article (Knack & Keefer, 1997, p. 1285). Both variables represent the social capital part of the cultural hypothesis. The small amount of observations limits the usable data for running regression analyses. Ciffob95 is another variable that stands out due to the lower amount of observations. This relative lower amount of observations is due to the low amount of observation conducted in the Geodata gathered by Gallup et al (Gallup, Sachs, & Mellinger, 1999). The variable landlock stands out because it is a dummy variable. The minimum and maximum scores of loggdp95, icrg82, and yr_sch are far apart. Those variables also have additionally high standard deviations, indicating relative big differences in scores across countries.

Table 1.

Description variables

	Observations	Mean	Standard deviation	Min	Max
Log GDP per capita (PPP) 1995	150	8,13	1,15	6,14	10,22
Institutional quality index	98	5,68	2,26	2,27	9,98
Openness to international trade	140	0,25	0,40	0	1
Percentage of land within 100 km of the coast or river	150	45,93	37,60	0	100
Percentage of population within 100 km of the coast or river	150	54,84	37,22	0	100
Landlock dummy	150	0,23	0,42	0	1
C.i.f./ f.o.b. shipping costs ratio	83	1,12	0,10	1,01	1,67
Falciparum malaria index 1994	149	0,29	0,41	0	1
Agricultural soil quality	146	13,89	9,89	0,15	55,07
Percentage of land in geographical tropics	150	0,49	0,48	0	1
Percentage of total population living in geographical tropics	138	0,28	0,36	0	1
Interpersonal trust	28	35,63	14,23	6,70	61,20
Strength of norms of civic cooperation	28	39,37	2,02	34,55	42,65
Average years of schooling	127	6,43	3,05	0,65	12,69

Notes: consult Appendix Table B for details variables

5.1.2 Description population

Table 2 provides an overview of the distribution of the main variables based on geographical regions. I choose to describe the data in such a way because the data consists of too much observations (countries) to individually assess and still provide a clear overview of the data. Dividing them in pre-organized clusters does provide this clear image. Most countries of the master dataset are already

match with their specific geographic region. However, not every country in the merged database is part of one of those geographic regions. The division used is based on data from the Geodata (Gallup, Sachs, & Mellinger, 1999). The merging of the data lead to a situation the sum of the geographic regions does not fit the total number of observation in the database. For this reason, I created a new geographic region representing the Middle East (mideast). This variable consists of the broad definition of countries that are part of this region, also known as the Greater Middle East or the Modern Middle East. Stretching the region from Morocco till Iran. In the database, it consists of the following countries: Lebanon; Turkey; Iraq; Iran; Saudi Arabia; United Arab Emirates; Afghanistan; Egypt; Morocco; Yemen; Tunisia; Algeria; Kuwait; Libya Arab Jamahiriya; Jordan; Syrian Arab Rep.; Oman; and Israel. All those countries were part of none of the geographic subgroups before they were appointed to this subcategory. Additionally, I reassigned Korea Dem. People's Rep.; Myanmar; and Lao PDR to East and Southeast-Asia (eseasia). I added Bosnia and Herzegovina and Yugoslavia to Transition countries (Former Soviet Union & Eastern Europe) (transit).

By creating this new variable mideast for geographic region and redefining the geographic region for some of the observations, only 4 observations do not fit to any of the geographic regions. Those countries: Australia; United States, Canada, and New Zealand do not fit any geographic region. No additional geographic region dummy can be created to mutually those observations. They do not share the same geographical region, and for this reason are not part of any geographic region dummy. Those observations represent western developed countries. For this reason, it is safe to assume that those countries have a higher loggdp95 than average.

The column 'Whole World' represents the total data sample or 150 observations. However, not all the variables contain this total amount of observations. The rest of the table consists of the division of the different geographical regions. Western Europe stands out for its relative high loggdp95 and its related low standard deviation. As expected Sub-Saharan Africa has the lowest log GDP, closely followed by South Asia. Western Europe has higher values of: icrg82, open6590, lndr100cr, pop100cr, soilsui1, trust, civic, and yr_sch than the average. Sub-Saharan Africa scores lower values than the average for all those variables. Additionally, Sub-Saharan Africa has a higher value for the variables: landlock, ciffob95, tropicar, troppop, and malfal94 than average. Western Europe scores lower than average for all those variables.

When comparing those very general scores for those two geographic regions, all assumption of the variables made in theory fit the sample data. This first view of the data does nevertheless not provide any information about statistical significance and internal relationship between the variables. Creating a first image of the data is the only function of comparing the means of those regions. It is important

to emphasize that only the geographic hypothesis stresses the importance of geographic factors. The other hypotheses link economic performances to other causes. Using geographic regions as dummies however creates a visual image of the worldwide distribution of the variables.

Just as pointed out in Table 1 the variables trust and civic once again form an obstacle. Were most variables completely or almost match the total amount of observations for the geographic region, the observations for the variables trust and civic are extremely low. In this way, only a few countries, which observation does include those variables, reflex the average for the total group. This creates a twisted view of the group average. Latin America & Carribean (latam) for example has a total N of 22, but only has 4 observations for the variable trust and civic. This also goes for other geographic regions. For this reason, the data described in Table 2 does not provide a weighted average of the sub sample. This is due to the low number of observation for those variables as already described in the first part of the descriptive analyses. The same goes for the variable ciffo95, but to less extent.

But there seems to be an additional problem with the variables trust and civic. The total sample consists of 28 observations concerning both variables, as described in Table 1 and in the whole world category of Table 2. Table 2 points out that 15 of those observations (that is more than 50%) are linked to countries in Western Europe. 15 of the total 16 countries linked to the geographic region Western Europe possess data for those two variables. Geographic regions as Sub-Saharan Africa and Transition countries are highly under represented.

Altogether Table 2 provides a clear image of the inordinate distribution of those two variables between the observation. The high degree of missingness combined with the disproportionate distribution of the observations leaves me no choice but to exclude this variable of the regression analyses. Including those variables in the regression could distort the regression outcome. This is mainly because better preforming countries (Western Europe) are highly overrepresented in the observations of trust and civic and could for this reason almost singly determine the relationship between trust, civic and the main dependent variable. Due to the loss of the variables trust and civic the social capital part of the cultural hypothesis loses its indicators. Only the average years of schooling variable, representing human capital, is left. The effects of this exclusion will be discussed later in this chapter when the relationship between the cultural hypothesis and first data analysis is discussed.

The lower amount of observations representing the group average for other variables does not provide an insurmountable obstacle for regression analysis. This is the case for icgr82 and yr_sch. Western Europe countries generally does not seem miss observations for the different variables. A possible explanation for this could be that information gathering in countries in Western Europe is easier

compared to information gathering in other countries. For this reason, countries in other geographic regions miss out on some observations but Western Europe does not.

Of course, it is important to remember that due the small errors, some data described in Table 2 does not form a generalizable image of the total subgroup. But this is not harmful for the research however because upcoming regression analyses will not be linked to those dummy geographical regions. Countries will be analysed as individual observations which counter the error of generalizability of subgroups.

Table 2.

Descriptive statistics population

Variable	Whole world	N	Western Europe	N	Sub-Saharan Africa	N	South Asia	N	Transition countries (FSU & EE)	N	Latin America & Caribbean	N	East and Southeast Asia	N	Middle East	N
Log GDP per capita (PPP) 1995	8,134	150	9,847	16	7,157	41	7,245	5	7,944	28	8,387	22	8,377	16	8,445	18
	1,145		*0,197*		*0,780*		*0,345*		*0,739*		*0,626*		*1,244*		*0,940*	
Institutional quality index	5,683	98	8,946	16	4,537	31	4,234	4	6,151	2	4,393	21	6,414	12	4,515	8
	2,257		*1,192*		*1,197*		*1,237*		*1,687*		*1,328*		*2,065*		*0,877*	
Openness to international trade	0,249	140	0,998	16	0,063	41	0,046	5	0,003	28	0,163	21	0,521	15	0,165	10
	0,397		*0,010*		*0,174*		*0,103*		*0,010*		*0,238*		*0,492*		*0,307*	
Percentage of land within 100	45,933	150	72,813	16	21,488	41	45,800	5	49,464	28	61,227	22	65,938	16	37,833	18
km of the coast or river	*37,598*		*27,345*		*29,280*		*4,805*		*39,554*		*33,772*		*33,127*		*33,289*	
Percentage of population	54,840	150	85,000	16	28,415	41	48,800	5	49,607	28	69,591	22	76,500	16	55,278	18
within 100 km of the coast or	*37,219*		*16,379*		*31,369*		*4,749*		*39,510*		*30,158*		*28,826*		*34,943*	
Landlock dummy	0,233	150	0,125	16	0,341	41	0,200	5	0,500	28	0,091	22	0,063	16	0,056	18
	0,424		*0,342*		*0,480*		*0,447*		*0,509*		*0,294*		*0,250*		*0,236*	
C.i.f./ f.o.b. shipping costs ratio	1,124	83	1,049	15	1,195	31	1,097	5	1,083	3	1,106	7	1,098	7	1,102	11
	0,100		*0,032*		*0,128*		*0,028*		*0,067*		*0,035*		*0,028*		*0,025*	
Falciparum malaria index 1994	0,290	149	0,000	16	0,820	40	0,243	5	0,000	28	0,085	22	0,361	16	0,084	18
	0,410		*0,000*		*0,328*		*0,180*		*0,000*		*0,217*		*0,355*		*0,203*	
Agricultural soil quality	13,438	146	16,169	16	8,805	40	16,798	5	21,014	27	15,953	22	8,897	14	13,874	18
	9,763		*8,144*		*7,631*		*6,549*		*11,952*		*11,292*		*6,971*		*5,288*	
Percentage of land in	0,487	150	0,000	16	0,926	41	0,383	5	0,000	28	0,825	22	0,712	16	0,178	18
geographical tropics	*0,477*		*0,000*		*0,222*		*0,416*		*0,000*		*0,342*		*0,448*		*0,316*	
Percentage of total population	0,276	138	0,000	16	0,689	40	0,099	5	0,000	22	0,259	21	0,202	15	0,106	15
living in geographical tropics	*0,361*		*0,000*		*0,308*		*0,166*		*0,000*		*0,289*		*0,298*		*0,178*	
Interpersonal trust	35,632	28	40,287	15	26,700	2	34,300	1	0,000	0	18,525	4	39,400	2	10,000	1
	14,229		*13,155*		*5,374*				*0,000*		*8,752*		*1,980*			
Strength of norms of civic	39,368	28	39,508	15	38,090	2	42,650	1	0,000	0	37,108	4	40,715	2	42,430	1
cooperation	*2,018*		*1,735*		*1,556*				*0,000*		*2,048*		*1,520*			
Average years of schooling	3,349	125	6,165	16	1,252	30	1,439	5	5,617	19	3,255	22	2,588	14	1,612	15
	2,606		*1,789*		*1,016*		*1,333*		*1,921*		*1,346*		*1,690*		*1,909*	

