UNIVERSITETET I AGDER

An Econometric Analysis of Convergence

Econometric methods applied to the theory of macroeconomics and economic growth

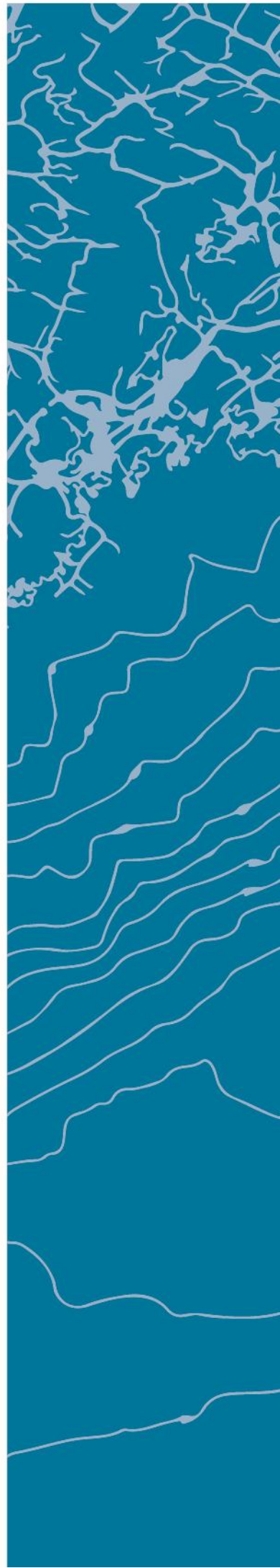JAN SEBASTIAN ROTHE

SUPERVISOR

Professor Jochen Jungeilges

University of Agder, 2018

School of Business and Law

PREFACE

The master thesis is strongly influenced by sound mathematical and statistical understanding that I gained during my bachelor program of Mathematical Finance at the University of Agder and additionally by courses in the field of econometrics and macroeconomics during my exchange period in Prague. The topic of my thesis was suggested to me by Professor Pavel Potužák of the University of Economics in Prague. Working on the thesis has been academically challenging as well as rewarding, and I have acquired a passionate interest in economic growth theory.

I would like to thank Professor Jochen Jungeilges for his excellent supervision and for promoting the program of Mathematical Finance which encouraged me to transition in 2013.

I would also like to thank my parents for their continuous support and encouragement.

Sebastian Rothe

Kristiansand, 01.06.2018

"The master-economist must possess a rare combination of gifts. He must reach a high standard in several different directions and must combine talents not often found together. He must be mathematician, historian, statesman, philosopher-in some degree. He must understand symbols and speak in words. He must contemplate the particular in terms of the general, and touch abstract and concrete in the same flight of thought. He must study the present in the light of the past for the purpose of the future." (Keynes, 1924)

ABSTRACT

This master thesis explores the concept of convergence in a macroeconomic perspective and applies econometric methods to economic growth theory.

Tests and analysis are performed using a dataset of national accounts from the rich database of The Penn World Tables version 9.0 and the statistical software Stata 15.1. Two sample selections are performed, with observations for 101 and 53 countries from 1970 to 2014.

The convergence classifications of β convergence, both absolute and conditional, as well as σ convergence are explained. The concepts of convergence are related to their respective research question. Do poorer economies tend to grow faster than richer economies? Do inequalities between poorer economies and richer economies tend to decrease? Do economies converge towards a common or unique steady state? Macroeconomic and economic growth theory is discussed and explained through neoclassical growth theory and new growth theory. The Solow model from neoclassical growth theory and the R&D model from new growth theory are mathematically derived and empirically tested to explore the dynamics of economic growth and to answer the question of the concept of absolute convergence. Other applied tests are growth-initial level regressions, which tests for β convergence, and standard deviation time series, which tests for σ convergence.

The research provides empirical evidence that poorer economies do tend to grow faster than richer economies, but with unreliable results due to issues of non-normality and heteroscedasticity. Empirical evidence also suggests that income dispersion of OECD countries is steadily increasing and that income dispersion of the full sample of 101 countries decreased from 1970 to 1988. The standard deviation time series test does not give a conclusive answer for the full sample after 1988. Due to issues of heteroscedasticity and autocorrelation, generalized least squares method is used to give the best linear unbiased estimator of the parameters of the Solow model. Empirical evidence show that capital's share is 60% and not 1/3 as the theory suggests. By adding human capital as in the theory of the augmented Solow model, empirical evidence shows a much lower capital's share of 20%. Individual heterogeneity suggests that countries follow unique paths to their own equilibrium level of economic growth given the parameters of the Solow model.

The resulting evidence from the conducted tests and analysis successfully provides satisfactory answers to the research questions of this master thesis.

CONTENTS

1 INTRODUCTION

Convergence is a concept of economic behavior in the theory of economic growth. The presence and empirical evidence of convergence has been greatly debated since the beginning of neoclassical growth theory. Many research papers found empirical evidence of absence of convergence and concluded that neoclassical growth theory was imperfect and should be rejected in favor of new growth theory. This motivated the start of theorizing and researching endogenous growth. However, neoclassical growth theory is still highly recognized and taught in academia of today, mainly due to its simplicity and the explanatory power of its parameters. This master thesis aims to apply econometric methods to the theory of macroeconomics and to gain insight in some of the shortcomings of economic growth theory. Studying economic growth is important to understand movements of the world income distribution and the welfare of individuals. The goal of economic growth research is to better understand the economic dynamics to enable pursuit of policies that increases standards of living and decreases world poverty.

1.1 RESEARCH QUESTIONS

The concept of convergence is associated with 3 research questions which again resembles different concepts of convergence. These are all interesting questions to analysts of convergence. The first question is a question of β convergence, the second question is a question of σ convergence and the third question is a question of absolute and conditional convergence.

1. Do poorer economies tend to grow faster than richer economies?
2. Do inequalities between poorer economies and richer economies tend to decrease?
3. Do economies converge towards a common or unique steady state?

1.2 RELEVANCE

Convergence has been widely researched for recent decades with diverging results. Different results have occurred due to variation in purpose and methodology used. This is because the question of convergence is interesting to both macroeconomic theorists and policy makers. Because of the magnitude of studies on the topic of convergence, it is helpful to be introduced to the convergence debate by the survey paper by Nazrul Islam (Islam, 2003). The survey paper briefly describes the different approaches to the study of convergences. The convergence debate started as a response to the neoclassical growth theory which was developed by Robert Solow

(Solow, 1956). A fundamental research paper that empirically addresses strengths and weaknesses of neoclassical growth theory is the research paper of Mankiw, Romer and Weil (Mankiw, Romer, & Weil, 1992). These two papers are included in two important textbooks of macroeconomic and economic growth theory by David Romer (D. Romer, 2012) and Barro and Sala-i-Martin (Barro & Sala-i-Martin, 2004).

1.3 STRUCTURE

The master thesis is structured in such a way that it should be perceived as both exploratory and descriptive research. The thesis seeks to describe advanced macroeconomic theory and econometric methods and to explore which econometric methods that are applicable to the questions of convergence. Some of the explored aspects might not be directly applied in the tests and analysis, but it provides an idea of how it could potentially be applied. The complexity of the theory explained varies which means that some aspects like averages and standard deviations are self-explanatory while matrix mathematics and stochastic processes requires a more advanced understanding.

Equations and mathematical derivations, called proofs, are generously used through most of the thesis. Graphs and regression outputs, including other test outputs in Stata, are provided in the chapter on tests and analysis. Equations, proofs, graphs and regression outputs are referenced where appropriate in the text. Equations and graphs are placed close to their reference while the proofs and regression outputs are placed in the appendix for convenience. The appendix also includes the Stata Do-file and the reflection notes.

The theory chapter "Economic growth theory" explaining what convergence is and the different concepts of convergence. The theory chapter briefly explains the role of neoclassical growth theory and new growth theory in the history of macroeconomic theory before technically and mathematically explaining two central models in detail, one from each theory.

The methodology chapter "Econometric methods" explains the mathematical statistics on which the econometric methods are created before explaining linear regressions, time series and panel data.

The chapter "Research approach" explains how the data is modified in preparation for conducting the tests and analysis.

2 ECONOMIC GROWTH THEORY

This chapter commences with the definition of the concept of convergence. Following that, the neoclassical growth theory, new growth theory and their relationship will be explained. Lastly, in separate subchapters, two specific models will be explained in detail and mathematically derived.

In mathematics, convergence is defined as an infinite series, a sum of infinite quantities of real numbers, that approaches a limit that can be expressed by a real number. A sequence is a collection of values of a variable which can be interpreted as a function or process of any natural number. The sequence is converging towards a convergent if the convergent is some constant that is equal to the limit of the function or process as the natural number goes to infinity (1). A series is an infinite summation of the values of a sequence and is converging if the sum is equal to some constant (2). If the values in the sequence are the same as for the series that converges then the convergent of the sequence is equal to zero (3). (Lorentzen, Hole, & Lindstrøm, 2010, p. 306-307, 314, 341)

$$\lim_{n \to \infty} x_n = c \tag{1}$$

$$\sum_{n=1}^{\infty} a_n = S \tag{2}$$

$$\lim_{n \to \infty} a_n = \lim_{n \to \infty}(S_n - S_{n-1}) = 0 \tag{3}$$

A series converges either conditionally or absolute (also called unconditional). The difference between absolute and conditional convergence is that taking the absolute value for each value in a conditional converging series will cause the series to diverge. On the contrast, doing this for each value in an absolute converging will not cause the series to diverge, the series will still be converging. This is because for an alternating series the sum of the positive values and the negative values is positive and negative infinity. (Lorentzen et al., 2010, p. 361)

In economics, the question of convergence explores the dynamics of growth of economies. Convergence is distinguished between multiple classifications. The classical classification is between β and σ convergence. β convergence is either absolute or conditional. Absolute convergence is a necessary, but not sufficient, condition for σ convergence which means that for an economy that is converging in σ is also converging absolute. (Sala-i-Martin, 1996, p. 1019-1020)

There is presence of β convergence if economies with lower initial levels of economic output grow faster than economies with higher initial levels of economic output. β convergence is typically tested by a growth-initial level regression where a negative value of the coefficient of β in the growth-initial level regression implies the presence of β convergence. If poor economies tend to grow faster per worker than rich economies without being conditioned on some other characteristic, then there is absolute convergence. If the growth rate of an economy is positively related to its distance from its steady state, then there is conditional convergence. In absolute convergence, all economies approach the same level of equilibrium. While in conditional convergence, all economies approach their own unique level of equilibrium. Another type of conditional convergence is club convergence, which is when economies approach similar levels of equilibrium if they are similar in terms of characteristics. However, it is difficult to distinguish between club convergence and conditional convergence empirically. (Islam, 2003, p. 315; Sala-i-Martin, 1996, p. 315)

There is presence of σ convergence if the dispersion of economies' real GDP per worker tends to decrease over time. The dispersion of real GDP per worker measures the development of distribution of income across countries and is statistically measured by standard deviation which is denoted by σ. (Sala-i-Martin, 1996, p.1020)

In modern macroeconomic theory, the neoclassical growth theory and new growth theory are the most recognized for explaining dynamics of economic growth. Neoclassical growth theory revolves around the contribution of Solow and Swan in 1956 (Solow, 1956). The Solow model (also called Solow-Swan model) specifies a production function that assumes constant returns to scale, diminishing returns to each input and some positive smooth elasticity of substitution between the inputs. The Solow model assume that savings rate, population growth and technological progress occurs outside of the model. The dependency on exogenous growth is a major weakness of the Solow model, despite causing a strongly admired simplicity in explaining economies and their dynamics. (Barro & Sala-i-Martin, 2004, p. 17)

A fundamental equation of the Solow model explains that economies with lower capital per worker tend to grow faster. This equation suggests that there is absolute convergence which has been empirically tested and shown to not be the case. Convergence in the Solow model has been empirically shown to be conditional, meaning that economies have their own steady state and that the distance from the steady state depends on some unobserved economic characteristics. The Solow model predicts a capital share which implies a speed of convergence that is too high to be realistic. To decrease the capital share to get a more appropriate capital

share is to include the concept of human capital. This gives the augmented Solow model. (Barro & Sala-i-Martin, 2004, p. 17)

New growth theory aims to explain long-term growth by endogenous growth models. Endogenous growth models assume non-diminishing constant returns to capital and labor and distinguish between physical and human capital. Paul M. Romer introduced such a model called the research and development (R&D) model (P. M. Romer, 1990).

The R&D model was developed in early 1990s to divide resources allocated between two sectors, the sector of output production and the sector of research and development. The equation for the sector of output production assumes constant returns to capital and labor. The equation for the sector of research and development does not assume constant returns to capital and labor. There is no restriction on the effect of the stock of knowledge on production of innovative ideas. This allows the possibility of increasing, constant and diminishing returns in the research and development sector. In case of increasing returns, past knowledge makes future ideas easier to accomplish. In the other case of decreasing returns, the easiest discoveries are made first, and innovative ideas are increasingly difficult to produce. (D. Romer, 2012, p. 103-104)

> *It has been generally thought that convergence was an implication of the neoclassical growth theory, while the new growth theories did not have this complication. (Islam, 2003, p. 309)*

The economic growth in the R&D model is either semi-endogenous or fully endogenous. In the case of semi-endogenous growth, the technological progress and capital growth rate converge to their equilibrium level where their respective growth rates, the growth rate of growth rate, are equal to zero. The long-run growth is an increasing function of population growth and parameters of the knowledge production function. In the case of fully endogenous growth, there is zero population growth and the growth rates of capital and knowledge are constant. In this case, the equilibrium that the growth rates of the economy are converging towards is unknown. The equilibrium depends on parameters that are difficult to derive and even more difficult to interpret. The fraction of labor force and capital stock used in research and development are among these parameters that affect the long-run growth. (D. Romer, 2012, p. 10)

2.1 THE SOLOW MODEL

In this subchapter, the Solow model is explained in greater detail and derived mathematically.

The Solow model proposes a production function consisting of four variables, the total output of the economy Y explained by capital K, labor L and knowledge A. All variables are functions of time t (1.1). (D. Romer, 2012, p. 10)

$$Y(t) = F(K(t), A(t)L(t)) \tag{1.1}$$

The production function holds two key features that imply that the ratio of capital to output will not show any positive or negative trend in the long run. First feature is that time is only affecting the output through the inputs of the function. Second feature is that the functions for knowledge and labor is multiplied, where the product of the two is referred to as effective labor. The knowledge in this composition of inputs is called labor-augmenting (also called Harrod-neutral). Other compositions of knowledge in the production function are called capital-augmenting (1.2) and Hicks-neutral (1.3). (D. Romer, 2012, p. 10)

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

$$Y(t) = A(t)F(K(t), L(t)) \tag{1.3}$$

A comprehensive assumption of the production function is constant returns to scale. Constant returns to scale is when capital and effective labor are multiplied by a positive constant c, and the expression is then equal to the composition of output multiplied by c (1.4). (D. Romer, 2012, p. 11)

$$F(cK(t), cA(t)L(t)) = cF(K(t), A(t)L(t)) \tag{1.4}$$

The assumption of constant returns to scale can be described as a combination of two lesser assumptions. The first assumption is that the multiplication by c does not change the composition of the function. This assumption state that that all advantages from divisions of labor have been exhausted which rules out Smiths' famous prediction of an increasing productivity from specialization. This assumption does not hold in cases of smaller economies where an increase in capital and effective labor causes the composition of output to change and causes a higher increase in output than the increase of capital and effective labor. (D. Romer, 2012, p. 11)

The second assumption is that other factors such as land and other natural resources are unimportant and does not affect the growth of the economy. This assumption state that land or

other resources are not as important as effective labor which rules out Malthus' famous prediction of that population growth is exponential and will eventually exceed the growth of the production of necessary resources which is arithmetic. (D. Romer, 2012, p. 11)

If the assumption of constant returns to scale holds, then the production function can be transformed to its' intensive form. The intensive form of the production function is derived by dividing the output and other factors by effective labor. From the assumption of constant returns to scale, the constant is set to be equal to 1 divided by effective labor. This gives output per effective worker as a function of capital per effective worker (1.5) (see Appendix: Proof 1). (D. Romer, 2012, p. 11)

$$y = f(k) \tag{1.5}$$

The intensive form of the production function (1.5) follows a set of assumptions. These include that the marginal product of capital is always positive but declines as capital per effective worker rises. Also, that if the capital per effective worker is equal to zero, the output per effective worker would also be zero. (D. Romer, 2012, p. 12)

$$f'(k) > 0 \tag{1.6}$$

$$f''(k) < 0$$

$$f(0) = 0$$

The Inada conditions are additional assumptions of the intensive form of the production function and assure that the path of the economy converges (Inada, 1963). The Inada conditions state that the marginal product of capital is infinitely large for an infinitely small capital per effective worker and that the marginal product is infinitely small for an infinitely large capital per effective worker. (D. Romer, 2012, p. 12)

$$\lim_{k \to 0} f'(k) = \infty \tag{1.7}$$

$$\lim_{k \to \infty} f'(k) = 0$$

The Cobb-Douglas production function is a commonly used and simple to analyze production function (1.8). It was developed by Charles W. Cobb and Paul H. Douglas in 1928 (Cobb & Douglas, 1928). The Cobb-Douglas production function with labor augmenting technological progress is represented as the total output explained by the capital powered by the capital share multiplied with knowledge and labor powered by 1 minus capital share. Capital share α is a

positive percentage. The Cobb-Douglas production function holds for all assumptions (see Appendix: Proof 2). (D. Romer, 2012, p. 12-13)

$$Y(t) = K(t)^\alpha (A(t)L(t))^{1-\alpha} \qquad (1.8)$$

Growth rates of a variable in the model refers to proportional rate of change, the derivative of the variable with regards to time, denoted with a dot above the variable, divided by the variable. The growth rate of labor and knowledge are given by the constant exogenous parameters population growth and technological progress, respectively. The assumption that labor and knowledge grow exponentially can be shown by solving the differential equations (1.9) (see Appendix: Proof 3). (D. Romer, 2012, p. 13-14)

$$\dot{L}(t) = nL(t) \qquad (1.9)$$

$$\dot{A}(t) = gA(t)$$

$$L(t) = L(0)e^{nt} \qquad (1.10)$$

$$A(t) = A(0)e^{gt}$$

The law of motion for capital explains that net investment, is equal to gross investment minus depreciation. The change in capital is equal to investment minus depreciated capital (1.11) (see Appendix: Proof 4). In the Solow model total savings is equal to gross investment in the long-run perspective and output is saved at an exogenous and constant rate s and capital depreciates at a rate δ. (D. Romer, 2012, 13-14)

$$\dot{K}(t) = sY(t) - \delta K(t) \qquad (1.11)$$

In the Solow model, the behavior of the economy is explained by the exogenous variables labor and knowledge, and the endogenous variable capital. The dynamics of capital per effective worker is derived from the equation of law of motion using the chain rule (1.12) (see Appendix: Proof 5). (D. Romer, 2012, p. 15-16)

$$\dot{k}(t) = sy(t) - (\delta + n + g)k(t) \qquad (1.12)$$

The growth rate of capital per effective worker converges to zero which is when the actual investment is equal to break-even investment. The steady state in the Solow model is a long-run equilibrium level that the economy converges towards. The equilibrium level is dependent on savings rate, population growth, technological growth, depreciation rate and capital share (1.13) (see Appendix: Proof 6). (D. Romer, 2012, p. 16-17)

$$k^* = \left(\frac{s}{n + g + \delta}\right)^{\frac{1}{1-\alpha}} \tag{1.13}$$

The Solow model implies that the parameter that is most important for economic growth is the savings rate. An increase in the savings rate will increase the actual investment and therefore increase the steady state level of output. The growth of capital per effective worker will then be positive until the new steady state is reached. The effect that an increase of the savings rate has on the long-run output of the Slow model can be derived by the elasticity of steady state output per effective worker to savings rate (1.14) (see Appendix: Proof 7). (D. Romer, 2012, p. 18)

$$E_{y^*/s} = \frac{\alpha}{1 - \alpha} \tag{1.14}$$

The speed of which the economy reaches its steady state is called the speed of convergence. The speed of convergence λ is measured by how quickly capital per effective worker moves to its steady state value (1.15) (see Appendix: Proof 8). (D. Romer, 2012, p. 25-26)

$$\lambda = (1 - \alpha)(n + g + \delta) \tag{1.15}$$

Convergence in the Solow model is assumed to be absolute, that all economies converges to the same steady state. This suggests a catch-up phenomenon where poorer economies grow faster than richer economies and hence catch-up in the long run. (D. Romer, 2012, p. 32)

The augmented Solow model includes another process of growth and distinguishes between physical capital K and human capital H (1.16). Human capital is measured by the total amount of productive services supplied by workers. The Cobb-Douglas production function suggested by the augmented Solow model can be transformed into intensive form in the same way as the previous production function because the assumption of constant returns to scale (1.17) (see Appendix: Proof 9 & Proof 10). (D. Romer, 2012, p. 16-17)

$$Y(t) = K(t)^{\alpha} H(t)^{\beta} (A(t)L(t))^{1-\alpha-\beta} \tag{1.16}$$

$$y(t) = k(t)^{\alpha} h(t)^{\beta} \tag{1.17}$$

The savings rates for physical and human capital per effective worker, s_k and s_h, are exogenous and constant. Further, the equations for the dynamics of physical and human capital per effective worker are explained by growth of physical and human capital per effective worker being equal to actual investment minus break-even investment (1.18) (see Appendix: Proof 11). (Barro & Sala-i-Martin, 2004, p. 59)

$$\dot{k}(t) = s_k y(t) - (n + g + \delta)k(t) \tag{1.18}$$

$$\dot{h}(t) = s_h y(t) - (n + g + \delta)h(t)$$

The augmented Solow model assumes diminishing returns to all capital which means that in the steady state the growth of physical and human capital per effective worker is equal to zero. Also, for both physical and human capital per effective worker in the steady state, the actual investment is equal to break-even investment. Steady state levels of capital per effective worker are dependent on two parameters in addition to those utilized in the Solow model, savings rate for human capital per effective worker s_h and human capital share β (1.19) (see Appendix: Proof 12). (Barro & Sala-i-Martin, 2004, p. 60)

$$k^* = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n + g + \delta} \right)^{\frac{1}{1-\alpha-\beta}} \tag{1.19}$$

$$h^* = \left(\frac{s_k^{\alpha} s_h^{1-\alpha}}{n + g + \delta} \right)^{\frac{1}{1-\alpha-\beta}}$$

Speed of convergence in the augmented Solow model can be derived from the growth rate of output per effective worker explained by the weighted average growth rate of physical and human capital per effective worker (1.20) (see Appendix: Proof 13). (Barro & Sala-i-Martin, 2004, p. 60-61)

$$\lambda = (1 - \alpha - \beta)(n + g + \delta) \tag{1.20}$$

The augmented Solow model solves some issues of the Solow model by suggesting that there is conditional convergence. Conditional convergence is present when each country converges to its own unique steady state depending on some other characteristic and if conditioned for this other characteristic then all countries would converge to the same steady state. In the case of the augmented Solow model, this other characteristic is human capital and if conditioned for human capital all countries would converge to the steady state of the Solow model's parameters. (Sala-i-Martin, 1996, p. 1027)

2.2 THE RESEARCH AND DEVELOPMENT MODEL

In this subchapter, the research and development (R&D) model of new growth theory will be explained in greater detail and mathematically derived.

The R&D model is an endogenous growth model proposed by D. Romer as a simplified model involving developments of P. Romer, Grossman and Helpman, and Aghion and Howitt (Aghion & Howitt, 1992; Grossman & Helpman, 1991; P. M. Romer, 1990). The R&D allocates resources into two sectors, the goods producing sector (2.1) and the knowledge producing sector (2.2). The shares of labor force and capital stock in the knowledge producing sector are a_L and a_K. Hence the share of labor force and capital stock in the goods producing sector is given by the respective remaining shares. Both shares are exogenous and constant. (D. Romer, 2012, p. 103)

$$Y(t) = \big((1 - a_K)K(t)\big)^{\alpha}\big(A(t)(1 - a_L)L(t)\big)^{1-\alpha} \tag{2.1}$$

$$\dot{A}(t) = B\big(a_K K(t)\big)^{\beta}\big(a_L L(t)\big)^{\gamma} A(t)^{\theta} \tag{2.2}$$

The savings rate in the R&D model, as in the Solow model, is exogenous and constant. The capital growth rate and technological progress is explained by g_K and g_A. To explain the dynamics of the economy in this model the growth rates of growth rates are derived (2.3) (see Appendix: Proof 14). (D. Romer, 2012, p. 104)

$$\frac{\dot{g}_K(t)}{g_K(t)} = (1 - \alpha)(g_A(t) + n - g_K(t)) \tag{2.3}$$

$$\frac{\dot{g}_A(t)}{g_A(t)} = \beta g_K(t) + \gamma n + (\theta - 1)g_A(t)$$

In equilibrium of the R&D model the growth rates of growth rates are equal to zero which predicts a steady growth in the long-run (2.4) (see Appendix: Proof 15). (D. Romer, 2012, p. 113-114)

$$g_K^* = g_A^* + n \tag{2.4}$$

$$g_A^* = \frac{\beta + \gamma}{1 - \theta - \beta} n$$

The long-run growth rate of output in the R&D model is converging to the same constant as the long-run growth rate of capital (2.5) (see Appendix: Proof 16). If the sum of knowledge and capital share is restricted under 1 (a hundred percent) then the model shows semi-endogeneity.

Then the long-run growth rate depends on the population growth and for a population growth of zero, there will also be zero growth rate of output. In the alternative case, where the sum of knowledge and capital share is equal to 1 and there is zero population growth then the growth rate of capital growth rate is equal to the growth rate of technological progress and the long-run growth is difficult to analyze. (D. Romer, 2012, p. 113-114)

$$g_Y^*(t) = g_K^*(t) = n \left(\frac{1 + \gamma - \theta}{1 - \theta - \beta} \right) \tag{2.5}$$

The equilibrium level of growth in the R&D model can explain persistent and increasing inequality between countries, thereby allowing economies to diverge.

3 ECONOMETRIC METHODS

This chapter explains econometric methods, from basic concepts of mathematical statistics to more complex concepts of linear regression, time series and panel data analysis.

Econometric methods are defined as the use of econometric models to understand quantitative data in economics and to achieve empirical evidence to economic theory. Econometric models are created by the application of mathematical statistics. Quantitative data are large collections of observations of a sample of a population.

3.1 MATHEMATICAL STATISTICS

This subchapter derives elements of mathematical statistics that are considered most relevant to econometric methods. These elements are mainly visual techniques and numerical summary measures from descriptive statistics and estimators and hypothesis testing from inferential statistics. Other elements explained are sample selection, variables and probability density functions.

The population is everyone that is relevant to what is researched and is often difficult to observe in its entirety. Therefore, a sample is used as convenience. Collecting the sample data using proper techniques is important for the sample to be representative of the population. Improper techniques might lead to the sample being different from the population which would give biased results. Selection bias occurs when the observed values differ in characteristics that influence the selection of the sample. If selection is random then there is no selection bias. Another method for avoiding selection bias is to use stratified sampling which entails separating the population into groups that are not overlapping in an observed characteristic. This method

avoids groups to be overestimated or underestimated in the full sample. However, it is still important to properly sample each group of the population. (Devore & Berk, 2012, p. 7)

A characteristic that is observed in the data is called a variable and is measured for each object or individual in the sample. The data is either univariate, bivariate or multivariate depending on how many variables that are included in the data. The variables are measured in numerical, categorical or string values. The variables in the sample are random if they for every outcome in the sample can be associated with a number. If the variable is random, it can then be defined as either discrete or continuous. A discrete random variable can only take on possible values in a defined set of outcomes. A random variable however, can take on any real number in an infinitely precise measure and the possibility for one exact value is equal to zero. (Devore & Berk, 2012, p. 3, 99)

Descriptive statistics aims to summarize and describe the data that is collected. Descriptive methods involve visual techniques and numerical summary measures. Numerical summary measures involve means, standard deviations and correlation coefficients which present locational properties of the data. The mean is the arithmetic average of a random variable and is called the sample mean when calculated for the sample (3.1). (Devore & Berk, 2012, p. 3-4, 24-25)

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{1}{N} \sum_{n=1}^{N} x_n \qquad (3.1)$$

The mean is highly affected in case there are extreme values for some observations. An alternative locational measure that is not affected by extreme values is the median. The sample median is either the middle value of all sorted values when the number of values is odd or the average of the two middle values for the sorted values if the number of values is even. Difference between calculated values for the mean and median is caused by skewness in the distribution of observed values. If there is no skewness, the mean and median are equal. (Devore & Berk, 2012, p. 27-28)

Standard deviation measures variability in the sample data and is measured by deviations from the mean. Deviations from the mean will be both negative and positive and will equal to zero after being summed. To avoid the effects of negative deviations, variance of the sample data is calculated first and then the standard deviation is calculated by the square root of the variance (3.2). (Devore & Berk, 2012, p. 32-35)

$$\widehat{\sigma_x^2} = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \bar{x})^2 \tag{3.2}$$

Both the mean and variance are important to explain the distribution of the observed values for a variable. Skewness is used to describe the lack of symmetry in the distribution of observations (3.3). The distribution shows the characteristic of a left-hand side tail for a negative skewness and a right-hand side for a positive value. (Devore & Berk, 2012, p. 121)

$$\widehat{S_x} = \frac{1}{(N-1)\widehat{\sigma_x^3}} \sum_{n=1}^{N} (x_n - \bar{x})^3 \tag{3.3}$$

Kurtosis is a measure for the relative quantity that is found within the tail(s) of the distribution (3.4). Values of kurtosis higher than 3 would imply that most of the observed values are found within that tail(s).

$$\widehat{K_x} = \frac{1}{(N-1)\widehat{\sigma_x^4}} \sum_{n=1}^{N} (x_n - \bar{x})^4 \tag{3.4}$$

The covariance is a measure of variability between two dependent random variables and is used to describe the strength of linear relationship between the two (3.5). A positive covariance signifies a positive linear relationship while a negative covariance signifies a negative linear relationship. A covariance close to zero signify that the two variables do not have a linear relationship while a covariance equal to positive or negative 1 signifies that there is positive or negative perfect linear relationship, respectively. (Devore & Berk, 2012, p. 247-249)

$$\widehat{C_{x,y}} = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \bar{x})(y_n - \bar{y}) \tag{3.5}$$