Notes: first number is the mean. Italic is standard deviations. N is number of observations.

5.2 Basic Regressions

This part of the analyses links the theoretical assumptions made based on the four main hypotheses to the data. The independent variable(s) representing the different hypotheses will be regressed against the key dependent variable. It provides a first image of the internal relationships between the different variables before conducting multiple regression analyses including all hypotheses. An additional argumentation for running those basic regressions is the fact that not every country has data for every variable. Analysing the different hypotheses on their own provides a bigger sample for testing the relationship between the variables. The number of useable observation declines when all the different variables are included. This is since not every country has data for all variables, but only the countries that possess data for all the variables can be used in the final regression.

Keep in mind that the upcoming regression only analyse the relationship between the variables representing the hypothesis and economic performance. Variables representing the other hypotheses and additional controls must be added in the regression before any statements of a correlation between the variables can be clarified. Those analyses will be discussed in the final multiple regression. For this reason, the upcoming regressions mainly add value through checking the strength, direction, significance, and R-square of the relationships between the variables. The figures additionally added provide a graphical overview of the relationship between the two variables and can be used to spot outliers. Linear relationships are assumed using scatterplots and OLS regressions. Figures are however only included where the hypothesis is represented by one single independent variable.

Before conducting the first analyses I checked if the linearity assumption of the relationship between the key independent variables and the dependent variable fits the data. I individually run regressions and created scatterplots using different independent variables and the dependent variable to check for linearity. I found that data fits the linearity assumption.

5.2.1 Institutional hypothesis

Figure 1 provide a bivariate linear OLS regression of the institution quality index and log GDP per capita (PPP) 1995. It provides a picture of the found correlation between the two variables. The total number of observations is 98. Figure 1 provides a scatter plot of the relationship between the variables. No extreme outliers seem to be present. The corresponding regression data is added in Table 3. As described by both theory and regression analyses or our sample, the quality of institutions and economic performance are positively related. Leaving controls aside, the coefficient of the relationship between institutions and log GDP is 0.418, what means that for every increase of percentage point of institutional quality log GDP per capita (PPP) 1995 increases with 0.418 points. The correlation between the variables seems to be highly significant with a p-value of less than 0.01. The R-square is

0.616, this means that the variance of institutional quality determines 61.6% of the variance in economic performance. This is a relative high number compared to the explanatory power of the basic regressions of the other hypotheses. The combination high significance and high R-square of the correlation creates high expectations for the effect of institutions on economic performance in multiple regressions. Per those statistics institutional theory seems to provide a very valid explanation of economic development when not controlling for other variables. This variable seems to fit the set expectations and is ready for additional regression analyses.

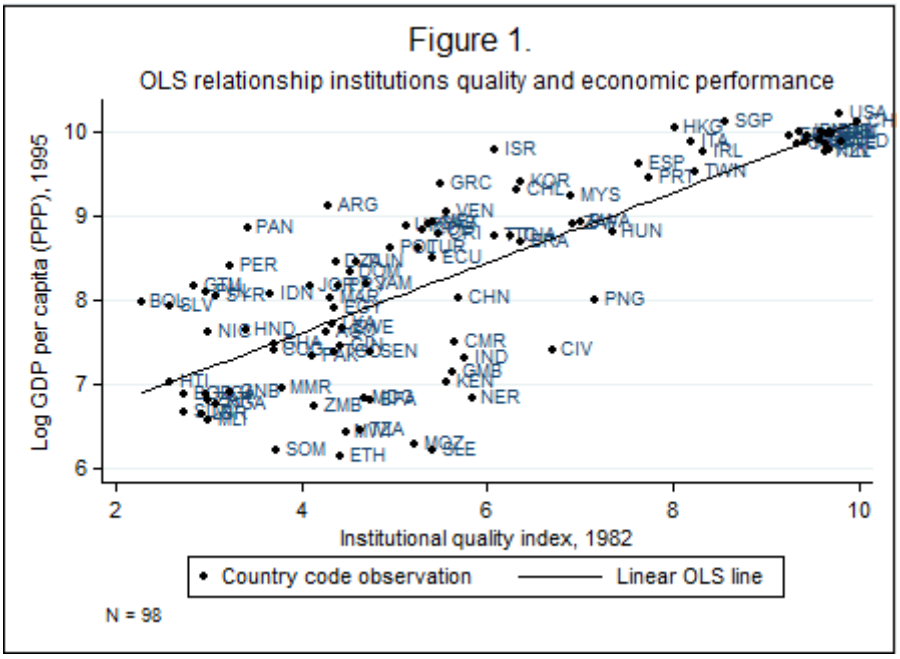

Table 3.	
Simple regression institutions quality and economic performance	
	Log GDP per capita (PPP), 1995
Institutional quality index	0.418***
	(12.42)
Intercept	5.952***
	(28.96)
N	98
R-square	0.616
t statistics in parentheses	
* p<0.05, ** p<0.01, *** p<0.001	

5.2.2 Policy hypothesis

Figure 2 provides a OLS regression of openness to international trade and log GDP per capita (PPP) 1995. The scatterplot present the visualisation of the relationship between the variables. The characteristics of this variable are described in the research design and clarify the necessary conditions that must be met for an observation to be classified as 'open'. Those strict conditions cause a lot of countries to be classified with zero. Other, generally better preforming, countries are classified with 1. Observations with a value of 0 for openness are scattered all over the y-axis representing economic performance. Countries with a value of 1 are mostly centred in the upper right corner, representing even higher economic performance. Corresponding regression data is added in Table 4. The scale of the variable openness is 0 till 1, while the scale of the y-axis representing log GDP per capita (PPP) 1995 in this figure varies from around 6 till 10. For this reason, the coefficient is very high. Simple regression analyses provide a coefficient of 2.079, controls left aside. This positive relationship is fits theory. More openness is generally linked to more economic performance. This coefficient is highly significant with a p-value of less than 0.01. The total number of observation is 140 and the R-square is 0.529. Just like the results of institutional regression, those results create high expectations for the multiple regressions.

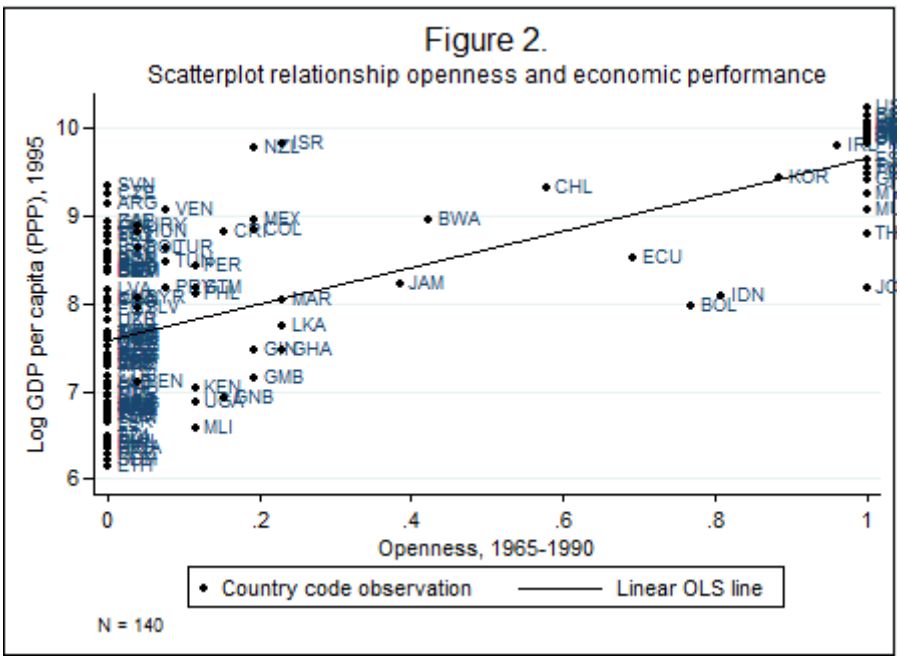

Table 4.	
Simple regression openness and economic performance	
	Log GDP per capita (PPP), 1995
Openness	2.079***
	(12.44)
Intercept	7.607***
	(97.32)
N	140
R-square	0.529
t statistics in parentheses	
* p<0.05, ** p<0.01, *** p<0.001	

5.2.3 Geographical hypothesis

The geographical hypothesis differs slightly from the other hypotheses in this research because it is represented by different variables. Upcoming regressions will provide an overview of how those variables individually and combined affect economic performance. As described in theory and translated into the research design, the geographical hypothesis is split up into variables representing geographic location and climate conditions. Those categories indicate different aspects of the geographic hypothesis and for this reason are regressed in separate models.