The concept of correlation coefficients was introduced by Francis Galton in 1888 and describes the strength of linear relationship between two variables (Galton, 1888) (3.6). If the variables are perfectly linearly related, then the coefficient takes a value of minus or positive 1. A coefficient between would signify that their relationship is not perfectly linear. (Devore & Berk, 2012, p. 249-250)

$$\widehat{\rho_{x,y}} = \frac{1}{(N-1)\widehat{\sigma_x}\widehat{\sigma_y}} \sum_{n=1}^{N} (x_n - \bar{x})(y_n - \bar{y}) \tag{3.6}$$

Visual techniques involve graph-based diagrams such as histograms and scatter plots. Histograms counts the frequency and then the density which is also called the relative frequency. The frequency is the number of times the same value occurs for a variable while the

density is the number of times the value occurs divided by the total number of observations of the variable in the dataset. The histogram then visualizes either the frequency or the density by bars. (Devore & Berk, 2012, p. 12-13)

Scatter plots uses coordinates of values for two variables and are useful for inference of the relationship between the two chosen variables. Scatter plots can show whether the relationship between the two variables is linear, exponential or polynomial. If the two variables follow a linear relationship, then the scatterplots show either decreasing or increasing one-to-one coordinates. If the two variables follow an exponential relationship, then there will be an increasing number of coordinates and variability for higher values. This could help determine the need for logarithmic transformation of variables. (Devore & Berk, 2012, p. 615-617)

The process of generalizing and analyzing the sample to draw reasonable conclusions of the population is called inferential statistics. Inferential statistics involves creating estimates and interval estimates using procedures such as point estimations, hypothesis testing and confidence intervals. The point estimate is the point in the sample that is best at explaining the true parameter of the population. For the average of the population, the parameter is the mean μ and is estimated by the point estimate which is the sample mean. (Devore & Berk, 2012, p. 332)

Estimators are the formulas and rules that are being used to calculate the estimate, usually shown by a denotation. Estimators are said to give the true parameter of the population plus some error of estimation (3.7). The quality of an estimator is measured by its unbiasedness, consistency and efficiency, which is measured by the error ϵ. (Devore & Berk, 2012, p. 334-335)

$$E[X] = \bar{x} + \epsilon \quad Kurt[X] = \widehat{K_x} + \epsilon \tag{3.7}$$

$$Var[X] = \widehat{\sigma_x^2} + \epsilon \quad Cov[X,Y] = \widehat{C_{x,y}} + \epsilon$$

$$Skew[X] = \widehat{S_x} + \epsilon \quad Corr[X,Y] = \widehat{\rho_{x,y}} + \epsilon$$

A hypothesis is an empirically testable research question and consists of a null hypothesis and one or more alternative hypotheses. The null hypothesis is a statement that something is true while the alternative hypothesis contradicts this statement. Through an empirical test there is only two possible outcomes, the null hypothesis is either rejected or failed to reject. The hypothesis testing procedure consists of specifying the test statistic and the rejection region. The null hypothesis is rejected if the estimated test statistic falls within the specified rejection region. A badly specified rejection region may result in type I error, rejecting the null hypothesis

when it is true, or type II error, failing to reject the null hypothesis when it is false. (Devore & Berk, 2012, p. 426-429)

The level of significance is the probability of type I error that is allowed in the hypothesis testing and the P-value is the probability of getting the same or greater value calculated by the test statistic given that the null-hypothesis is true. If the P-value is lower than the significance level, then the null hypothesis is rejected. If the P-value is greater than the significance level, then the null hypothesis cannot be rejected. The P-value can also be referred to as the lowest acceptable significance level for the null hypothesis to be rejected. (Devore & Berk, 2012, p. 456-459)

The probability that a continuous random variable will take on a value within a specific interval can be explained by the integral of the continuous random variable's probability density function (3.8). An important probability density function is the normal distribution (3.9). (Devore & Berk, 2012, p. 160, 179)

$$P(a \leq X \leq b) = \int_a^b f(x)dx \tag{3.8}$$

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.9}$$

The central limit theorem states that for any population that is normally distributed, the arithmetic average will also be normally distributed for any sample size. Also, if the population is not normally distributed, the distribution averages for different samples will be more normally distributed than the distribution for the population. Therefore, for a large sample size the arithmetic average of the population will be asymptotically normal. (Devore & Berk, 2012, p. 298)

Commonly used test statistics are Z, T, χ^2 and F. The rejection region defines values of the test statistic of which the null hypothesis is rejected. The rejection region is the area under the curve of the probability density function and is either upper tailed, lower tailed or two-tailed. The boundaries of the rejection region are determined by the significance level of the test. (Devore & Berk, 2012, p. 428)

The Z-statistic follows a standard normal probability density function (3.10). By the central limit theorem, the Z-statistic require a sample size larger than 30. The probability of Z for the population being equal or less than the test statistic is given by the cumulative distribution

function (3.11). The Z-statistic can be calculated, and the p-value can be found using a program or by checking a table for the standard normal curve areas (3.12). (Devore & Berk, 2012, p. 181)

$$f(z; 0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{3.10}$$

$$\Phi(z) = P(Z \geq z) = \int_{-\infty}^{z} f(x; 0,1) dx \tag{3.11}$$

$$z = \frac{\bar{x} - \mu_0}{\widehat{\sigma_x}/\sqrt{N}} \tag{3.12}$$

The T-statistic is used when there is less than or equal to 30 observations in the sample. The T-statistic follows a Student's T probability density function with ν degrees of freedom (3.13). The gamma function is an infinite integral of a positive value α with only positive values (3.14). The T-statistic has N minus 1 number of degrees of freedom. The Z-statistic and T-statistic are estimated in similar fashion (3.15). (Devore & Berk, 2012, p. 320-321)

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{3.13}$$

$$\Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1} e^{-x} dx \tag{3.14}$$

$$t = \frac{\bar{x} - \mu_0}{\widehat{\sigma_x}/\sqrt{N}} \tag{3.15}$$

A random variable has a chi-squared distribution with parameter ν for number of degrees of freedom if the probility density function is a function of the gamma density and has only positive values (3.16). The chi-squared statistic is estimated by summing all cells of the table where the observed frequency minus the expected frequency squared is divided by the expected frequency (3.17). The null hypothesis is rejected if the estimated chi-squared is larger than $\chi^2_{\alpha,\nu}$. (Devore & Berk, 2012, p. 318)

$$f(x; \nu) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \tag{3.16}$$

$$\chi_\nu^2 = \nu \frac{\widehat{\sigma^2}}{\sigma^2} \tag{3.17}$$

A random variable that follows a F-distribution has a probability density function with gamma functions, two numbers of degrees of freedom for two independent chi-squared distributed random variables and only positive values (3.18). The F-statistic is estimated from two independent chi-squared random samples with number of degrees of freedom equal two each samples number of observation minus one (3.19). For a value higher than $F_{\alpha,\nu1,\nu2}$, the null hypothesis is rejected. (Devore & Berk, 2012, p. 323)

$$f(x; \nu_1, \nu_2) = \frac{\frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1}}{\left(1 + \frac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}} \tag{3.18}$$

$$F_{\nu_1,\nu_2} = \frac{\nu_2 \chi_{\nu_1}^2}{\nu_1 \chi_{\nu_2}^2} = \frac{\widehat{\sigma_1^2}\sigma_2^2}{\widehat{\sigma_2^2}\sigma_1^2} \tag{3.19}$$

The T-, χ^2- and F-statistic can all be explained by a sequence of independent standard normal random variables (3.20). (Devore & Berk, 2012, p. 325)

$$\chi_\nu^2 = Z_1^2 + Z_2^2 + \cdots + Z_\nu^2 = \sum_{n=1}^{\nu} Z_n^2 \tag{3.20}$$

$$T_\nu = \frac{Z_{\nu+1}}{\sqrt{\frac{Z_1^2 + Z_2^2 + \cdots + Z_\nu^2}{\nu}}} = \frac{Z_{\nu+1}}{\sqrt{\frac{1}{\nu}\sum_{n=1}^{\nu} Z_n^2}}$$

$$F_{\nu_1,\nu_2} = \frac{\nu_2 \sum_{n=1}^{\nu_1} Z_{n+\nu_2}^2}{\nu_1 \sum_{n=1}^{\nu_2} Z_n^2}$$

3.2 LINEAR REGRESSIONS

In this subchapter, the linear regression model will be explained by ordinary least squares, the Gauss-Markov theorem and goodness-of-fit measures.

The linear regression model aims to find evidence for a linear relationship between a dependent variable y, called the regressand, and independent variables x_n, called regressors (4.1). By using data from the sample, the model estimates the parameters of the population β_n, called regression coefficients. The error of estimation is given by the error term ε_n. The linear regression model

can also be written in matrix form where y and ε are N-dimensional vectors, the β is a M-dimensional vector and X is a matrix of N×M dimension (4.2). (Verbeek, 2012, p. 12-15)

$$y_n = \beta_1 + \beta_2 x_{n,2} + \beta_3 x_{n,3} + \cdots + \beta_M x_{n,M} + \epsilon_n \qquad (4.1)$$

$$y = X\beta + \epsilon \qquad (4.2)$$

In the sampling process, by stating that every new sample will give the same X matrix, it is assumed that each independent variable is deterministic, which means that they are fixed and non-stochastic. However, this assumption is only perfectly true in laboratory experiments. (Verbeek, 2012, p. 13)

Ordinary least squares (OLS) is an approach to minimize the sum of squared approximation errors which gives the best linear approximation of a random variable. The sum of squared approximation errors can be written as a function of the coefficients (4.3). The formulae for best linear approximation of the coefficients is found by minimizing the function (see Appendix: Proof 17) (4.4). (Verbeek, 2012, p. 7-9)

$$f(\beta) = (y - X\beta)'(y - X\beta) \qquad (4.3)$$

$$\hat{\beta} = (X'X)^{-1}X'y \qquad (4.4)$$

The Gauss-Markov theorem was developed by Carl Friedrich Gauss and Andrey Markov and state under which conditions the OLS estimator is a good estimator for the true unknown parameter of the population. The first assumption says that the expected value of the error term is zero, which is an assumption for unbiasedness. The second assumption is that the error terms and independent variables are independent. The third assumption is that all error terms have constant variance, which means that there is homoscedasticity. The fourth and last assumption says that there is zero correlation between the error terms, which means that there is no autocorrelation. The first, third and fourth assumption together state that the error terms are uncorrelated drawings from a normal distribution with zero mean and σ^2 in constant variance. (Verbeek, 2012, p. 15)

The Gauss-Markov theorem can be written in matrix form where I is an identity matrix of N×N dimension (4.5). The OLS estimator holds for these assumptions (see Appendix: Proof 18). If for a test result all Gauss-Markov assumptions hold then the estimator is said to be the best linear unbiased estimator (BLUE). (Verbeek, 2012, p. 15-17)

$$E[\epsilon|X] = E[\epsilon] = 0 \qquad (4.5)$$

$$Var[\epsilon|X] = Var[\epsilon] = \sigma^2 I$$

The Gauss-Markov assumption for normality and homoscedasticity can be tested by residual diagnostics after a linear regression model is estimated. A standardized normal probability plot can be used to determine the distribution of residuals relative to a normal distribution (D'Agostino & Belanger, 1990) and a residual versus fitted values scatterplot can be used to determine the variance of residuals.

For an estimated linear regression model, it is of interest to measure how well the model fit the observed values. A common measure for goodness-of-fit is called the R-squared, which measures how much of the variance of the observations that is explained by the model. The R-squared takes a value equal to or between 1 and 0, where 1 means that the model fits perfectly to the observed values and 0 means that the model does not explain any of the variations in the observed values. There are several ways of measuring the R-squared. The straight-forward way is to estimate the average of sum of squared differences between the estimated values and the arithmetic average divided by the average of the sum of squared differences between observed values and the arithmetic average (4.6). Another way of measuring the R-squared can be derived as the remaining percentage of variance of the observed values that are unexplained in the residuals (4.7) (see Appendix: Proof 19). (Verbeek, 2012, p. 20-21)

$$R^2 = \frac{\widehat{\sigma_{\widehat{y}_n}^2}}{\widehat{\sigma_{y_n}^2}} = \frac{\frac{1}{N-1}\sum_{n=1}^{N}(\widehat{y}_n - \bar{y})^2}{\frac{1}{N-1}\sum_{n=1}^{N}(y_n - \bar{y})^2} \qquad (4.6)$$

$$R^2 = 1 - \frac{\widehat{\sigma_{e_n}^2}}{\widehat{\sigma_{y_n}^2}} = 1 - \frac{\frac{1}{N-1}\sum_{n=1}^{N}e_n^2}{\frac{1}{N-1}\sum_{n=1}^{N}(y_n - \bar{y})^2} \qquad (4.7)$$

For models with intercept, these two formulas give identical results. On the other hand, in the absence of an intercept the two formulas will give different results. In this case it is useful to use another alternative formula, which measures the uncentered R-squared (4.8). The uncentered R-squared is in most cases higher than the standard measures. (Verbeek, 2012, p. 21)

$$R_{uncentered}^2 = \frac{\sum_{n=1}^{N}\widehat{y}_n^2}{\sum_{n=1}^{N}y_n^2} = 1 - \frac{\sum_{n=1}^{N}e_n^2}{\sum_{n=1}^{N}y_n^2} \qquad (4.8)$$

For models with many regressors, the R-squared will be higher because of more regressors alone, even if the additional regressors have no real explanatory power. Adjusted R-squared is a measure that corrects the variance estimates in the standard R-squared for the degrees of freedom (4.9). The adjusted R-squared is always smaller than the standard R-squared unless the model consists of only an intercept, the number of degrees of freedom is equal to 1. The adjusted R-squared is not restricted to the same interval of the standard R-squared. Therefore, for a high number of degrees of freedom, the adjusted R-squared can give negative results. (Verbeek, 2012, p. 22)

$$\bar{R}^2 = 1 - \frac{\frac{1}{N-M}\sum_{n=1}^{N} e_n^2}{\frac{1}{N-1}\sum_{n=1}^{N}(y_n - \bar{y})^2} \tag{4.9}$$

A simplified method for measuring R-squared and adjusted R-squared is to use the error sum of squares, denoted SSE, and the total sum of squares, denoted SST (4.10). (Devore & Berk, 2012, p. 632-634)

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{y})^2}{\sum_{n=1}^{N}(y_n - \bar{y})^2} \tag{4.10}$$

$$\bar{R}^2 = 1 - \frac{(N-1)SSE}{(N-M-1)SST} = 1 - \frac{(N-1)\sum_{n=1}^{N}(y_n - \hat{y})^2}{(N-M-1)\sum_{n=1}^{N}(y_n - \bar{y})^2}$$

3.3 TIME SERIES

In this subchapter, time series analysis will be explained by decomposition, transformations, ARIMA processes and the Box-Jenkins method.

Time series analysis is an econometric method that dedicates itself to explain, model and forecast one or few economic variables that are generated by a process over time. Time series analysis uses quantitative data with annual, quarterly or monthly frequency. For financial values, the frequency can be even higher.

Time series' composition can often be distinguished between a deterministic, a stationary and a seasonal component. The seasonal component will not be included in this master thesis. The deterministic component of a time series often involves a trend, referred to as a deterministic trend, which can be explained by some constant and a mathematical function of the time variable t (5.1). The function can for example be linear, quadratic, polynomial or any additive

or multiplicative combination of functions. The main idea behind the trend component is that it is the long-run equilibrium as time goes to infinity. However, this is only true if the time series show deterministic tendencies. If the time series show stochastic tendencies, then it will in the long-run divert from the long run trend. (Heij, De Boer, Franses, Kloek, & Van Dijk, 2004, ch. 7)

$$T_t = c + \beta_1 f_1(t) + \beta_2 f_2(t) + \cdots + \beta_N f_N(t) \qquad (5.1)$$

Stationary processes, also called statistical processes, is the part of the time series that can only be described in terms of statistical properties which involves a probability distribution with a constant mean and a constant variance. A stationary component can often be identified by calculating autocorrelations which are short-run relations between successive values in the stationary component. A stationary process with all autocorrelations equal to zero is called white noise and has the same properties as the error term ε_t (also called disturbance term). These properties are zero mean, homoscedasticity and no autocorrelation. The error term is said to be independently and identically distributed with zero mean and σ^2 in variance. (George E. P. Box, Jenkins, Reinsel, & Ljung, 2015, p. 22-24)

It is often of interest or necessary to transform time series. Transformations can in many cases allow for a wider range of applications of models to the time series. A transformation is the process of applying a mathematical function to each value of the time series which often can help avoid difficulties in fitting a model to the observed values. These difficulties may include violations of statistical properties of the error term. The goal of the transformation is to avoid these violations by either linearizing or stationarizing the time series. (George E. P. Box et al., 2015, p. 96)

By distinguishing between the deterministic and stationary component, it is assumed that they are additive components. If the components are multiplicative then a logarithmic transformation is necessary (5.2). A logarithmic transformation is one of many power transformations that can help linearize the data. (Heij et al., 2004, ch. 7)

$$\log(Y_t) = \lim_{\lambda \to 0} \frac{Y_t^\lambda - 1}{\lambda} \qquad (5.2)$$

Differentiation can be used to make a time series stationary by removing trends, both stochastic and deterministic. Absolute growth is called the first difference and shows the exact difference between each observation (5.3). Relative growth is the percentage change of each observation from the respected previous observation (5.4). Logarithmic transformation and differentiation

can be used together to approximate the relative growth (5.5) (see Appendix: Proof 20). (Heij et al., 2004, ch. 7)

$$\Delta Y_t = Y_t - Y_{t-1} \tag{5.3}$$

$$\frac{Y_t - Y_{t-1}}{Y_{t-1}} \tag{5.4}$$

$$\Delta \log(Y_t) \approx \frac{\Delta Y_t}{Y_{t-1}} \tag{5.5}$$

For a time series with deterministic trend, the time series will converge to a trend line in the long-run and shocks will have transitory effects. In contrast, for a time series with stochastic trend, the time series will not converge to the trend line in the long-run and shocks will have permanent effects. Unit root tests are important to determine if a time series exhibit a deterministic or stochastic trend. In presence of a unit root, the time series exhibit a stochastic trend. If there is no unit root, then the time series exhibit the property of mean reverting behavior to an attractor which is the expected trend of the series. (Heij et al., 2004, ch. 7)

The Dickey-Fuller test unit root test developed by David Dickey and Wayne Fuller in 1979 (Dickey & Fuller, 1979). The Dickey-Fuller test considers an autoregressive process of order 1 and tests the null hypothesis that Φ is equal to one or the alternative hypothesis that Φ is less than one (5.6). The augmented Dickey-Fuller test is an extended test to consider autoregressive processes of order p (5.7). The null hypothesis in the augmented Dickey-Fuller test is that the sum of all Φ is equal to one and the alternative hypothesis is that the sum of all Φ is less than one. (Heij et al., 2004, ch. 7)

$$Y_t = \alpha + \Phi Y_{t-1} + \epsilon_t \tag{5.6}$$

$$Y_t = \alpha + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + \epsilon_t \tag{5.7}$$

A stationary process X_t with significant autocorrelation can be explained as an autoregressive process of order p denoted AR(p) (5.8) or as a moving average process of order q denoted MA(q) (5.9). Moving average model is the inverse of the autoregressive model and is called the invertible when being expressed as an autoregressive model of infinite order. An autoregressive moving average process is a combination of the two processes denoted ARMA(p, q). An autoregressive moving average process provides a more accurate approximation of higher order of autoregressive and moving average processes. (George E. P. Box et al., 2015, p. 52-53)

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \epsilon_t \qquad (5.8)$$

$$X_t = \epsilon_t + \Theta_1 \epsilon_{t-1} + \Theta_2 \epsilon_{t-2} + \cdots + \Theta_q \epsilon_{t-q} \qquad (5.9)$$

A non-stationary process may be stationary when differentiated d times. The process is then said to be integrated at d^{th} order. The process is then an autoregressive integrated moving average denoted ARIMA(p, d, q). (George E. P. Box et al., 2015, p. 90-91)

The Box-Jenkins method is an iterative approach to the construction of ARIMA models. It was developed by George Box and Gwilym Jenkins in 1970 (George E. P. Box et al., 2015). The approach involves three comprehensive steps: identification, estimation and diagnostics checking.

Identification methods aims to understand the data, how it was generated and to identify a model that should be further investigated. The first stage of identification is to determine stationarity of the time series. This is done by differencing the time series or extracting any deterministic trend from the time series. The autocorrelation function (ACF) and partial autocorrelation function (PACF) are analyzed to determine the behavior of the time series. (George E. P. Box et al., 2015, p. 177-182)

Stationary processes are assumed to have constant covariance between values Y_t and Y_{t-k} where k is called the degree of lag. If this holds for all values of t then there is autocovariance (5.10). Autocorrelation at lag k is given by its proportion of autocovariance at lag k relative to autocovariance at lag 0 (5.11). The partial autocorrelation function at lag k is defined as the correlation between the residuals from the linear regression assuming zero mean and the regression adjusted for intermediate variables (5.12). (George E. P. Box et al., 2015, p. 24-25)

$$\gamma_k = Cov[Y_t, Y_{t-k}] \qquad (5.10)$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov[Y_t, Y_{t-k}]}{Var[Y_t]} \qquad (5.11)$$

$$\Phi_{k,k} = Corr[Y_t - \hat{Y}_t, Y_{t-k} - \hat{Y}_{t-k}] \qquad (5.12)$$

The graphs of the autocorrelation and partial autocorrelation function with confidence intervals are helpful for determining the order of the autoregressive and/or moving average process. The confidence intervals can be calculated by Bartlett's formula (Bartlett, 1946).

Diagnostic checking involves checking for ways to improve the model. Residual diagnostics are helpful for checking the model's efficiency in explaining the data. The Ljung-Box test

(Ljung & Box, 1978) (5.13) is a modification of the Portmanteau lack-of-fit test and the simpler Box-Pierce test (G. E. P. Box & Pierce, 1970). The test measures the distribution of residual autocorrelations.

$$\tilde{Q} = n(n+2) \sum_{k=1}^{K} (n-k)^{-1} r_k^2(\hat{a})$$
(5.13)

Further testing for model adequacy can be performed with the Breusch-Godfrey test (Breusch, 1978; Godfrey, 1978), also called Lagrange multiplier (LM) test for serial correlation, the Durbin-Watson test for autocorrelation (Durbin & Watson, 1971), the autoregressive conditional heteroscedasticity test (ARCH) and White's test for heteroscedasticity (White, 1980). The ARCH and White's test considers the squared residuals as the dependent variable. The ARCH test regresses the squared residuals on lagged squared residuals and a constant while White's test regresses the squared residuals on the cross product of the original regressors and a constant. Jarque-Bera test is a goodness-of-fit test (Jarque & Bera, 1980). It tests if the skewness and kurtosis of the residuals resembles that of a normal distribution.

3.4 PANEL DATA

In this subchapter, panel data analysis will be explained, and different linear panel data regression models and diagnostics tests will be derived.

Panel data (also called longitudinal data) is characterized by large datasets where the number of units is much larger than the number of observations per unit. When the number of observations per unit corresponds to observations over time then the panel data exhibits properties of time series. To prepare panel data, both number of units and number of observations per unit is specified. It is then checked for missing values. If there are missing values, the panel data is called unbalanced. In some tests, it is required that the panel data is strongly balanced, meaning that the number of observations per unit is consistent and that there are no missing values. (Stock & Watson, 2012, p. 390)

Pooled regression models are ordinary least square regression models performed on panel data. This model for panel data assumes that all units have identical marginal effects of independent variables. This can only be true if there are no unobservable characteristics which is not true for most cases. In case of unexplained variations over units, individual heterogeneity, the recommended solution is to use robust and clustered standard errors. However, this solution gives better standard errors at the expense of reliability of the results. Other regression models

for panel data explains the individual heterogeneity across units by including unit-specific effects, denoted α_i (6.1). (Verbeek, 2012, p. 373)

$$y_{n,t} = \beta_1 + \beta_2 x_{2,n,t} + \cdots + \beta_M x_{M,n,t} + \alpha_n + u_{n,t} \tag{6.1}$$

Fixed effects regression model treats the unit-specific effects as intercepts that vary for each unit and can therefore be rewritten as the summed product of the unit-specific intercept times a dummy for each unit (6.2). This specific model is called the least squares dummy variable (LSDV) model. The fixed effects regression model assumes that variables are uncorrelated to the error term for all units and observations, which imply that the variables are strictly exogenous, independent of past, present and future values of the error term. The fixed effects regression model estimates parameters based on the differences within dimensions of the data, it does not explain differences across the observed units. Greene's test is a modified Wald test for heteroscedasticity in a fixed effects regression model and is a postestimation residual diagnostic test (Greene, 2012). (Stock & Watson, 2012; Verbeek, 2012, p. 377-378)

$$y_{n,t} = \beta_1 + \beta_2 x_{2,n,t} + \cdots + \beta_M x_{M,n,t} + \alpha_1 d_{1,n} + \alpha_2 d_{2,n} + \cdots + \alpha_N d_{N,n} + u_{n,t} \tag{6.2}$$

Random effects regression model treats the unit-specific effects as random factors that are independently and identically distributed over individuals (6.3). The error term is consisting of two components, the unit-specific residual and the remainder. The unit-specific residuals are assumed not to vary over time and the remainder is assumed to be uncorrelated over time. (Verbeek, 2012, p. 381-383)

$$y_{n,t} = \beta_1 + \beta_2 x_{2,n,t} + \cdots + \beta_M x_{M,n,t} + \epsilon_{n,t} \tag{6.3}$$

where $\epsilon_{n,t} = \alpha_n + u_{n,t}$

The Hausman test was developed by J. A. Hausman in 1978 (Hausman, 1978) and tests whether the fixed effects or random effects should be used by testing if they are significantly different. The Hausman test statistic has an asymptotic chi-squared distribution with the number of degrees of freedom equal to the number of elements in β (6.8). (Verbeek, 2012, p. 384-386)

$$\xi_H = \left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right)' \left(\hat{\sigma}^2_{\hat{\beta}_{FE}} - \hat{\sigma}^2_{\hat{\beta}_{RE}}\right)^{-1} \left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right) \tag{6.8}$$

A good test to decide whether to use random effects regression or a pooled regression is the Breusch-Pagan Lagrange multiplier (LM) test (6.9) (Breusch & Pagan, 1980). It is a test for individual heterogeneity with null hypothesis of zero variance across units.

$$LM = \sqrt{\frac{NT}{2(T-1)}} \left(\frac{\sum_{n=1}^{N}\left(\sum_{t=1}^{T} e_{n,t}\right)^2}{\sum_{n=1}^{N}\sum_{t=1}^{T} e_{n,t}^2} - 1 \right) \qquad (6.9)$$

Wooldridge's test is a test for serial correlation of non-systematic errors of a linear panel data model (Drukker, 2003; Wooldridge, 2010). The test involves regressing the first differenced variables of the model and performing a Wald's test of the null hypothesis that the coefficient of lagged residuals, correlation between sequential differenced error terms, is equal to -0,5. A rejected null hypothesis implies the presence of autocorrelation.

In cases of structure within the error term, there are problems with both heteroscedasticity and autocorrelation. The assumptions of Gauss-Markov (4.5) no longer hold and the OLS estimator is therefore no longer the best estimator. In these cases, a more efficient estimator is the generalized least squares (GLS) estimator. Generalized least squares assumes a different error covariance matrix (6.4). The Ψ is a positive definite matrix and when it is not equal to the identity matrix then there are non-spherical error terms. By taking the variance of the OLS estimator, it is shown that it is unbiased but not efficient (6.5) (see Appendix: Proof 22). (Verbeek, 2012, p. 381-383)

$$Var[\epsilon|X] = \sigma^2 \Psi \qquad (6.4)$$

$$Var[\hat{\beta}|X] = \sigma^2 (X'X)^{-1} X' \Psi X (X'X)^{-1} \qquad (6.5)$$

Generalized least squares aims to transform the model such that it retains β as a linear parameter vector and creates a new error term which meets the Gauss-Markov assumptions of homoscedasticity and no autocorrelation. In the derivation of the generalized least squares estimator, the Ψ is assumed to be known. It can be shown that this assumption is sufficient to transform the regression (see Appendix: Proof 23). Then by applying the OLS method on the transformed regression model, the best linear unbiased estimator is then estimated by the generalized least squares estimator (6.6). (Verbeek, 2012, p. 96-97)

$$\hat{\beta}_{GLS} = (X' \Psi^{-1} X)^{-1} X' \Psi^{-1} y \qquad (6.6)$$

In most cases, Ψ is not known and therefore must be estimated first. This can be done by feasible generalized least squares (FGLS) introduced by D. Cochrane and G. H. Orcutt in 1949 (Cochrane & Orcutt, 1949). (Stock & Watson, 2012, p. 648; Verbeek, 2012, p. 97)

Another estimator that can be used when there is presence of heteroscedasticity of the OLS estimator is the weighted least squares (WLS) estimator. The derivation of the weighted least

squares estimator is like the derivation of the GLS estimator, but in the WLS the error covariance matrix is explained by the form of heteroscedasticity (6.7). (Stock & Watson, 2012, p. 725-726; Verbeek, 2012, p. 99)

$$\Psi = Diag[h_n^2] \tag{6.7}$$

4 RESEARCH APPROACH

The program used to conduct the research is the statistical software package Stata version 15.1 and the data is from The Penn World Table version 9.0 (PWT9.0). PWT9.0 is a database with information on national accounts for 182 countries from 1950 to 2014. The database was developed and released by the Groningen Growth and Development Centre of the university of Groningen in 2015 (Feenstra, Inklaar, & Timmer, 2015). The database exhibits properties of both time series and panel data. Each country is specified as a unit and observations per country is sorted by a yearly frequency. Because of annual frequency, there will not be a seasonal component to analyze and all differentiation will be yearly only. Because of observations only until 2014, there are no reason to forecast. However, this is the most recent and comprehensive database that is available today.