Geographic location

This section covers the following variables: lnd100cr, pop100cr, landlock, and ciffob95. Table 5 present the regression results for those variables against log GDP per capita (PPP) in 1995 and are split up in different sub regressions to check their individual effect on economic performance. Some variables are linked because of their high degree of internal correlation. Model 1 and Model 2 contain variables indirectly related to transaction costs (lnd100cr, pop100cr, and landlock). Model 2 includes the direct measurement of transport costs to the variables of Model 1 (ciffob95). All variables are combined in the Model 3.

Theory prescribes the negative influence of transport costs on economic performance. The variables use in the models are (indirect) indicators of transport costs. Regression results show that the percentage of population in geographic tropics has a slightly positive significant effect on economic performance. Both in Model 1 (0.0238) as in Model 3 (0.0247) where CIF/ FOB shipping costs are included. This effect does not hold for the comparable variable percentage of land in geographical tropics. Results indicate a insignificant negative effect for this variable. CIF/ FOB shipping costs are in both models negatively related to log GDP per capita and highly statistical significant. This variable is the most direct indicator of transport costs and the regression result fit the theory. All the geographical

variables in Model 3 combined reflexed an R-square of 0.527. Indicating that over 50 percent of the variation in log GDP per capita is associated with the variation in the geographical variables. The landlocked variable takes on an insignificant but positive coefficient of 0.198. This outcome does not fit theory. Per theory being landlocked supposes to negatively affect economic performance through the increasing transportation costs. Adding additional variables to the model could possible change the direction of the effect. Table 5 provides a first glimpse of regression while not including additional controls. Pearson's Correlation Coefficient Index displays the relationship between percentage of land within 100 km of the coast or river and percentage of the population within 100 km of the coast or river. The variables are highly and significantly correlated (correlation: 0.9206, p-value: 0.000)[6].

Table 5.

Multiple linear regression geographic location factors and economic performance

	Log GDP per capita (PPP), 1995		
	Model 1	Model 2	Model 3
Percentage land within 100 km of coast or river	-0.00851		-0.00886
	(-1.51)		(-1.44)
Percentage population within 100 km of coast or river	0.0238***		0.0247***
	(3.88)		(3.51)
Landlock dummy	0.0338		0.198
	(0.14)		(0.63)
CIF/ FOB shipping costs ratio		-7.422***	-4.855***
		(-6.56)	(-4.38)
Intercept	7.211***	16.64***	12.72***
	(35.46)	(13.04)	(9.61)
N	150	83	83
R-square	0.271	0.347	0.527
adj. R-square	0.256	0.339	0,503

t statistics in parentheses

* p<0.05, ** p<0,01, * p<0,001

Climate conditions

Table 6 present the results of multi linear regression of climate conditions and log GDP per capita (PPP) 1995. The following variables are used: malfal94, soilsui1, tropicar, and troppop. Model 1 contains the falciparum malaria index 1994, Model 2 the soil suitability, and Model 3 the percentage of land and population in geographic tropics. Model 4 combines all those variables representing climate conditions.

[6] See Appendix Table C

Theory links high prevalence of disease to a low economic performance. Regression results show that the variable falciparum malaria index has a large negative effect on log GDP per capita in both of its models. Both are also highly significant with a corresponding p-value lower than 0,001. The falciparum malaria index individually has a R-square of 0.372. Combining percentage of land and population in geographic tropics has almost the same explanatory power (0.369). The R-square of Model 4 is 0.498. This number is comparable to the explanatory power of table 5 regressing geographic location factors.

Soil suitability has only a very low correlation coefficient when individually regressed (0.0125). But this coefficient turns negative (-0.0248) and becomes highly significant when the other variables are added in the regression. This negative effect on economic performance does not fit theory. Per theory an increase in percentage of soil suitability would cause economic performance to increase. Additional regressions will point out if including additional controls changes the effect of soil suitability on the dependent variable, but it seems unlikely those regression results will point out that this variable is a decisive cause of economic performance.

Regression results show that percentage of land and population in geographic tropics do fit theory. In Model 4 both variables negatively related to economic performance (land: -0.158, population: -0.770). The coefficient of percentage of population in geographic tropics is statistical significant. The level of significance decreases, but stays significant at a level of 0.05 when additional variables are included. An explanation for this reaction is that this effect could decrease because of the including of the falciparum malaria index. Pearson's Correlation Coefficient Index indicates a high significant correlation between percentage of land in geographic tropics and falciparum malaria index (coefficient: 0,7156, p-value: 0.000) and percentage of the population in geographic tropics (coefficient: 0,7631, p-value: 0.000) [7]. Representing a high correlated between those variables. The falciparum malaria index could be the explanatory power behind the significant negative effect that the variable percentage of people living in geographic tropics has on economic performance. This variable could possibly be the hidden element that causes the variable population in geographic tropics to be significant in the first place. The Pearson's Correlation Coefficient indicates that the percentage of land in geographic tropics and the percentage of population in geographic tropics are also highly and significant correlated (correlation: 0,7190, p-value: 0.000) [8].

[7] See Appendix Table C
[8] See Appendix Table C

Table 6.				
Multiple linear regression climate conditions and economic performance				
	Log GDP per capita (PPP), 1995			
	Model 1	Model 2	Model 3	Model 4
Falciparum malaria index 1994	-1.693***			-1.390***
	(-9.33)			(-4.94)
Agricultural soil suitability		0.0125		-0.0248**
		(1.31)		(-3.16)
Percentage of land in geographic tropics			-0.310	-0.158
			(-1.32)	(-0.65)
Percentage of population in geographic tropics			-1.596***	-0.770*
			(-5.16)	(-2.33)
Intercept	8.636***	7.931***	8.800***	9.238***
	(94.97)	(49.03)	(77.69)	(53.30)
N	149	146	138	135
R-square	0.372	0.012	0.369	0.498
adj. R-square	0.368	0.005	0.360	0.483
t statistics in parentheses				
* p<0.05, ** p<0,01, * p<0,001				

5.2.4 Cultural hypothesis

The cultural hypothesis can be distinguished in the social capital and the human capital subdivision. Due to discoveries made in the descriptive analyses, variables linked to the social capital cannot be used for regression analyses as intended. The variables trust and civic will not be included in the final regressions because of their high degree of missingness. This part will however provide the regression results of the relationship between those independent variables and log GDP per capita (PPP) 1995. In some way, this provides the only option to compare the existing theory with my sample. Results of those analyses cannot confirm or falsify theory because of the not inclusion of the variables in the final regression, but it does profit an image of the relationship with the main dependent variable.

Model 1 and Model 2 display the results for regressing interpersonal trust and strength of norms of civic cooperation against log GDP per capita in that order. Model 3 combines all independent variables. On first sight trust in society appears to have a small positive effect on log GDP per capita (0.029). Strength of norms of civic cooperation has an negatable insignificant effect (0.0088). The coefficient even becomes negative when more variables are added to the model (-0.0577). Theory argued for a positive relationship where norms of civic cooperation causes economic performance. Model 2 shows that strength of norms of civic cooperation also has no explanatory power of the dependent variable (R-square: 0.000). When only regression years of schooling against log GDP per capita the total number of observations highly increases to 125 observations, close to the total number of observations in the

database. The relationship has a coefficient of 0.304 and is highly significant. When this variable is included with the other variable in Model 5 results change dramatically. Trust in society loses its significant relationship and the coefficient also drops dramatically. Model 4 represents the statistical data for simple regressing years of schooling and economic performance. Both in this model as in Model 5 years of schooling has a positive and significant effect. Model 4 has a R-square of almost 50 percent, and this R-square increase as shown in Model 5 when the additional social capital variables are added.

Regression analyses on the data sample fits theory speculating for the relationship between human capital and economic performance. However, the data highly apposes theory that argues for the importance of social capital as cause of economic performance. The effects are neglectable and non-significant. Theory speculated social capital as cause of economic performance. Theory also describes interpersonal trust and strength of norms of civic cooperation both as measurement of the same concept, social capital. But when checking the bivariate correlation between those variables the correlation coefficient is just 0.3808[9]. This is not a very big interaction. This coefficient is nonetheless significant (p-value: 0.046).

As repeatedly mentions, my data sample does however not provide room for valid scientific research of the effects of social capital on economic performance due to the low number of observations of the variables. Based on those regression analyses data human capital does highly correlates with economic performance but social capital does not.