4.1 VARIABLES

The variables that are included in the Stata work file are:

Label	Name
Country name	country
Year	year
Population (in millions)	pop
Number of persons engaged (in millions)	emp
Human capital index, see note hc	hc
Real GDP at constant 2011 national prices (in mil. 2011US$)	rgdpna
Real consumption at constant 2011 national prices (in mil. 2011US$)	rconna
Capital stock at constant 2011 national prices (in mil. 2011US$)	rkna
Average depreciation rate of the capital stock	delta

Real and constant 2011 national prices are good for comparison between countries. Nominal and current national prices show bigger differences in values due to effects of inflation of prices. In real and constant prices, the effects of inflation have been excluded. Prices in purchasing power parity is also effective, but highly fluctuating in a day to day basis and therefore is not as accurate for yearly observations. The human capital is measured by average years of education. From the variables from the database it is of interest to create these new variables:

Label	Name
Real GDP per capita (in 2011US$)	rgdppc
Real GDP per worker (in 2011US$)	y_t
Real capital stock per worker (in 2011US$)	k_t
Consumption per worker (in 2011US$)	c
Savings rate (%)	s
Population growth (%)	n
Technological progress (%)	g
OECD country (dummy)	OECD

Real GDP per capita and Real GDP per worker have their own interesting interpretations. As the real GDP per capita is a measure of the average welfare in a country, the real GDP per worker is a measure of the average income levels in a country. Both are interesting, but from the theory of the Solow model it is more correct to use real GDP per worker as an estimator for output per effective worker.

The variable real GDP per capita is derived by the real GDP divided by population. The variable real GDP per worker is derived by the real GDP divided by people employed. Savings rate is derived from real GDP per worker minus consumption per worker which is real consumption divided by people employed. Population growth is derived from the growth rate of population. Technological progress is derived from the growth rate of the employment rate which is derived from the people employed divided by the population.

Some of the generated variables have their respective logarithmic transformations. This is to be able to use linear regression models as the output is explained by a multiplicative relationship of inputs. The logarithmic transformation of real GDP per capita and real GDP per worker, also allows for derivation of annual growth rates.

The OECD dummy is created to be able to compare the full sample to OECD countries exclusively. This is because the result can be dependent on certain unobserved characteristics of the countries, and OECD countries are assumed to be similar in terms of many of these characteristics. This gives a more reliable result, but also less relevant to answer the question of interest. OECD stands for The Organization for Economic Co-operation and Development and there are 35 countries that are members today.

4.2 SAMPLE SELECTION

The dataset includes 182 countries out of the 195 countries recognized by the United Nations today. To prepare the panel data, unit index is specified as country and observation for each unit, the time index, is specified by the year. Since country is a string variable, it must first be encoded to a numerical variable.

For the 182 countries in the dataset, not all countries have observed values for the variables of interest. Also, some countries do not have observed values for all the years that are needed. The problem with missing values in the dataset can be solved by creating balanced panels by sampling out countries and years without missing values for an interval of years.

The method is to maximize the number of observations by the number of years and countries included. Initial requirements are that the latest year included is always 2014, the minimum of observed values for each country are always 30, meaning from 1985. The last requirement is that for all panel data the number of countries included must exceed the number of years included. The goal of the sample selection is to maximize observations of the necessary variable given that the panel is balanced and that none of the requirements are broken.

The process of the sample selection involves counting observed values for each country up until 2014 and to then create a histogram which shows the frequency of countries by number of observations per country. It is then possible to choose countries with sufficient number of observations by how many years that are to be included.
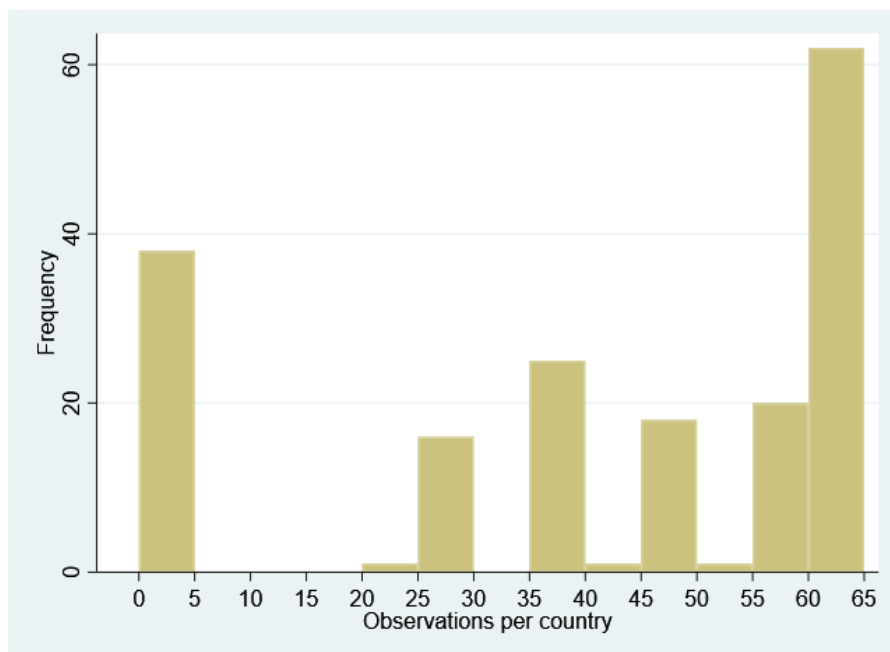
<div align="right">Graph 1</div>

The histogram shows the quantity of countries by observations per country (Graph 1), where each observation is a year of no missing values of all the original variables included in the work file. 38 countries are excluded from the sample due to 0 observations because they are not observed for one or more variables or/and missing value(s) for the year 2014. There are 48 countries with no missing values for the full range of 65 observations. By setting a requirement of 35 observations per country, 4445 observations are included. While for a requirement of 45 observations per country, 4545 observations are included. It is of interest to maximize the number of observations and therefore the requirement of 45 observations per country is exercised and 101 countries, including 29 OECD countries, are included in the sample.

A second sample selection is constructed from the first sample and the reason will be explained later. This sample include 53 countries with 44 observations per country and 2332 observations in total.

5 TESTS AND ANALYSIS

As mentioned earlier, real GDP per capita and real GDP per worker have their own interesting interpretations. In the theory of the Solow model it is more correct to use real GDP per worker, but in many previous cases the real GDP per capita has been used. This is because the available data on population exceed the data on employment. The choice of whether to use per capita or per worker affects the results and it is therefore of interest to look at some of the empirical

differences of the two. The graph shows time series of average real GDP per capita and average real GDP per worker (Graph 2). The real GDP per worker has more variability.
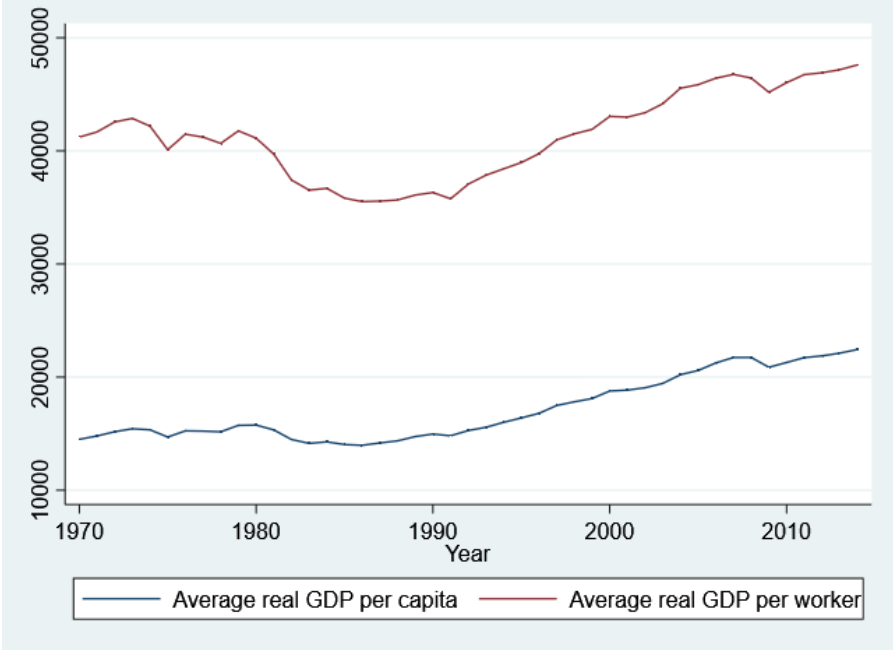
Graph 2

The scatterplot shows average population growth and average technological progress (Graph 3). The average population growth seems to follow a downward somewhat cyclical trend. Technological progress is more fluctuating and does not follow a clear trend.
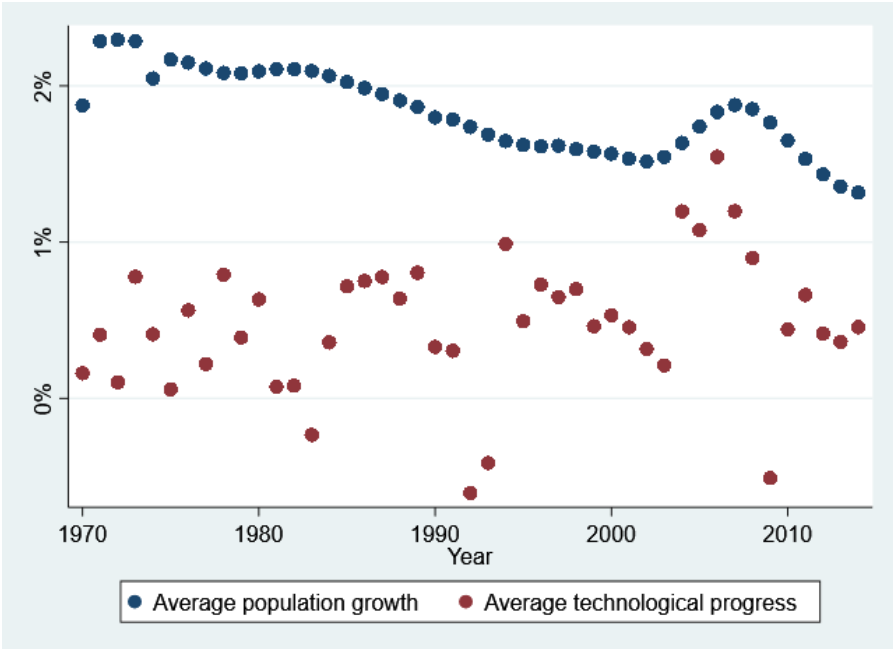
Graph 3

The growth-initial level regression is a test for β convergence which regresses the annual economic growth explained by initial levels of economic output (7.1). If the test shows significant negative coefficient, then there is indication of β convergence and the coefficient

would imply that a percentage decrease in initial levels of economic output is estimated to cause a percentage increase in annual economic growth.

$$\frac{\ln\left(\frac{y_{n,T}}{y_{n,0}}\right)}{T-1} = \alpha + \beta \ln y_{n,0} + \epsilon \tag{7.1}$$

The growth-initial level regression (Graph 4) shows evidence of β convergence because the coefficient of the linear regression is negative, equal to –0,0056 (see Appendix: Regression output 1). The result says that poorer economies grow faster than richer economies and the result is highly significant, but the R-squared is at 20% which is low. Residual diagnostics show non-normality and heteroscedasticity. The standardized normal probability plot shows symmetric heavy tails (Graph 5). Plotting residuals against fitted values show an irregular variance of the residuals (Graph 6). Breusch-Pagan test for heteroscedasticity rejects the null hypothesis of homoscedasticity at a 1,07% significance level while White's test rejects the null hypothesis at a 5,62% significance level. The problems with the residuals indicate an unreliable test result and a lot of unexplained variation of observations. This motivates the use of robust standard errors in the regression which relaxes the assumption of heteroscedasticity. Performing the regression with the option for robust standard errors the 95% confidence interval of coefficients are wider.
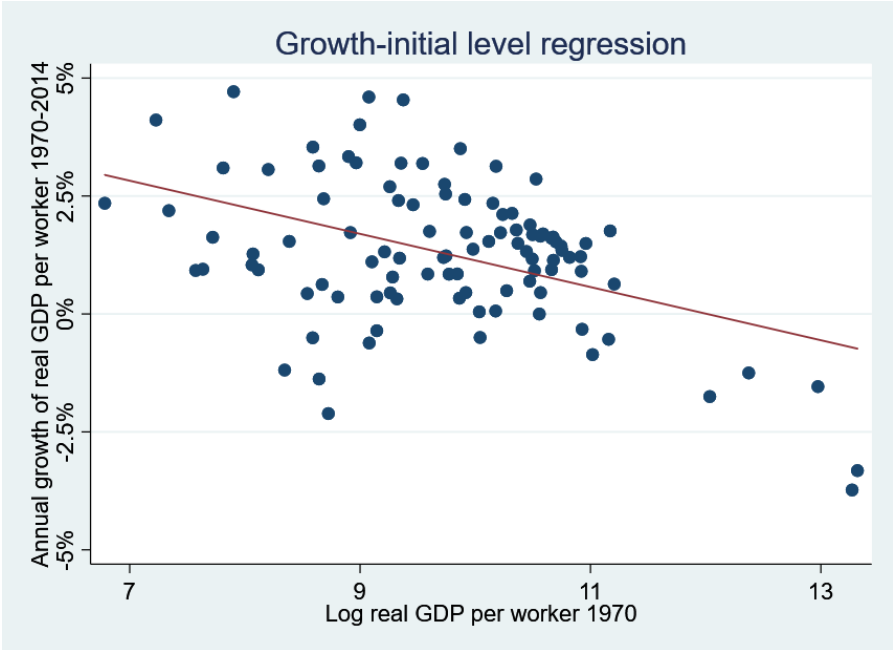
Graph 4
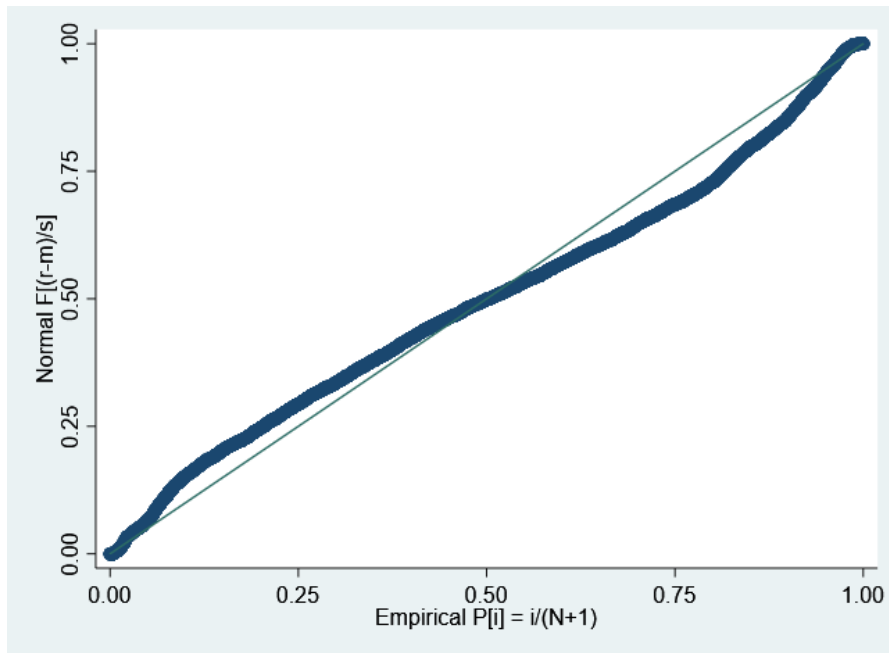
Graph 5
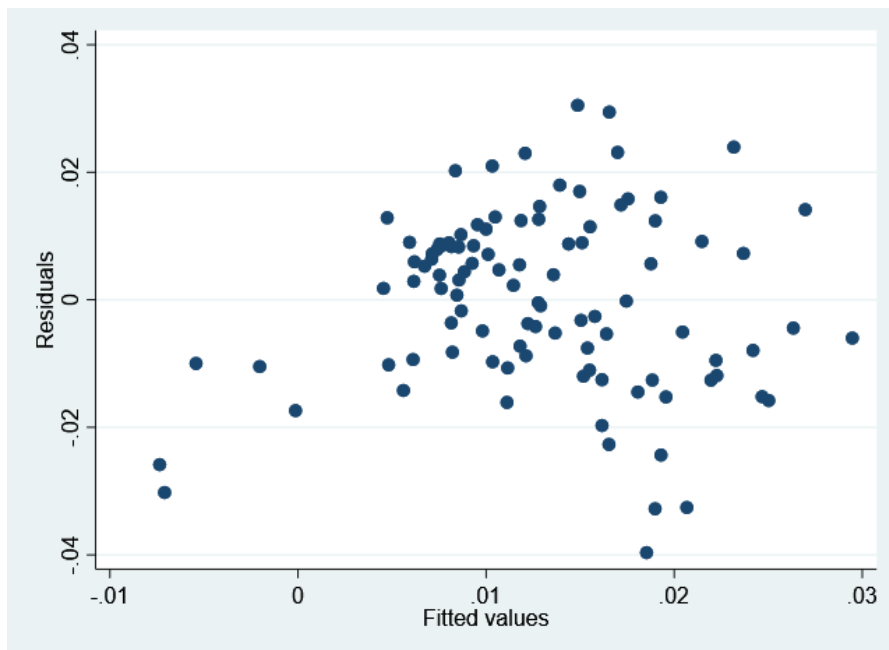
Graph 6

Performing the growth-initial level regression test exclusively for OECD countries (Graph 7) shows evidence of β convergence, a highly significant β-coefficient of –0,0103 (see Appendix: Regression output 2). In this test the R-squared is 38,78% which is higher than for the full sample. The residual diagnostics show non-normality and homoscedasticity. The standardized normal probability plot shows heavy tails (Graph 8). Plotting residuals against fitted values show a somewhat constant variance of the residuals (Graph 9). White's test for heteroscedasticity fails to reject the null hypothesis of homoscedasticity at an 80,99% significance level while Breusch-Pagan test rejects the null hypothesis of constant variance of

residuals at a 2,52% significance level. Since the purpose of looking at OECD countries exclusively is to look at countries similar in unobserved characteristics, it is of interest to look at countries that may seem different in behavior by a leverage versus squared residuals plot is interesting (Graph 10). Two countries with high leverages to low squared residuals are Poland and Hungary, Also, Switzerland and Turkey show higher leverages to low squared residuals. The growth-initial level regression test performs better when done for OECD countries exclusively. From an analytical perspective, the test performs better for countries similar in some unobserved characteristics.
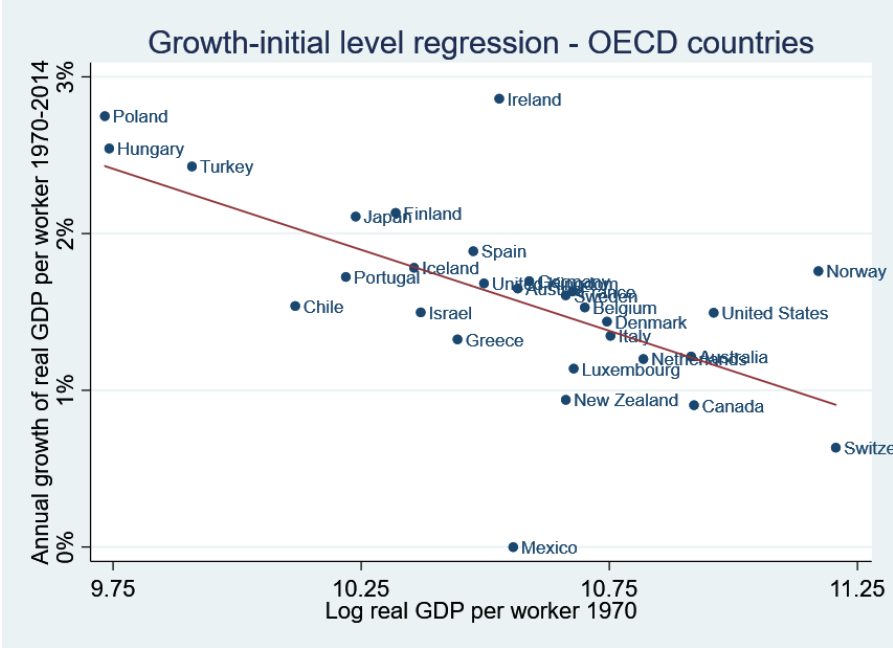
Graph 7
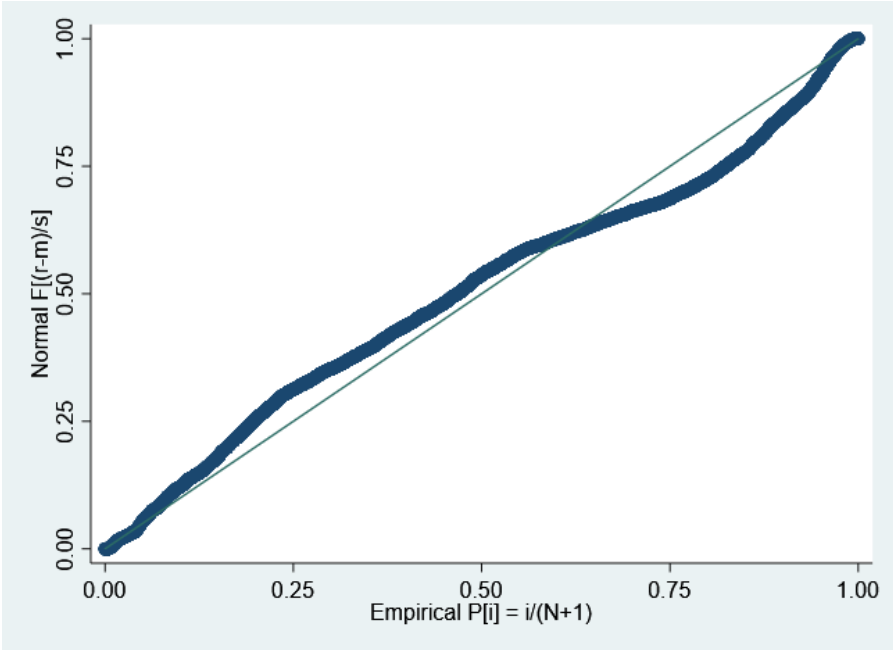
Graph 8
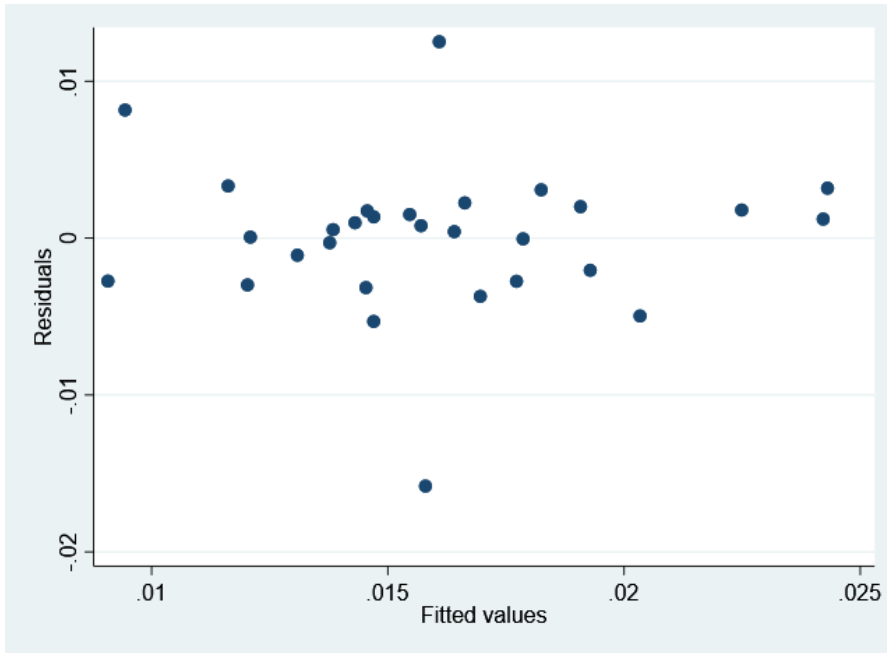
Graph 9
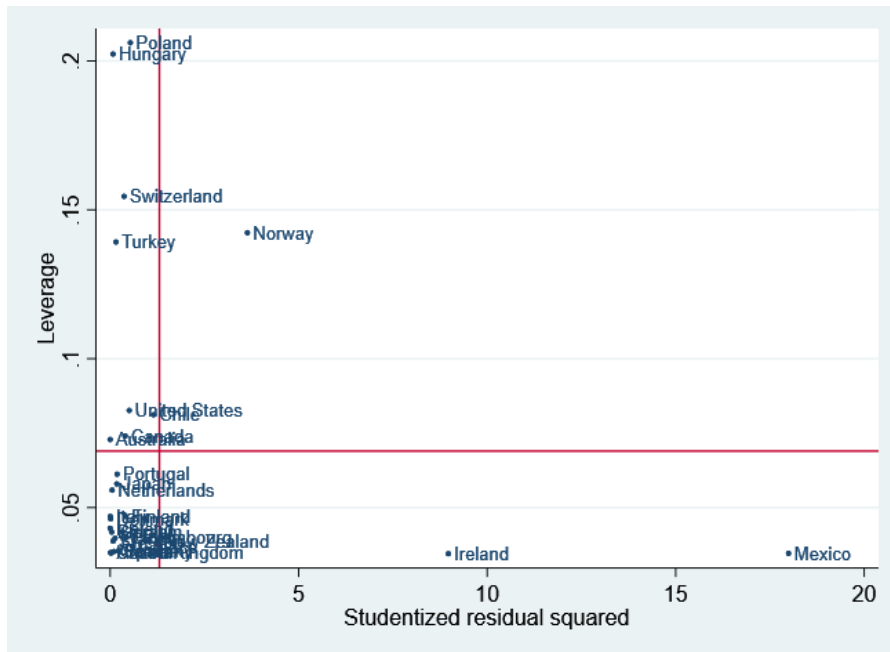
Graph 10

There is σ convergence if the measured standard deviation of real GDP per worker decreases over time. The test for σ convergence can be written as that subsequent values of standard deviation are lower (7.2). By examining the behavior of the standard deviation time series, the presence of σ convergence is inferred.

$$\hat{\sigma}_{y_t} > \hat{\sigma}_{y_{t+1}} \tag{7.2}$$

The graph shows time series of the standard deviation of real GDP per worker, for the full sample and the OECD countries exclusively (Graph 11). For the OECD countries the inequality

is much lower than for the full sample. The standard deviation of real GDP per worker for OECD countries shows that there is no σ convergence, but rather σ divergence. The inequality among OECD countries is increasing. For the full sample, there is a lot more variation. Inequality seems to be decreasing drastically between 1970 and 1988, increasing until 2004 and decreasing again until 2009. This could however, be showing convergence to a steady level of inequality, meaning that in the long-run there will always be a deterministic amount of inequality between countries.
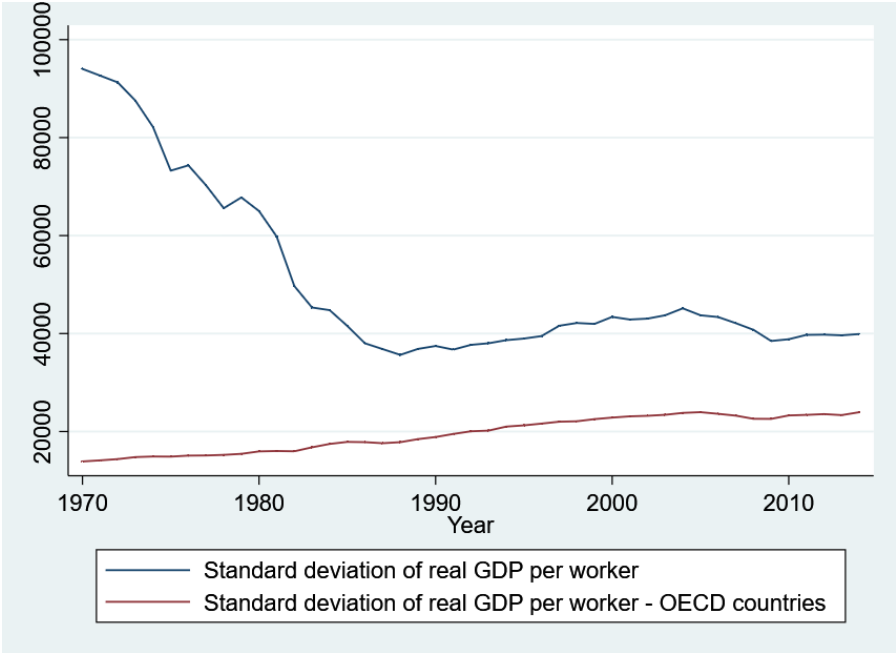
Graph 11

Absolute convergence is when all economies follows a similar path, while for conditional convergence there must be included some other characteristic for this to be true and therefore each countries path is unique if not conditional on this characteristic. So far, it has been shown that the growth-initial level regression is only a good test for countries with similar characteristics and countries with similar characteristics most likely diverge in the sense of dispersion. This would all imply that there are greater economic characteristics that must be included to determine the trend of economic growth.