[9] See Appendix Table C

Table 7.					
Multiple linear regression social and human capital and economic performance					
	Log GDP per capita (PPP), 1995				
	Model 1	Model 2	Model 3	Model 4	Model 5
Interpersonal trust	0.029**		0.0333**		0.00625
	(2.98)		(3.18)		(0.83)
Strength of norms of civic cooperation		0.0088	-0.0806		-0.0577
		(0.11)	(-1.09)		(-1.38)
Years of schooling				0.304***	0.197***
				(11.01)	(4.57)
Intercept	8.448***	9.133**	11.47***	7.262***	10.50***
	(22.69)	(2.92)	(4.11)	(62.10)	(6.62)
N	28	28	28	125	27
R-square	0.254	0.000	0.288	0.496	0.639
adj. R-square	0.225	-0.038	0.231	0.492	0.592
t statistics in parentheses					
* p<0.05, **p<0.01, *** p<0.001					

5.3 Interconnectedness independent variables

The limitation of possible interaction multicollinearity is already mentioned in research design. One of the assumptions of linear regression is that there is little of no multicollinearity in the data (Sykes, 1993). The first indicator in checking for multicollinearity is through analysing the Pearson Bivariate Correlation Coefficient index for all independent variables. This index shows the bivariate relationship between all the independent variables. All correlation coefficients in the Pearson Bivariate Correlation Coefficient must be smaller than 1. The Pearson Correlation Coefficient matrix of all independent variables can be found in Appendix Table C. None of the coefficients of the independent variables have a correlation score of 1. Some of the correlation do come close, what presents a high correlation between the two independent variables. Therefor I also checked the Variance Inflation Factor (VIF) to test for possible multicollinearity. The VIF indicates the degree of multicollinearity and measures to what extent the variation of the other independent variables increases as effect of adding the concerning variable into the model. The rule of thumb is that if the VIF is exceeds 10 there is an indication for multicollinearity of that the (StatisticsSolutions, 2017). This is also indicated by the 1/VIF function, what is another indicator of the same results. Here the rule of thumb is that the 1/VIF function should not be smaller than 0,10. The VIF results are presented in Appendix Table D. None of the value exceed the VIF score of 10. This indicates that there is no reason to belief there is a problematic multicollinearity between the independent variables.

5.4 Multiple Regressions

Where the basic regressions only run regressions freestanding from other variables, additional controls are needed to provide a clear image of the underlying relationships between the independent variables and log GDP per capita (PPP) 1995. In upcoming pages the main representative variables for each hypothesis will be used while additionally controlling for the effect of the variables of the other hypothesis and using additional control variables. This multiple regression analysis, including all additional variables and controls, will help to provide an answer to the research question. Interpreting those results could provide an idea of the main causes of economic performance. Merging all the datasets created a situation where not every country has values for every variable. Only those countries that possess variables for every variable can be regressed in the final model when all those variables are included. This decreases to total sample population used in the regressions.

Table 8 presents the results from the final regression analysis. It consists of all the main independent variables found suitable for multiple regression analyses and additionally includes an external war dummy, Western Europe dummy, Sub-Saharan Africa dummy and "other" geographic region dummy as controls. All those variables are regressed against log GDP per capita (PPP) 1995 using a total of 5 models. Every model differs slightly through the inclusion of other independent variables or controls.

Model **1** is a multivariate linear regression analysing the relationship between the independent variables suitable for regression analyses (as described in the basic regressions) and the dependent variable, economic performance. The independent variables included are the following: institutional quality index; openness to international trade; percentage of land within 100 km of the coast or river; percentage of population within 100 km of the coast or river; landlock dummy; CIF/FOB shipping costs ratio; falciparum malaria index 1994; agricultural soil quality; percentage of land in geographical tropics; percentage of total population in geographical tropics; and the average years of schooling. As discussed the variables interpersonal trust and strength of norms of civic cooperation are not included due to their low number of observations. This model tests the individual effect of each variable of economic performance, while holding the rest constant. Model 1 is constructed through analysing 62 countries. All correlations coefficients are expected to be biased upwards due to reversed causality. This is expected because the OLS regression captures both the positive impact of the independent variables on economic performance as the other way around.

This model explains around 88 percent of the variation in the dependent variable (R-square: 0.887, Ajd. R-square: 0.862). This very high explanatory power reveals that the model includes the right variables to explain cross-country variation in economic performance. Identifying that the theories described by the hypotheses of economic development capture the fundamental aspects in this dataset. This measurement of explanatory power does include all independent variables as described

in the multivariate regression model. Analysing the individual outcomes of the independent variables helps to point out which of those factors are crucial in describing the variation. Most results of this regression are consistent with the initial findings of the basic regressions.

Results show that the institutional quality index variable is highly significant with a p-value of less than 0.01 and a corresponding t-value of 3.21. It holds is its positive correlation with log GDP per capita (PPP) as described in Table 3 of the basic regressions. The correlation coefficient of 0.161 indicates that every increase of one point on the institutional quality index corresponds with an increase of approximately 16 percent points of the log GDP per capita (PPP) 1995. This coefficient decreased compared to the bivariate regression results of Table 3 (0.418) due to the inclusion of additional variables. This indicates that institutions alone are not able to explain the variation in economic performance, but additional variables should be included. The results do still fit theory that the provision of good institutions creates a situation which allows the economy to flourish. For this reason, I can confirm the set hypothesis by theory that economic performance is linked to the quality of institutions. The research design does however not provide the possibility to determine the direction of this correlation. Consequently, I cannot confirm the causal relationship between the existence of good institutions and economic performance.

Openness to international trade is significantly correlated to log GDP per capita (PPP) 1995 at a 95 percent confidence level. The positive relationship indicates that being open to international trade increases Log GPD per capita (PPP) with more than 50 percent points (coefficient: 0.519) compared to being completely closed to international trade. Indicating that being open to international trade is associated with high economic development. This fits theory that openness of a country's economy, as influenced by its macro-economy policy, positively affect its economic performance. The direction of the correlation unfortunately not be distinguished out of those results.

All variables related to geographical location are insignificant. The percentage of population (coefficient: 0.00305) and land (coefficient: -0.000861) within 100 km to the coast of river both have a neglectable relationship with log GDP per capita (PPP) 1995. Although both variables are insignificant and have neglectable coefficients the percentage of population is positively related to economic performance and the percentage of land within 100 km of the coast or river is negatively associated. The direction of the coefficient contradicts theory. Theory argues for the positive effect of both variables on economic performance through their decreasing effect on transport costs. The results correspond to the finding of Table 5, except for the insignificant relationship between the percentage of population within 100 km to the coast of river. Without controlling for the additional variable added in this model this variable was found to be significantly correlated to economic performance at a 99-

percentage confidence level (see Table 3). The landlock dummy and CIF/ FOB shipping costs ratio are both negatively but insignificantly associated with log GDP per capita (PPP) 1995. Being landlocked corresponds to a coefficient of -0.181 and CIF/ FOB shipping costs at a coefficient of -0.754 when regressed against log GDP per capita (PPP). Regression results of Table 3 reported a positive insignificant effect. Meaning that adding more variables to the model changed the direction of the landlock coefficient. Although the results are not robust, this direction of the association better fits theory arguing that being landlocked negatively influences economic performance. Due to the insignificant outcomes of the variables data does not support the effects of the geographic location on economic performance in any way. Results indicating that transport costs are not fundamental causes of economic performance.

The climate condition subpart of the geographic hypothesis shows more promising results. The falciparum malaria index variable is negatively associated with log GDP per capita (PPP) 1995 with a corresponding coefficient of -0.871. There is a high significant correlation at a 99 percent confidence level. Those data results roughly correspond to the basic regression run for climate conditions on economic performance as described in Table 6. The coefficient did decline from -1.390 to -0.871 but stayed significant. Regression results argue for the correlation between falciparum malaria index, as indicator of prevalence of disease, and economic performance as argued in theory.

The basic regression already argued for the subsumed effect of population in geographic tropics and the falciparum malaria index. The variable percentage of population in geographic tropics loses the significance it has in Table 6, but the coefficient stays negative (-0.105). The variable percentage of land in geographic tropics stays insignificant but the direction of the coefficient changes to positive (coefficient Table 6: -0.158; coefficient Table 8: 0.0586). This could however be due to change because of the lack of significance, for this reason I cannot draw any additional conclusions out of those specific results. Agricultural soil quality is insignificantly but marginal negatively related to economic performance with a coefficient of -0.00858. This negative association corresponds to the findings of Model 7 of Table 6. Both results contradict theory that agricultural productivity is crucial for economic development.

Human capital presented by the initial level of education, measured by average years of schooling 1965 is significant correlated with log GDP per capita (PPP) 1995. For every additional year of education log GDP per capita (PPP) increases with 0.0792 in Model 1. The significance and direction of the correlation are both comparable with the results of the basic regression testing for the non-controlled effects of social and human capital on economic performance as displayed in Table 7. Only the strength of both elements decreases when additional variables are added to the model. Indicating that initial level of

schooling does significantly relates to economic performance, but that adding additional variables declines some of its explanatory power.

Ideally results of regression models both include a high level of significance and a high degree of explanatory power. To check for the explanatory power of the individual significant variables of Model 1 of Table 8 I compare the R-square data of the bivariate regressions of the individual independent variables. This is not the best method of checking the individual influence of the significant variables, but it does provide comparable data. The significance of the variables is shown by the stars next to the correlation coefficient and the t-values in Table 8 Model 1 and the separate bivariate regression models. This data confirms the significant relation of four independent variables with the dependent variable. The R-square indicates how much of the variance in the dependent variable is predicted by the independent variable. Table 3, 4, 6, and 7 provide the R-square of the bivariate regression of the significant variables. Those tables provide consecutive data for the institutional quality index (R-square: 0.616), openness to international trade (R-square: 0.529), falciparum malaria index (R-square: 0.372), and average years of schooling (0.496). A high R-square means that the observed outcomes and the observed predicted values is relatively small. Implicating that the independent variable is relatively good in predicting the coherent dependent variable. The additional value of examining the R-square should not be overrated, but it does provide an idea of how big the explanatory power of each independent variable is. The amount of observations in the bivariate institutional quality index regression is with a total of 92 observations a bit lower than the other bivariate regression models. Compared to openness to international trade with 140 observations, falciparum malaria index with 149 observations and average years of schooling with 127 observations[10]. This could affect the R-square.