The data is first fitted to the Cobb-Douglas production function as a pooled linear regression (7.3).

$$\ln(Y_{n,t}) = \beta_0 + \beta_1 \ln(K_{n,t}) + \beta_2 \ln(A_{n,t} L_{n,t}) + \epsilon_{n,t} \qquad (7.3)$$

The pooled linear regression is highly significant, and the result suggest a capital's share of 80,2% (see Appendix: Regression output 3). The graph shows a scatterplot of real GDP per

worker for each country (Graph 12). The red dots show mean values for each country and the connected line shows the across country variance. The graph implies that there is individual heterogeneity, which is a strong appeal to use a fixed effects linear regression model.
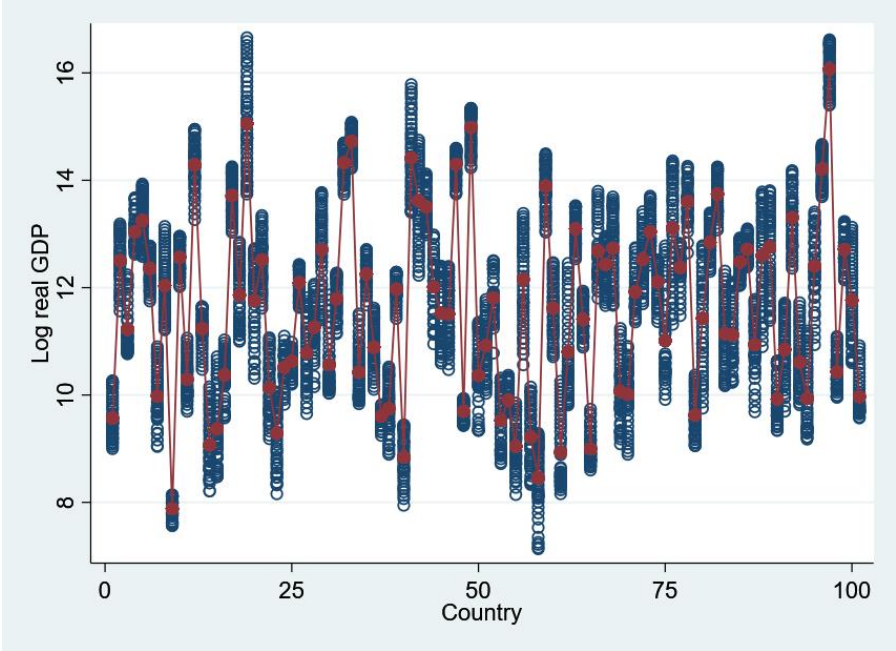
<div align="right">Graph 12</div>

Across countries variance are included by including unit-specific intercepts and a dummy for each country which gives the fixed effects regression model (7.4).

$$\ln(Y_{n,t}) = \alpha_n D_n + \beta_1 \ln(K_{n,t}) + \beta_2 \ln(A_{n,t} L_{n,t}) + \epsilon_{n,t} \qquad (7.4)$$

The result of the fixed effects linear regression shows a highly significant β_1-coefficient of 0,6232 and β_2-coefficient of 0,3542 (see Appendix: Regression output 4). There seems to be correlation between unit-specific intercepts and independent variables of 0,2276. This implies that the use of a random effects regression model is not reasonable in this case. Greene's test for heteroscedasticity show strong presence of heteroscedasticity which implies that the estimator is not efficient and that robust standard errors or GLS should be considered. Wooldridge's test for autocorrelation rejects the null hypothesis of no first order autocorrelation in the panel data. By running a GLS regression with heteroscedasticity and panel specific autocorrelation structure, the model is highly significant with highly significant β_1-coefficient of 0,7338 and β_2-coefficient of 0,2728 (see Appendix: Regression output 5).

By testing that coefficient of the logarithm of capital and the coefficient of the logarithm of effective labor is equal to 1, the assumption of constant returns to scale is empirically tested.

The null hypothesis is that there are constant returns to scale and the test statistic fail to reject the null hypothesis at a 23,6% significance level.

From the Solow model, the long run trend is explained by the steady state. If the real GDP per effective worker converges to the steady state, then the trend must be deterministic and explained by the steady state (7.5) (see Appendix: Proof 24).

$$\ln(y_{n,t}) = (1 - e^{-\lambda t}) \frac{\alpha}{1 - \alpha} \ln\left(\frac{s_{n,t}}{n_{n,t} + g_{n,t} + \delta_{n,t}}\right) + e^{-\lambda t} \ln(y_{n,t-1}) + \epsilon_{n,t} \tag{7.5}$$

The model explains that in the long-run, when t goes to infinity, the component $e^{-\lambda t}$ will be equal to zero and ln $y_{n,t}$ will be explained by the steady state alone. When observing a single country, the speed of convergence λ is measure of the economy's distance from its own steady state. When observing multiple countries, the speed of convergence λ is a measure of the speed of which countries are closing the gap of differences between rich and poor countries.

A large problem in the neoclassical growth theory is that the models fail to consider negative values of savings rate, population growth and technological progress. Negative savings rates occur when the annual average of private consumption exceeds the annual average of private income. Negative population growth and negative technological growth are not uncommon and depreciation rate is always positive. These problems occur because of logarithmic transformations which generate missing values in the sample when the sum of population growth, technological progress and depreciation rate are negative values. This creates the need for another sample data selection within the sample, where these negative rates do not occur. Therefore, the previously mentioned second data sample selection will be used.

Performing the pooled regression model for the steady state, the result show highly significant coefficients of 0,0067 and 0,9944 which implies a capital's share of 54,5% (see Appendix: Regression output 6). Individual heterogeneity suggests that the fixed effects regression model is appropriate (Graph 13).
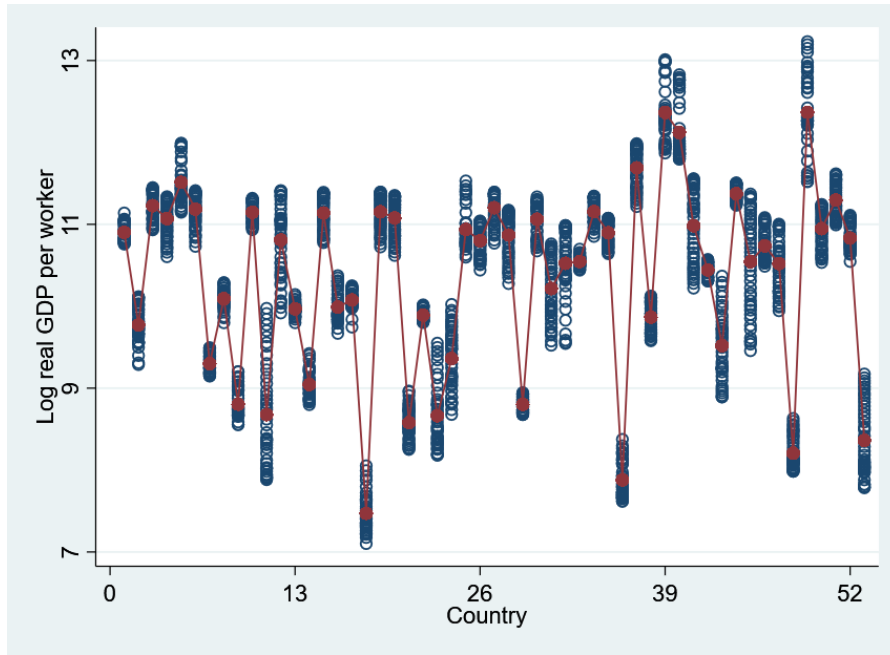
Graph 13

The fixed effects regression model for the steady state shows highly significant coefficients 0,0186 and 0,976 which implies a capital's share of 43,7% (see Appendix: Regression output 7). The correlation between unit-specific intercepts and independent variables is equal to 0,7063. Residual diagnostics show non-normality and heteroscedasticity. Greene's test for heteroscedasticity rejects the null hypothesis of homoscedasticity which implies that the estimator is not efficient and that robust standard errors or GLS should be considered. By running a GLS regression for the steady state with heteroscedasticity and panel specific autocorrelation structure, the result show highly significant coefficients of 0,0109 and 0,9927 which implies a capital's share of 60% (see Appendix: Regression output 8).

It is stated in the neoclassical growth theory that a reasonable capital's share is equal to 1/3 which means that the results that have been presented so far is unsatisfactory, even despite the non-normality, heteroscedasticity and autocorrelation. Therefore, there is strong appeal to add human capital and to use the augmented Solow model. When human capital is added, the long-run trend can be derived from the augmented Solow model (7.6) (see Appendix: Proof 25). Since there is no reasonable way to derive savings rate of human capital from the available data, human capital is used a measure of the steady state of human capital.

$$\ln y_{n,t} = \left(1 - e^{-\lambda t}\right) \frac{\alpha}{1 - \alpha - \beta} \ln\left(\frac{s_{k_{n,t}}}{n_{n,t} + g_{n,t} + \delta_{n,t}}\right) + e^{-\lambda t} \ln y_{n,t-1}$$

$$+ \left(1 - e^{-\lambda t}\right) \frac{\beta}{1 - \alpha - \beta} \ln\left(\frac{s_{h_{n,t}}}{n_{n,t} + g_{n,t} + \delta_{n,t}}\right) + \epsilon_t$$

(7.6)

The result of a fixed effects regression model shows significant coefficients of 0,0181, 0,9722 and 0,0196 (the coefficient of the logarithmic transformation of average years of education is significant at a 0,3% level) which implies a α of 27,6% and a β of 29,9% (see Appendix: Regression output 9). Greene's test for heteroscedasticity rejects the null hypothesis of homoscedasticity which implies that the estimator is not efficient and that robust standard errors or GLS should be considered. By running a GLS regression for the steady state with human capital and with heteroscedasticity and panel specific autocorrelation structure, the result show highly significant coefficients of 0,0096, 0,9875 and 0,0264 which implies a capital's share of 19,9% and a human capital's share of 54,4%. (see Appendix: Regression output 10).

The equilibrium level of growth of output in the R&D model depends solely on population growth. The Hausman test and the Breusch Pagan test show preference for the random effects regression model (see Appendix: Regression output 11). Greene's test for heteroscedasticity rejects the null hypothesis and robust standard errors are included in the model. The results show highly significance, but a low R-squared.

6 CONCLUSION

To conclude, the research has tested for the presence of convergence. The presence of β convergence was tested by a growth-initial level regression. First for the full sample of 101 countries and then exclusively for OECD countries. The test result showed evidence of β convergence which implies that poorer countries tend to grow faster than richer countries. In contradiction, the model was diagnosed with non-normality and heteroscedasticity showing signs of a non-reliable test result that is generalizing and affected by extreme values. The test performs better for the OECD where the intention is to compare countries that are similar in unobserved characteristics.

The presence of σ convergence was tested by time series of the standard deviation of real GDP per worker for the full sample of 101 countries and exclusively for OECD countries. The result showed a steady increase in standard deviation for OECD countries, implying that inequalities between richer and poorer countries within the OECD are increasing. This means that countries within the OECD are diverging in the sense of income dispersion. For the full sample of 101 countries, the result showed a significant decrease in standard deviation between 1970 and 1988 with mixed interpretations for years until 2014. It is difficult to make a conclusion about income

dispersion and inequality for the 101 countries in recent years from the time series of standard deviation of real GDP per worker for the full sample.

Absolute and conditional convergence was tested through the theory of the Solow model. The results show similar empirical weaknesses of the Solow model as previous research. However, by including a measure for human capital by the average of years of education, the results show a more satisfactory capital's share. Because of difficulties of heteroscedasticity and autocorrelation, it is appropriate to use a generalized least squares method to estimate the best linear unbiased estimator. The strong presence of individual heterogeneity between countries implies that countries converge conditionally rather than absolute.

The resulting evidence from the conducted tests and analysis has thus successfully provided satisfactory answers to the research questions of this master thesis.

Results of the research in this thesis revisit some conclusions that motivated the start of new growth theory. The R&D model was tested, but not given a thorough analysis. From the random effects regression model of growth rate of GDP and population growth, the model did not seem to explain more than the Solow model.

There are tools of time series analysis beyond those exploited in this thesis. Time series analysis is important for understanding underlying processes and it would be of interest to do convergence analysis of one or few economies.

Convergence has proven to be an interesting topic to study by applying econometric methods. For further research it would be of interest to include other models and variables to explain economic growth.

7 Appendix

7.1 Proofs

Proof 1: Intensive form transformation

Left hand side: $\frac{1}{AL}Y = \frac{Y}{AL} = y$

Right hand side: $\frac{1}{AL}F(K,AL) = F\left(\frac{1}{AL}K, \frac{1}{AL}AL\right) = F\left(\frac{K}{AL}, \frac{AL}{AL}\right) = F\left(\frac{K}{AL}, 1\right) = f(k)$

Proof 2: Cobb-Douglas assumptions

Constant returns to scale:

$$F(cK, cAL) = (cK)^\alpha (cAL)^{1-\alpha} = c^\alpha c^{1-\alpha} K^\alpha (AL)^{1-\alpha} = cF(K,AL)$$

Intensive form:

$$f(k) = \left(\frac{K}{AL}\right)^\alpha \left(\frac{AL}{AL}\right)^{1-\alpha} = k^\alpha 1^{1-\alpha} = k^\alpha$$

Diminishing returns to capital:

$$f'(k) = \alpha k^{\alpha-1} > 0$$

$$f''(k) = \alpha(\alpha-1)k^{\alpha-2} < 0$$

Inada conditions:

$$\lim_{k\to 0} f'(k) = \lim_{k\to 0} \alpha k^{\alpha-1} = \infty$$

$$\lim_{k\to 0} f'(k) = \lim_{k\to\infty} \alpha k^{\alpha-1} = 0$$

Proof 3: Solving growth rates as differential equations

$$\frac{dL(t)}{dt} = nL(t) \qquad\qquad \frac{dA(t)}{dt} = gA(t)$$

$$\int \frac{1}{L(t)} dL(t) = \int n \, dt \qquad\qquad \int \frac{1}{A(t)} dA(t) = \int g \, dt$$

$$\log(L(t)) = nt + c \qquad\qquad \log(A(t)) = gt + c$$

$$L(t) = e^{(nt+c)} \qquad\qquad A(t) = e^{(gt+c)}$$

$$L(0) = e^{n*0+c} = e^c \qquad\qquad A(0) = e^{g*0+c} = e^c$$

$$=> L(t) = L(0)e^{nt} \qquad\qquad => A(t) = A(0)e^{gt}$$

Proof 4: Law of motion for capital

$$K_t = K_{t-1} + I_{t-1} - \delta K_{t-1}$$

$$net\ investment\ \Delta K_t = gross\ investment\ I_{t-1} - depreciation\ \delta K_{t-1}$$

Proof 5: The dynamics of capital per effective worker

$$\dot{k} = \left(\frac{\dot{K}}{AL}\right) = \frac{\dot{K}AL - K(\dot{AL})}{(AL)^2} = \frac{\dot{K}}{AL} - \frac{K}{(AL)^2}(\dot{A}L + A\dot{L}) = \frac{sY - \delta K}{AL} - \frac{K}{AL}\left(\frac{\dot{A}}{A} + \frac{\dot{L}}{L}\right)$$

$$= sy - \delta k - k(g + n) = sy - (n + g + \delta)k$$

Proof 6: Steady state level of capital per effective worker

$$sy^* = (n + g + \delta)k^*$$

$$sk^{*\alpha} = (n + g + \delta)k^*$$

$$\frac{k^*}{k^{*\alpha}} = \frac{s}{n + g + \delta}$$

$$k^{*1-\alpha} = \frac{s}{n + g + \delta}$$

$$k^* = \left(\frac{s}{n + g + \delta}\right)^{\frac{1}{1-\alpha}}$$

$$y^* = k^{*\alpha} = \left(\frac{s}{n + g + \delta}\right)^{\frac{\alpha}{1-\alpha}}$$

Proof 7: Derivation of elasticity of output to savings rate

$$E_{y^*/s} = \frac{\partial y^*}{\partial s} * \frac{s}{y^*} = \frac{\partial y^*}{\partial k^*} * \frac{\partial k^*}{\partial s} * \frac{s}{k^{*\alpha}} = \alpha k^{*\alpha-1} * \frac{1}{1-\alpha}\left(\frac{s}{n+g+\delta}\right)^{\frac{1}{1-\alpha}-1} \frac{1}{n+g+\delta} * \frac{s}{k^{*\alpha}}$$

$$= \alpha k^{*\alpha-1} \frac{1}{1-\alpha} k^* \frac{s}{n+g+\delta}^{-1} \frac{s}{n+g+\delta} k^{*-\alpha}$$

$$= \frac{\alpha}{1-\alpha} k^{*\alpha-1} k^{*1-\alpha} \frac{s}{n+g+\delta}^{1-1} = \frac{\alpha}{1-\alpha}$$