Through compare the R-squares of the bivariate models with the t-value and correlation coefficients presented in Model 1 of Table 8, I can analyse which factors particularly determine the cross-country differences in economic performance. All four variables are significant, and have a high R-square when individually regressed against log GDP per capita (PPP) 1995. The different scales of the variables make their effect on economic performance a bit harder to compare. But when rescaling variables, the effects are comparable. Indicating that all four of those independent variables can be view as fundamental determiners of economic performance in this dataset.

[10] See Table 1

Model 2 includes the external war dummy variable to the regression of Model 1. The effects of war on the economy have long been subject to discussion (Moffatt, 2017). Initially, it is obvious that contributing to war effects the economy. War affects the production capacity, which could destroy productive input like physical and human capital (Koubi, 2005, p. 68). This negatively affects economic performance. An argument related to the positive effect of war on the economy is that war gives the economy a boost. Demand for certain supplies increases, what could lead to a higher employment rate and initially higher economic performance (Moffatt, 2017). Research has pointed out that war indeed influences the economy, but the direction of the relationship is highly subject to the characteristics of the war (Koubi, 2005).

Regression results show the following. The total number of observation decreases with 1, meaning that one of the original observations does not include data for this variable. Results show that including the dummy in this regression does not influence the outcome. Coefficients of independent variables only minimally change and there is no change at all in the significance levels. The coefficient of the external war dummy variable (0.143) is positive, what marks a positive effect of external war on economic performance. This outcome would indicate that being involved in an external war positively correlates with economic performance. It is however non-significant. What means there is no reason to assume a linear relationship exists and is not due to chance. The R-square of the model increases with 0.2% to 0.889. This is already a minimal increase of explanatory power, but the adjusted R-square shows no increase in explanatory power. This indicates that adding this dummy variable to the regression does not contribute in any way to the effectiveness of the model.

Model 3 possesses the same independent variables as previous models but does not excludes the geographic region Sub-Saharan Africa of the sample population. This decreases the total observations to 44 countries. This model is included because of the information presented in the descriptive analyses. Table 2 describes the dataset using geographic regions as dummies, indicating that Sub-Saharan Africa is the worst preforming geographic region. By excluding all the countries that are part of the Sub-Saharan Africa region I check if the relationship between the independent and dependent variables stay the same. Model 3 shows that most results are robust to dropping all the Sub-Saharan Africa countries from the sample. The coefficient of the institutional quality index increases to 0.183 compared to Model 1 and stays significant with an increasing t-value to 3.35. Indicating that the estimate of institutions is now bigger and more precise. The coefficient of openness to international trade slightly decreases to 0.497 but remains significant. The effect of the falciparum malaria index marginally increases but stays significant at a 95 percent confidence level. The variable average years of schooling does lose its significance and the coefficient also decreases (coefficient: 0.0567). Illustrating that excluding the Sub-Saharan Africa countries change the regression estimates.

Model 4 adds continental dummies to the regression. Adding the Western Europe, Sub-Saharan Africa and 'other' geographic region dummy to the model creates the opportunity to check for the individual effect of those regions on economic performance. The 'other' geographic regions includes all geographic regions that are not Western Europe of Sub-Saharan Africa. All dummies are insignificant. This fits theory that there is no need an 'Africa proxy'. The insignificant results of the dummies indicate that being part of those geographic regions does not determine economic performance. Different values for the significant independent variables do. The institution quality index, falciparum malaria index and average years of schooling remain significant. Openness to international trade losses its significance when adding the continental dummies. The Western Europe dummy has a positive coefficient (0.0938), Sub-Saharan Africa a negative coefficient (-0.0838) and the 'other' geographic regions also has a negative coefficient (-0.0902). As mentioned none of the dummies are significant, meaning that regression cannot rule out the effect of change on the results.

The descriptive analyses provided an overview of the total observations of each variable. In order to increase the number of observations in the regression Model 5 excludes the CIF/ FOB shipping costs ratio. The amount of observations increases to a total of 86. This model help to check for the robustness of the regression when the sample population increases. The CIF/ FOB shipping costs is excluded because it has no significant relationship with log GDP per capita (PPP) in additional models and has the highest proportion of missingness of observations. Dropping additional variables would harm the regression to heavily. The results of this model stay close to the baseline results of Model 1. The coefficient of the falciparum malaria index and average years of schooling increase compared to Model 1. The explanatory power of the model decreases marginally when excluding this variable (R-square: 0.874). All independent variables that where significantly correlated with log GDP per capita (PPP) in the first model still are, all with a higher corresponding t-value. Meaning that the level of significance increased. Model 5 emphasizes the results of Model 1.

All additional controls added in Model 2 to 5 only marginally change the outcomes of the regression results of Model 1. Indicating that the results of Model 1 are robust to controlling for those additional factors.

Table 8.					
Multiple regression analyses economic performance					
	Log GDP per capita (PPP), 1995				
	Total sample	Sample including war-torn	Sample without Sub-Saharan Africa	Sample with continent dummies	Sample without CIF/ FOB shipping costs ratio
	Model 1	Model 2	Model 3	Model 4	Model 5
Institutional quality index	0.161**	0.162**	0.183**	0.157**	0.167***
	(3.21)	(3.06)	(3.35)	(2.89)	(4.36)
Openness to international trade	0.519*	0.520*	0.497*	0.436	0.493**
	(2.52)	(2.53)	(2.50)	(1.69)	(2.97)
Percentage of land within 100 km of the coast or river	-0.000861	-0.000130	0.00132	-0.000621	-0.000587
	(-0.21)	(-0.03)	(0.35)	(-0.12)	(-0.18)
Percentage of population within 100 km of the coast or river	0.00305	0.00217	0.00155	0.00225	0.000196
	(0.60)	(0.42)	(0.32)	(0.38)	(0.05)
Landlock dummy	-0.181	-0.236	-0.154	-0.199	-0.203
	(-0.84)	(-1.08)	(-0.60)	(-0.88)	(-1.20)
CIF/FOB shipping costs ratio	-0.754	-0.833	-1,530	-0.720	
	(-0.95)	(-1.03)	(-0.65)	(-0.86)	
Falciparum malaria index 1994	-0.871**	-0.936**	-0.912*	-0.887*	-1.051***
	(-3.09)	(-3.25)	(-2.48)	(-2.50)	(-5.51)
Agricultural soil quality	-0.00858	-0.00951	-0.00752	-0.00894	-0.00528
	(-0.95)	(-1.06)	(-0.87)	(-0.96)	(-0.76)
Percentage of land in geographical tropics	0.0220	0.0418	0.0565	0.129	0.0586
	(0.09)	(0.16)	(0.23)	(0.38)	(0.30)
Percentage of total population in geographical tropics	-0.105	-0.0245	-0.0971	-0.167	-0.208
	(-0.23)	(-0.05)	(-0.17)	(-0.33)	(-0.59)
Average years of schooling	0.0792*	0.0788*	0.0567	0.0845*	0.0921**
	(2.07)	(2.04)	(1.54)	(2.04)	(2.87)
External war dummy		0.143			
		(0.56)			
Western Europe dummy				0.0938	
				(0.35)	
Sub-Saharan Africa dummy				-0.0838	
				(-0.22)	
"Other" geographic region dummy				-0.0902	
				(-0.32)	
Intercept	8.131***	8.250***	8.887**	8.178***	7.421***
	(8.56)	(8.41)	(3.34)	(8.08)	(24.21)
N	62	61	44	62	86
R-square	0.887	0.889	0.894	0.888	0.874
adj. R-square	0.862	0.862	0.857	0.855	0.857
t statistics in parentheses					
* p<0.05, ** p<0.01, *** p<0.001					

Chapter 6: Discussion & Conclusion

The aim of this inquiry was to answer the research question: '*What are the fundamental determinants of cross-country differences in economic performance?*'. Hoping to contribute to the academic literature of development economics. Existing academic literature provided useful points of departure in search of the causes of cross-country world inequality. Conducted analyses tested for the association between different groups of variables, representing different economic development hypotheses, and economic performance.

Of the four economic development hypotheses set up in academic literature all four in some way contribute to explaining the variation in cross-country economic performance. Multiple regression analyses providing evidence for linear relationships between particular variables and economic performance. My dataset points out that the quality of institutions, openness to international trade, falciparum malaria index, and the average years of schooling all statistical significantly and to a mentionable extent influence economic performance. Recognizing that four independent variables, representing four different schools of thought, are significant related to economic performance indicates than none of those hypotheses of economic development is comprehensive. Data points out that those different elements all contribute in explaining cross-country differences in economic performance.

Good institutions create incentives for economic activity and decrease transaction costs, what translate in a higher economic performance. Institutions form the structure of the economy. Additionally, the lack of good institutions causes a country's economy to preform worse. A country's macro-economic policy additionally contributes to explaining differences cross-country economic performance. The openness of a country's economy to international trade is also significantly related to its economic performance. Open economies tend to perform better than closed economies. This corresponds to theory emphasizing the importance of liberal trade regimes. Any governmental policy harming openness of an economy leads to a lower degree of economic performance. Supplementary human capital also highly relates to economic performance. High degree of schooling relates to high economic performance. Every additional increase in human capital account for a higher degree of performance. This corresponds to theory. Prevalence of disease is also pointed out as a relevant factor in determining economic performance. There is a clear negative correlation. Indicating that a high disease burden relates to a relatively lower economic performance as pointed out in theory.

Not all theory did however prove to be appropriate for analysing economic performance. No signals for the importance of geographical location and its effect on economic performance through transport costs is found. Data rejects the influence of a country's geographic location as determining factor as

described by the geography hypothesis. All related variables represent an insignificant association. Social capital is also not found to be significantly correlated with economic performance. The effect of social capital is not ruled out due to measurement errors, but basic regressions conducted to check for the existence of any effect did not found relevant interactions.

6.1 Limitations

Although this thesis is carefully conducted, the used research design and additional limitations that emerged during the research progress restrain the academic value of this thesis. This thesis contributed to the search for fundamental determinants of cross-country differences in economic performance but an accurate answer to the fundamental research questions is still out of reach. Due to the set-up of the research design checking for causal relationships is not possible**.** Meaning that the research question cannot be answered completely. However, regression analyses do provide evidence for significant relationships between different independent variables and economic performance. Those correlations are analysed through theory and argue for the importance of their represented elements. This limitation is directly linked to the problem of reversed causality. Because the research design cannot determine the direction of the causality the possibility exists that economic performance additionally affects the independent variable used for determining the determinants of economic performance. This is very likely. Literature already discussed the two-sided function of many of the relationships. Rich countries will inevitably possess higher standards of living due to their better preforming economies. Those higher standards of living again positively affect economic performance. This interaction further complicates the search for fundamental causes of economic performance.

Additional problems arise in directly analysing the four main hypotheses of economic performance in regression models. The idea was to create a platform to objectively assess the effect of each hypothesis on economic performance. This required a merge of different datasets to include the variables that indicate the different theories linked to the hypotheses. Not every country does however possess observations for every different variable. This lead to a situation where only approximately fifth percent of the total sample size could be included in the final multivariate regression. This high degree of missingness lowers the reliability of the regression results.

The last limitation concerns the generalizability of the research results. This thesis analyses the fundamental causes of economic performance in the period 1960 to 1995. It includes most world countries, making it generalizable for the whole worlds cross-country differences in economic performance. This generalizability is only harmed by the degree of missingness in country specific observations. Outcomes of this thesis cannot directly be generalized to periods outside the analysed

period 1960 to 1995. Additional data for those years must be analysed to check if the found correlations remain.

This thesis helped to pinpoint the crucial elements related to economic performance but additional research is needed to specify the relationship and direction of the found correlations. This thesis does specify correlations that need to be considered in the search of the fundamental causes of cross-country differences in economic performance. Further research should be conducted to continue the search for fundamental determinants of economic performances. Focussing on the factors found to be significantly associated in this thesis. This should be done by searching for exogenous variation in the main independent variables. In this way, the problem of (reversed) causality can be countered. The fundamental causes could be viewed as given variables, influencing the endogenous variable economic performance. Creating the possibility to determine the direction of the correlation. Using panel data of cross-country differences in economic performance could also contribute to finding causal relationships.

Appendix

Appendix Table A							
Country specific data							
Country	Abbreviation country name	Log GDP per capita (ppp), 1995	Geographic region	Country	Abbreviation country name	Log GDP per capita (ppp), 1995	Geographic region
Afghanistan	AFG	6,397	mideast	Lebanon	LBN	8,626	mideast
Albania	ALB	7,098	transit	Lesotho	LSO	7,299	safri
Algeria	DZA	8,464	mideast	Liberia	LBR	6,646	safri
Angola	AGO	7,628	safri	Libya Arab Jamahiriy	LBY	8,781	mideast
Argentina	ARG	9,136	latam	Lithuania	LTU	8,352	transit
Armenia	ARM	7,604	transit	Madagascar	MDG	6,847	safri
Australia	AUS	9,897	-	Malawi	MWI	6,428	safri
Austria	AUT	9,965	eu	Malaysia	MYS	9,236	eseasia
Azerbaijan	AZE	7,309	transit	Mali	MLI	6,578	safri
Bangladesh	BGD	6,879	sasia	Mauritania	MRT	7,523	safri
Belarus	BLR	8,348	transit	Mauritius	MUS	9,057	safri
Belgium	BEL	9,985	eu	Mexico	MEX	8,935	latam
Benin	BEN	7,096	safri	Moldova, Republic of	MDA	7,377	transit
Bolivia	BOL	7,977	latam	Mongolia	MNG	7,509	transit
Bosnia and Herzegovina	BIH	6,397	transit	Morocco	MAR	8,041	mideast
Botswana	BWA	8,948	safri	Mozambique	MOZ	6,285	safri
Brazil	BRA	8,697	latam	Myanmar	MMR	6,957	eseasia
Bulgaria	BGR	8,479	transit	Namibia	NAM	8,545	safri
Burkina Faso	BFA	6,818	safri	Nepal	NPL	6,944	sasia
Burundi	BDI	6,488	safri	Netherlands	NLD	9,895	eu
Cambodia	KHM	6,492	eseasia	New Zealand	NZL	9,768	-
Cameroon	CMR	7,502	safri	Nicaragua	NIC	7,620	latam
Canada	CAN	9,990	-	Niger	NER	6,835	safri
Central African Republic	CAF	7,327	safri	Nigeria	NGA	6,776	safri
Chad	TCD	6,790	safri	Norway	NOR	10,011	eu
Chile	CHL	9,320	latam	Oman	OMN	9,192	mideast
China	CHN	8,030	eseasia	Pakistan	PAK	7,351	sasia
Colombia	COL	8,841	latam	Panama	PAN	8,860	latam
Congo	COG	7,404	safri	Papua New Guinea	PNG	8,008	eseasia
Costa Rica	CRI	8,798	latam	Paraguay	PRY	8,169	latam
Cote d'Ivoire	CIV	7,412	safri	Peru	PER	8,416	latam
Croatia	HRV	8,355	transit	Philippines	PHL	8,092	eseasia
Cuba	CUB	7,170	latam	Poland	POL	8,634	transit
Czech Republic	CZE	9,241	transit	Portugal	PRT	9,473	eu
Denmark	DNK	10,001	eu	Romania	ROM	8,381	transit
Dominican Republic	DOM	8,353	latam	Russian Federation	RUS	8,398	transit
Ecuador	ECU	8,511	latam	Rwanda	RWA	6,358	safri
Egypt	EGY	7,918	mideast	Saudi Arabia	SAU	9,201	mideast
El Salvador	SLV	7,937	latam	Senegal	SEN	7,398	safri
Eritrea	ERI	6,346	safri	Sierra Leone	SLE	6,220	safri
Estonia	EST	8,384	transit	Singapore	SGP	10,133	eseasia
Ethiopia	ETH	6,142	safri	Slovak Republic	SVK	8,846	transit
Finland	FIN	9,803	eu	Slovenia	SVN	9,337	transit
France	FRA	9,960	eu	Somalia	SOM	6,215	safri
Gabon	GAB	8,913	safri	South Africa	ZAF	8,914	safri
Gambia	GMB	7,149	safri	Spain	ESP	9,618	eu
Georgia	GEO	7,402	transit	Sri Lanka	LKA	7,729	sasia
Germany	DEU	9,923	eu	Sudan	SDN	6,685	safri
Ghana	GHA	7,477	safri	Sweden	SWE	9,862	eu
Greece	GRC	9,397	eu	Switzerland	CHE	10,121	eu
Guatemala	GTM	8,166	latam	Syrian Arab Rep.	SYR	8,052	mideast
Guinea	GIN	7,458	safri	Taiwan	TWN	9,540	eseasia
Guinea Bissau	GNB	6,915	safri	Tajikistan	TJK	6,727	transit
Haiti	HTI	7,029	latam	Tanzania	TZA	6,461	safri

Honduras	HND	7,659	latam	Thailand	THA	8,778	eseasia
Hong Kong	HKG	10,064	eseasia	The fmr Yug Rep Macedonia	MKD	6,780	transit
Hungary	HUN	8,819	transit	Togo	TGO	7,394	safri
India	IND	7,323	sasia	Trinidad & Tobago	TTO	8,770	latam
Indonesia	IDN	8,080	eseasia	Tunisia	TUN	8,466	mideast
Iran	IRN	8,585	mideast	Turkey	TUR	8,638	mideast
Iraq	IRQ	7,601	mideast	Turkmenistan	TKM	7,640	transit
Ireland	IRL	9,778	eu	Uganda	UGA	6,877	safri
Israel	ISR	9,802	mideast	Ukraine	UKR	7,801	transit
Italy	ITA	9,895	eu	United Arab Emirates	ARE	9,664	mideast
Jamaica	JAM	8,208	latam	United Kingdom	GBR	9,869	eu
Japan	JPN	10,007	eseasia	United States	USA	10,215	-
Jordan	JOR	8,166	mideast	Uruguay	URY	8,888	latam
Kazakhstan	KAZ	8,061	transit	Uzbekistan	UZB	7,806	transit
Kenya	KEN	7,026	safri	Venezuela	VEN	9,053	latam
Korea Dem.People's Rep.	PRK	6,824	eseasia	Vietnam	VNM	7,279	eseasia
Korea,Republic of	KOR	9,422	eseasia	Yemen	YEM	6,683	mideast
Kuwait	KWT	9,741	mideast	Yugoslavia	YUG	7,601	transit
Kyrgyz Republic	KGZ	7,580	transit	Zaire	ZAR	6,819	safri
Lao PDR	LAO	7,082	eseasia	Zambia	ZMB	6,749	safri
Latvia	LVA	8,154	transit	Zimbabwe	ZWE	7,676	safri

Notes: sample countries sorted on alphabet. Total population of 150 countries. Geographic regions: Western Europe (eu); Sub-Saharan Africa (safri); South Asia (sasia); Transition countries (FSU & EE) (transit); Latin America & Caribbean (latam); East and Southeast Asia (eseasia); and Middle East (broad definition) (mideast)

Appendix Table B

Summary variables: data descriptions and sources

Variable	Description	Source(s)	Abbreviation in dataset
Dependent variable			
Economic performance	**Log GDP per capita 1995**: controlled for Purchasing Power Parity. Created by taking the log of GDP per capita 1995.	Geodata: Gallup, et al (1999). Original source: World Bank (1998)	loggdp95
Independent variables			
Institutions	**Institutional quality index**: Average of 5 sub-indexes of the International Country Risk Guide (ICRG), consisting of: the rule of law, the bureaucratic quality index, the corruption in government index, the risk of expropriation index, and the government repudiation of contracts index. Measured on a scale from 0 to 10. High scores indicate high level of institutional quality. The first ICRG data available for each country is used. Most countries contain data from the year 1982 and additionally the year's 1984 and 1985 are used.	Geodata: Gallup, et al (1999): as described by Knack and Keefer (1995). Original source: Political Risk Services Group (PRSGroup, 2017)	icrg82
Openness	**Openness to international trade**: Share of a country's economy classified as open in the period 1965 till 1990. Scale 0 to 1. High scores represent a high degree of openness. A country's economy is classified as open when the following criteria are met: An average tariff rates that does not exceed 40%; An average quota and licensing coverage of import below 40%; A black-market exchange rate premium below 20%; No extreme controls on exports; and is not considered as a socialist country per the standard of Kornai.	Geodata: Gallup, et al (1999): as described by Sachs and Warner (1997). Original source: Sachs and Warner (1995)	open6590
Geographic location: land	**Percentage of land within 100 km of the coast or river**: Percentage of land fitting that criteria measured by dividing this area through the total country's territory. Exludes Artic and sub-Artic region coasline above the winter extent of sea ice. Scale represents percentage from 0 to 100.	Geodata: Gallup, et al (1999). Original sources: Rand McNally and Co. (1980), Britannica Online (1998), and Encyclopedia Encarta (1998), Environmental Systems Research Institute (ESRI) (1996),	lnd100cr
Geographic location: population	**Percentage of population within 100 km of the coast or river:** Percentage of population fitting that criteria measured by dividing the proportion of population living in this area through the total country's population. Scale represents percentage from 0 to 100.	Geodata: Gallup, et al (1999), Original sources: Rand McNally and Co. (1980), Britannica Online (1998), and Encyclopedia Encarta (1998), Environmental Systems Research Institute (ESRI) (1996), Tobler et al. (1995)	pop100cr

PTO

Landlock	**Landlock dummy**: Equal to 1 when the county has direct access to the coast	Geodata: Gallup, et al (1999)	landlock
Transport costs	**C.i.f./ f.o.b. shipping costs ratio**: The abbreviation c.i.f. stands for 'costs of insurance and freight' and includes those costs to the total value. C.i.f. is divided by f.o.b.. The abbreviation f.o.b. stands for 'free on board' and measures the total costs of the transported good or services till shipment.	Geodata: Gallup, et al (1999), as described by Radelet and Sachs (1998). Original source: International Monetary Fund (IMF)	ciffob95
Prevalence of disease	**Falciparum malaria index 1994**: Percentage of a country's land area subject to falciparum malaria. Falciparum malaria is the deadly version of malaria. Scale represents percentage from 0 to 1.	Geodata: Gallup, et al (1999) Original source: World Health Organization (WHO, 1997)	malfal94
Agricultural productivity	**Agricultural soil quality**: mean of the estimate of the percentage of very suitable soil for the six rainfed crops. Scale represents percentage from 0 to 100.	Agric data: PortlandStateUniversity (2017), Matthews (1983). Original source: FAO (1995)	soilsui1
Tropic region: land	**Percentage of land in geographical tropics**: Percentage of land fitting that criteria measured by dividing this area through the total country's territory. The geographic tropics is the area between the Tropic of Cancer (23° 45¢ north latitude) and the Tropic of Capricorn (23° 45¢ south latitude). Covering the latitudes where the sun is directly overhead at least once a year. Scale represents percentage from 0 to 1.	Geodata: Gallup, et al (1999). Original source: Environmental Systems Research Institute (ESRI) (1996)	tropicar
Tropic region: population	**Percentage of total population living in geographical tropics**: Percentage of population fitting that criteria measured by dividing the proportion of population living in this area through the total country's population. The geographic tropics is the area between the Tropic of Cancer (23° 45¢ north latitude) and the Tropic of Capricorn (23° 45¢ south latitude). Covering the latitudes where the sun is directly overhead at least once a year Scale represents percentage from 0 to 1.	Geodata: Gallup, et al (1999). Original source: Environmental Systems Research Institute (ESRI) (1996), Tobler et al. (1995)	troppop
Social capital: trust	**Interpersonal trust**: Percentage of respondents that answers 'most people can be trusted' to the following survey question: "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" Scale represents percentage from 0 to 100. High scores represent a high degree of trust.	Knack and Keefer (1997). Original source: World Value Surveys (1981 and 1990-1991)	trust
Social capital: civic norms	**Strength of norms of civic cooperation**: Mean of survey answers of 5 statements about civic norms with a number in the scale 1-10. Scale is transformed to a scale from 0 to 100, representing percentage. High scores represent a high degree of cooperation.	Knack and Keefer (1997). Original source: World Value Surveys (1981 and 1990-1991)	civic
Human capital: initial level of schooling	**Average years of schooling**: Measured through analysing the population from age 25 and older. Average taken from total years of obtained schooling. Conducted for the year 1965.	Years of schooling dataset: Barrol and Lee (2013)	yr_sch
Control variables			
Geographic region	**Geographic region dummies**: Equal to 1 when the country is part of the geographic region. Every country is part of only one region. Geographic regions: Western Europe; Sub-Saharan Africa; South Asia; Transition countries (FSU & EE); Latin America & Caribbean; East and Southeast Asia; Middle East (broad definition)	Geodata: Gallup, et al (1999). With additional countries classified by the author	eu; safri; sasia; transit; latam; eseasia; mideast
War-torn	**External war dummy**: Equal to 1 when the country has participated in at least one external war in the period of 1960 till 1985	Geodata: Gallup, et al (1999). Original source: Barro and Lee (1994)	wardum
	Notes: additional detailed data descriptions can be found in the (original) source(s) and in the research design		

Appendix Table C													
Bivariate Pearson's Correlation Coefficient Index: independent variables													
	icrg82	open6590	lnd100cr	pop100cr	landlock	ciffob95	tropicar	troppop	malfal94	soilsui1	trust	civic	yr_sch
icrg82	1												
open6590	0,7342*	1											
	0												
lnd100cr	0,2806*	0,3233*	1										
	0,005	*0,000*											
pop100cr	0,3636*	0,3951*	0,9206*	1									
	0,000	*0,000*	*0,000*										
landlock	-0,0662	-0,2080*	-0,4348*	-0,5492*	1								
	0,517	*0,014*	*0,000*	*0,000*									
ciffob95	-0,4755*	-0,4274*	-0,3311*	-0,4300*	0,3828*	1							
	0,000	*0,000*	*0,002*	*0,000*	*0,000*								
tropicar	-0,5907*	-0,2230*	-0,1191	-0,1353	-0,06	0,5097*	1						
	0,000	*0,008*	*0,147*	*0,099*	*0,466*	*0,000*							
troppop	-0,4696*	-0,3345*	-0,5566*	-0,5984*	0,2693*	0,6245*	0,7190*	1					
	0,000	*0,000*	*0,000*	*0,000*	*0,001*	*0,000*	*0,000*						
malfal94	-0,4467*	-0,3257*	-0,3292*	-0,3496*	0,0624	0,5685*	0,7156*	0,7631*	1				
	0,000	*0,000*	*0,000*	*0,000*	*0,450*	*0,000*	*0,000*	*0,000*					
soilsui1	0,1324	-0,0371	0,3460*	0,2514*	-0,0256	-0,2072	-0,3987*	-0,4395*	-0,4270*	1			
	0,198	*0,668*	*0,000*	*0,002*	*0,759*	*0,065*	*0,000*	*0,000*	*0,000*				
trust	0,6694*	0,6175*	0,0548	0,4241*	0,0371	-0,4956*	-0,4363*	-0,3973*	-0,2483	-0,2306	1		
	0,000	*0,001*	*0,782*	*0,025*	*0,851*	*0,016*	*0,020*	*0,036*	*0,203*	*0,238*			
civic	0,1943	0,1359	-0,0642	0,0982	0,2522	-0,1337	-0,2203	-0,1429	0,0372	-0,12	0,3808*	1	
	0,322	*0,490*	*0,746*	*0,619*	*0,195*	*0,543*	*0,260*	*0,468*	*0,851*	*0,543*	*0,046*		
yr_sch	0,7276*	0,4453*	0,4106*	0,4275*	-0,0192	-0,4785*	-0,5695*	-0,5708*	-0,5819*	0,3035*	0,6192*	0,2013	1
	0,000	*0,000*	*0,000*	*0,000*	*0,831*	*0,000*	*0,000*	*0,000*	*0,000*	*0,001*	*0,001*	*0,314*	
Notes: first number is the correlation, scaled from -1 to 1. Italic is p-value. Variables written in abbreviation as used in dataset. Consult Appendix Table B for complete variable description													

Appendix Table D		
Variance Inflation Factor (VIF) test for multicollinearity		
Variable	VIF	1/VIF
pop100cr	9,28	0,1077
troppop	8,3	0,1204
lnd100cr	6,72	0,1488
tropicar	4,48	0,2231
malfal94	4,26	0,2345
icrg82	4	0,2502
yr_sch	3,82	0,2616
open6590	2,63	0,3800
soilsui1	1,97	0,5065
ciffob95	1,84	0,5434
landlock	1,74	0,5747
Mean VIF	4,46	
Notes: test for multicollinearity by analysing the main independent variables.		

References

Acemoglu, D. (2009). *Introduction to Modern Economic Growth.* Princeton, New Jersey: Princeton University Press.

Acemoglu, D., & Robinson, J. (2012). *Why Nations Fail: The origins of power, prosperity and poverty.* London: Profile Books Ltd.

Acemoglu, D., Gallego, F., & Robinson, J. (2014). Institutions, human capital, and development. *Annual Review of Economics*(6), 875-912.

Acemoglu, D., Johnson, S., & Robinson, J. (2001). The colonial origins of comparative development: An empirical investifation. *The American Economic Review, 91*(5), 1369-1401.

Ahlfeld, S., Hemmer, H., & Lorenz, A. (2005). The economic growth debate: geography versus institutions: Is there anything really new? *Discussion papers in development economics*(34).

Albouy, D. (2008, June). The colonial origins of comparative development: An investigation of the settler mortality data. *National Bureau of Economic Research*.

Angrist, J., & Pischke, J.-S. (2015). *Mastering 'Metrics: the path from cause to effect.* Woodstock: Princeton University Press.

Babbie, E. (2012). *The Practice of Social Research* (Thirteenth Interantional Edition ed.). Wadsworth: Cengage Learning.

Barro, R. (1996). Determinants of economic growth: a cross-country empirical study. *National Bureau of Economic Research*.

Barro, R. (2001, May). Human capital and growth. *The American Economic Review, 91*(2), 12-17.

Barro, R., & Lee, J. (2013). A new dataset of educational attainment in the world: 1950-2010. *Journal of Development Economics, 104*, 184-198.

Barro, R., & Lee, J.-W. (1994, January). *Data set for a Panel of 138 Countries*. Retrieved from nber publications: http://admin.nber.org/pub/barro.lee/ascii/readme.txt

Baumol, W., Litan, R., & Schramm, C. (2007). What drives economic growth. In *Good capitalism, bad capitalism, and the economics of growth and prosperity* (pp. 35-59). New Haven & London: Yale University PRess.

Bils, M., & Klenow, P. (2000, Dec). Does schooling cause growth? *The American Economic Review, 90*(5), 1160-1183.

Collier, P., & Gunning, J. (1999). Why has Africa grown slowly? *The Journal of Economic Perspectives, 13*(3), 3-22.

Costanza, R., Hart, M., Talberth, J., & Posner, S. (2009). Beyond GDP: The need for new measures of progress. (4).

Domar, E. (1946, April). Captial expansion, rate of growth, and employment. *Econometrica*, 137-147.

Easterly, W., & Levine, R. (2002). Tropics, germs, and crops: how endownments influence economic development. *National Bureau of Economic Research*.

Economist. (2014, September). The headwinds return. *The Economist*.

FoodAgricultureOrganization. (1995). *The digital soils map of the world.* Rome: United Nations. Retrieved from http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/

Frankel, J., & Romer, D. (1999). Does trade cause growth? *The American Economic Review, 89*(3), 379-399.

Gallup, J., Sachs, J., & Mellinger, A. (1999, August). Geography and Economic Development. *International Regional Science Review, 22*(2), 179-232.

Gallup, J., Sachs, J., & Mellinger, A. (2001, March). The geography of poverty and wealth. *Scientific America, 284*(3), 70-75.

Gaulier, G., Mirza, D., Turban, S., & Zignago, S. (2008). International transportation costs around the world: a new cif/fob rates dataset. *CEPII*.

Glaeser, E. (1993). Why does schooling generate economic growth. *Economic Letters: Elsevier*(44), 333-337.

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

Harrison, A. (1996). Openness and growth: A time series, cross-county analysis for developing countries. *Journal of Development Economics, 48*, 419-447.

Harrod, R. (1939, March). An essay in dynamic theory. *Economic Journal*, 13-33.

Knack, S., & Keefer, P. (1995, November). Institutions and economic performance: Cross-country test using alternative institutional indicators. *Economics and Politics, 7*(3), 207-227.

Knack, S., & Keefer, P. (1997, November). Does social capital have an economic payoff? A cross-country investigation. *The Quartely Journal of Economics, 112*(4), 1251-1288.

Koubi, V. (2005). War and Economic Performance. *Journal of Peace Research, 42*(1), 67-82.

Landes, D. (1998). *Wealth and Poverty of Nations: Why some are so rich and some so poor.* New York: W.W. Norton & Company.

Lawn, P. (2003, February). A theoretical foundation of support: the Index of Sustainable Economic Welfare (ISEW), Genuine Progress Indicator (GPI), and other related indexes. *Ecological Economics, 44*(1), 105-118.

Lee, K., Pesaran, M., & Smith, R. (1997). Growth and convergence in a multi-county empirical stochastic solow model. *Journal of Applied Econometrics, 12*(4), 357-392.

Maddison, A. (2001). *The World Economy: A millennial perspective.* OECD.

Mankiw, N., Romer, D., & Weil, D. (1992). A contribution to the empirics of economic growth. *Quartely Journal of Economics, 107*(2), 407-437.

Matthews, E. (1983). Global vegetation and land use: new high resolution database for climate studies. *Journal of Climate and Applied Meteorology, 22*, 474-487. Retrieved from https://www.pdx.edu/econ/jlgallup/country-geodata

Moffatt, M. (2017, March 26). *Are Wars Good for the Economy?* Retrieved from ThoughtCo.: https://www.thoughtco.com/are-wars-good-for-the-economy-1148174

Montesequieu, C. d. ([1748] (1989)). *The spirit of the laws.* New York: Cambrigde University Press. Retrieved from https://books.google.nl/books?hl=nl&lr=&id=Nr_3CwAAQBAJ&oi=fnd&pg=PT8&dq=montesquieu+the+spirit+of+law&ots=opIHtnowuQ&sig=PvYQC6Ype3jiZ6uS3Gp4Z-cMrA8#v=onepage&q=climate&f=false

North, D. (1990). *Institutions, Institutional Change and Economic Performance.* Cambridge: Camridge University Press.

North, D. (1991). Institutions. *The Journal of Economic Perspectives, 5*(1), 97-112.

North, D., & Thomas, R. (1973). *The rise of the western world: A new economic history.* Cambridge: Cambridge University Press. Retrieved from https://books.google.nl/books?id=MEh_54l-dK8C&printsec=frontcover&hl=nl#v=onepage&q&f=false

PortlandStateUniversity. (2017). *Country geography data.* College of Urban & Public Affairs: Economics, Portland. Retrieved from https://www.pdx.edu/econ/jlgallup/country-geodata

PRSGroup. (2017, May 29). *Methodologies: ICRG*. Retrieved from prsgroup web site: https://www.prsgroup.com/about-us/our-two-methodologies/icrg

Putman, R. (2000). Social capital: Measurement and consequences. Kennedy School of Government: Harvard University.

Sachs, J. (2003, Februari). Institutions don't rule: direct effects of geography on per capita income. *National bureau of economic research. NBER Working Paper No. 9490*.

Sachs, J. (2005). *The end of Poverty: Economic possibilities for our time.* New York: The Penguin Press.

Sachs, J., & Warner, A. (1995, August). Economic reform and the process of global integration. *Brookings Papers on Economic Activity*, 1-118.

Sachs, J., & Warner, A. (1997). Sources of slow growth in African economies. *Journal of African Economies, 6*(3), 335-376.

Sarkar, P. (1998, April). The catch-up debate: A statistical investigation. *The Helen Kellogg Institute for International Studies*(252).

Solow, R. (1994). Perspectives on growth theory. *The Journal of Economic Perspectives, 8*(1), 45-54.

Solow, R. (2007). The last 50 years in growth theory and the next 10. *Oxford Review of Economic Policy, 23*(1), 3-14.

StatisticsSolutions. (2017, June 1). *Assumptions of Linear Regression*. Retrieved from Statistics Solutions: http://www.statisticssolutions.com/assumptions-of-linear-regression/

Stockhammer, E., Hochreiter, H., Obermayr, B., & Steiner, K. (1997, April). The index of sustainable economic welfare (ISEW) as an alternative to GDP in measuring economic welfare: The results of the Austrian (revised) ISEW calculation 1955-1992. *Ecological Economics: Elsevier, 21*(1), 19-34.

Sykes, A. (1993). An Introduction to Regression Analysis. *Coase-Sandor Institute for Law & Economics Working Paper*(20).

Van Schaik, A. (2002). *Social capital in the European values study surverys.* Paris: OECD.

Wallace, P., & La Mond, A. (1977). Chapter 8: A dynamic theory of racial income differences. In G. Loury, *Women, Minorities & Employment Discrimination* (pp. 153-188). Lexington: Lexington Books. Retrieved from https://www.brown.edu/Departments/Economics/Faculty/Glenn_Loury/louryhomepage/papers/A%20Dynamic%20Theory.pdf