Proof 8: Speed of convergence

$$\dot{k} = \frac{\partial \dot{k}(k)}{\partial k}(k-k^*)$$

$$\lambda = -\frac{\partial \dot{k}(k)}{\partial k}$$

$$\dot{k}(t) = -\lambda(k(t) - k^*)$$

$$\frac{\partial k(t)}{\partial t} = -\lambda(k(t) - k^*)$$

$$\int \frac{1}{k(t) - k^*} \partial k(t) = \int -\lambda \, \partial t$$

$$\ln(k(t) - k^*) = -\lambda t + c$$

$$k(t) - k^* = e^{-\lambda t + c}$$

$$k(0) - k^* = e^{-\lambda*0+c} = e^c$$

$$k(t) = k^* + e^{-\lambda t}(k(0) - k^*)$$

$$\frac{\partial \dot{k}(k)}{\partial k} = sf'(k^*) - (n+g+\delta) = \frac{(n+g+\delta)k^*}{f(k^*)}f'(k^*) - (n+g+\delta)$$

$$= (n+g+\delta)(k^{1-\alpha}\alpha k^{\alpha-1} - 1) = (n+g+\delta)(\alpha-1)$$

$$\lambda = (1-\alpha)(n+g+\delta)$$

Proof 9: Constant returns to scale

$$cY(t) = (cK(t))^\alpha (cH(t))^\beta \big(cA(t)L(t)\big)^{1-\alpha-\beta} = c^\alpha c^\beta c^{1-\alpha-\beta} K(t)^\alpha H(t)^\beta \big(A(t)L(t)\big)^{1-\alpha-\beta}$$
$$= cK(t)^\alpha H(t)^\beta \big(A(t)L(t)\big)^{1-\alpha-\beta}$$

Proof 10: Intensive form transformation

Left hand side: $\frac{1}{A(t)L(t)} Y(t) = y(t)$

Right hand side: $(\frac{1}{A(t)L(T)} K(t))^\alpha (\frac{1}{A(t)L(T)} H(t))^\beta \left(\frac{1}{A(t)L(T)} A(t)L(t)\right)^{1-\alpha-\beta} =$

$k^\alpha h^\beta 1^{1-\alpha-\beta} = k^\alpha h^\beta$

Proof 11: Dynamics of physical and human capital

$$\dot{k} = \left(\frac{\dot{K}}{AL}\right) = \frac{\dot{K}AL - K(\dot{AL})}{(AL)^2} = \frac{\dot{K}}{AL} - \frac{K}{(AL)^2}\big(\dot{A}L + A\dot{L}\big) = \frac{s_k Y - \delta K}{AL} - \frac{K}{AL}\left(\frac{\dot{A}}{A} + \frac{\dot{L}}{L}\right)$$
$$= s_k y - \delta k - k(g+n) = s_k y - (n+g+\delta)k$$

$$\dot{h} = \left(\frac{\dot{H}}{AL}\right) = \frac{\dot{H}AL - H(\dot{AL})}{(AL)^2} = \frac{\dot{H}}{AL} - \frac{H}{(AL)^2}\big(\dot{A}L + A\dot{L}\big) = \frac{s_h Y - \delta H}{AL} - \frac{H}{AL}\left(\frac{\dot{A}}{A} + \frac{\dot{L}}{L}\right)$$
$$= s_h y - \delta h - h(g+n) = s_h y - (n+g+\delta)h$$

Proof 12: Steady state levels of physical and human capital per effective worker

$$s_k y^* = (n+g+\delta)k^* \qquad\qquad s_h y^* = (n+g+\delta)h^*$$

$$s_k k^{*\alpha} h^{*\beta} = (n+g+\delta)k^* \qquad\qquad s_h k^{*\alpha} h^{*\beta} = (n+g+\delta)h^*$$

$$k^{*1-\alpha} = \frac{s_k}{n+g+\delta} h^{*\beta} \qquad\qquad h^{*1-\beta} = \frac{s_h}{n+g+\delta} k^{*\alpha}$$

$$k^* = \left(\frac{s_k}{n+g+\delta} h^{*\beta}\right)^{\frac{1}{1-\alpha}} \qquad\qquad h^* = \left(\frac{s_h}{n+g+\delta} k^{*\alpha}\right)^{\frac{1}{1-\beta}}$$

$$k^* = \left(\frac{s_k}{n+g+\delta} \left(\frac{s_h}{n+g+\delta} k^{*\alpha} \right)^{\frac{\beta}{1-\beta}} \right)^{\frac{1}{1-\alpha}}$$

$$k^* = \frac{s_k^{\frac{1}{1-\alpha}}}{(n+g+\delta)^{\frac{1}{1-\alpha}}} \left(\frac{s_h^{\frac{\beta}{1-\beta}}}{(n+g+\delta)^{\frac{\beta}{1-\beta}}} k^{*\frac{\alpha\beta}{1-\beta}} \right)^{\frac{1}{1-\alpha}}$$

$$k^* = \frac{s_k^{\frac{1}{1-\alpha}}}{(n+g+\delta)^{\frac{1}{1-\alpha}}} \frac{s_h^{\frac{\beta}{(1-\alpha)(1-\beta)}}}{(n+g+\delta)^{\frac{\beta}{(1-\alpha)(1-\beta)}}} k^{*\frac{\alpha\beta}{(1-\alpha)(1-\beta)}}$$

$$k^{*1-\frac{\alpha\beta}{(1-\alpha)(1-\beta)}} = \frac{s_k^{\frac{1}{1-\alpha}} s_h^{\frac{\beta}{(1-\alpha)(1-\beta)}}}{(n+g+\delta)^{\frac{1}{1-\alpha}+\frac{\beta}{(1-\alpha)(1-\beta)}}}$$

$$k^{*\frac{(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)}-\frac{\alpha\beta}{(1-\alpha)(1-\beta)}} = \frac{s_k^{\frac{1}{1-\alpha}} s_h^{\frac{\beta}{(1-\alpha)(1-\beta)}}}{(n+g+\delta)^{\frac{(1-\beta)}{(1-\alpha)(1-\beta)}+\frac{\beta}{(1-\alpha)(1-\beta)}}}$$

$$k^{*\frac{1-\alpha-\beta+\alpha\beta-\alpha\beta}{(1-\alpha)(1-\beta)}} = \frac{s_k^{\frac{1}{1-\alpha}} s_h^{\frac{\beta}{(1-\alpha)(1-\beta)}}}{(n+g+\delta)^{\frac{1-\beta+\beta}{(1-\alpha)(1-\beta)}}}$$

$$k^{*\frac{1-\alpha-\beta}{(1-\alpha)(1-\beta)}} = \frac{s_k^{\frac{1}{1-\alpha}} s_h^{\frac{\beta}{(1-\alpha)(1-\beta)}}}{(n+g+\delta)^{\frac{1}{(1-\alpha)(1-\beta)}}}$$

$$k^{*\frac{1-\alpha-\beta}{(1-\alpha)(1-\beta)}} = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n+g+\delta} \right)^{\frac{1}{(1-\alpha)(1-\beta)}}$$

$$k^{*\frac{1-\alpha-\beta}{(1-\alpha)(1-\beta)}} = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n+g+\delta} \right)^{\frac{1}{(1-\alpha)(1-\beta)}}$$

$$k^* = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n+g+\delta} \right)^{\frac{(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)(1-\alpha-\beta)}}$$

$$k^* = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n+g+\delta} \right)^{\frac{1}{1-\alpha-\beta}}$$

Proof 13: Speed of convergence

$$\frac{\partial \dot{k}}{\partial k} = s_k \frac{\partial y}{\partial k} - (n+g+\delta) = \frac{(n+g+\delta)k}{y} \frac{\partial y}{\partial k} - (n+g+\delta)$$

$$= (n+g+\delta) \frac{k}{k^{\alpha}h^{\beta}} \left(\alpha k^{\alpha-1}h^{\beta} - 1 \right) = (n+g+\delta)(\alpha-1)$$

$$\frac{\partial \dot{h}}{\partial h} = s_h \frac{\partial y}{\partial h} - (n+g+\delta) = \frac{(n+g+\delta)h}{y} \frac{\partial y}{\partial h} - (n+g+\delta)$$

$$= (n+g+\delta) \frac{h}{k^{\alpha}h^{\beta}} \left(\beta k^{\alpha}h^{\beta-1} - 1 \right) = (n+g+\delta)(\beta-1)$$

$$\lambda = -\frac{\partial \left(\frac{\dot{y}}{y} \right)}{\partial \log(y)} = (n+g+\delta) - \frac{\partial \big((n+g+\delta)(\alpha-1) \big)}{\partial \alpha} - \frac{\partial (n+g+\delta)(\beta-1)}{\partial \beta}$$

$$= (1-\alpha-\beta)(n+g+\delta)$$

Proof 14: Growth rate of growth rate

For capital:

$$\dot{K}(t) = sY(t)$$

$$\dot{K}(t) = s\big((1-a_K)K(t) \big)^{\alpha} \big(A(t)(1-a_L)L(t) \big)^{1-\alpha}$$

$$g_K(t) = \frac{\dot{K}(t)}{K(t)} = s(1-a_K)^{\alpha} K(t)^{\alpha-1} \big(A(t)(1-a_L)L(t) \big)^{1-\alpha}$$

$$\ln\big(g_K(t)\big) = \alpha \ln\big(s(1-a_K)\big) + (\alpha-1)\ln(K(t)) + (1-\alpha)\ln\big(A(t)(1-a_L)L(t)\big)$$

$$\frac{d\big(\ln(g_K(t))\big)}{dt} = \frac{\dot{g}_K(t)}{g_K(t)} = 0 + (\alpha-1)\frac{\dot{K}(t)}{K(t)} + (1-\alpha)\left(\frac{\dot{A}(t)}{A(t)} + 0 + \frac{\dot{L}(t)}{L(t)} \right)$$

$$\frac{\dot{g}_K(t)}{g_K(t)} = (\alpha-1)g_K + (1-\alpha)(g_A+n) = (1-\alpha)(g_A(t)+n-g_K(t))$$

For knowledge:

$$\dot{A}(t) = B\big(a_K K(t)\big)^{\beta}\big(a_L L(t)\big)^{\gamma} A(t)^{\theta}$$

$$\frac{\dot{A}(t)}{A(t)} = g_A(t) = B\big(a_K K(t)\big)^{\beta}\big(a_L L(t)\big)^{\gamma} A(t)^{\theta-1}$$

$$\ln\big(g_A(t)\big) = \ln B + \beta \ln\big(a_K K(t)\big) + \gamma \ln\big(a_L L(t)\big) + (\theta - 1)\ln\big(A(t)\big)$$

$$\frac{d\big(\ln\big(g_A(t)\big)\big)}{dt} = \frac{\dot{g}_A(t)}{g_A(t)} = 0 + \beta\left(0 + \frac{\dot{K}(t)}{K(t)}\right) + \gamma\left(0 + \frac{\dot{L}(t)}{L(t)}\right) + (\theta - 1)\left(\frac{\dot{A}(t)}{A(t)}\right)$$

$$= \beta g_K(t) + \gamma n + (\theta - 1)g_A(t)$$

Proof 15: Equilibrium growth rate of capital and knowledge

$$(1 - \alpha)\big(g_A^*(t) + n - g_K^*(t)\big) = 0$$

$$g_K^* = g_A^* + n$$

$$\beta g_K^*(t) + \gamma n + (\theta - 1)g_A^*(t) = 0$$

$$g_A^*(t) = \frac{\beta g_K^*(t) + \gamma n}{1 - \theta}$$

$$g_A^*(t) = \frac{\beta(g_A^*(t) + n) + \gamma n}{1 - \theta}$$

$$(1 - \theta - \beta)g_A^*(t) = (\beta + \gamma)n$$

$$g_A^*(t) = \frac{\beta + \gamma}{1 - \theta - \beta}n$$

Proof 16: Equilibrium growth rate of output

$$Y(t) = \big((1 - a_K)K(t)\big)^{\alpha}\big(A(t)(1 - a_L)L(t)\big)^{1-\alpha}$$

$$\ln\big(Y(t)\big) = \alpha \ln\big((1 - a_K)K(t)\big) + (1 - \alpha)\ln\big(A(t)(1 - a_L)L(t)\big)$$

$$g_Y(t) = \frac{\dot{Y}(t)}{Y(t)} = \alpha \frac{\dot{K}(t)}{K(t)} + (1 - \alpha)\big(\frac{\dot{A}(t)}{A(t)} + \frac{\dot{L}(t)}{L(t)}\big)$$

$$g_Y^*(t) = \alpha g_K^*(t) + (1 - \alpha)(g_A^*(t) + n)$$

$$g_Y^*(t) = \alpha \left(\frac{\beta + \gamma}{1 - \theta - \beta} n + n \right) + (1 - \alpha) \left(\frac{\beta + \gamma}{1 - \theta - \beta} n + n \right) = \frac{\beta + \gamma}{1 - \theta - \beta} n + n$$

$$= n \left(\frac{\beta + \gamma}{1 - \theta - \beta} + \frac{1 - \theta - \beta}{1 - \theta - \beta} \right) = n \left(\frac{1 + \gamma - \theta}{1 - \theta - \beta} \right) = g_K^*(t)$$

Proof 17: The OLS estimator

$$f(\beta) = (y - X\beta)'(y - X\beta) = y'y - 2y'X\beta + \beta'X'X\beta$$

$$\frac{\partial f(\beta)}{\partial \beta} = -2(X'y - X'X\beta) = 0$$

$$X'X\beta = X'y$$

$$\beta = (X'X)^{-1}X'y$$

Proof 18: Properties of the OLS estimator

$$E[\hat{\beta}] = E[(X'X)^{-1}X'y] = E[\beta + (X'X)^{-1}X'\epsilon] = E[\beta] + E[(X'X)^{-1}X']E[\epsilon] = \beta$$

$$Var[\hat{\beta}] = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] = (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1}$$

Proof 19: Alternative R-squared formulae

$$R^2 = \frac{\widehat{\sigma_{\hat{y}_n}^2}}{\widehat{\sigma_{y_n}^2}} = \frac{\widehat{\sigma_{y_n - e_n}^2}}{\widehat{\sigma_{y_n}^2}} = \frac{\widehat{\sigma_{\hat{y}_n}^2}}{\widehat{\sigma_{y_n}^2}} - \frac{\widehat{\sigma_{e_n}^2}}{\widehat{\sigma_{y_n}^2}} = 1 - \frac{\widehat{\sigma_{e_n}^2}}{\widehat{\sigma_{y_n}^2}}$$

Proof 20: Relative growth rate

$$\Delta \log(Y_t) = \log \left(\frac{Y_t}{Y_{t-1}} \right) = \log \left(\frac{Y_{t-1} + \Delta Y_t}{Y_{t-1}} \right) = \log \left(1 + \frac{\Delta Y_t}{Y_{t-1}} \right) \approx \frac{\Delta Y_t}{Y_{t-1}}$$

Proof 21: Stochastic trend

$$Y_t = Y_0 + \sum_{i=1}^{t} \Delta Y_i = Y_0 + \sum_{i=1}^{t} (\beta + \epsilon_i) = Y_0 + \beta t + \sum_{i=1}^{t} \epsilon_i$$

Proof 22: Variance of heteroscedastic OLS estimator

$$Var[\hat{\beta}|X] = Var[(X'X)^{-1}X'\epsilon|X] = (X'X)^{-1}X'Var[\epsilon|X]X(X'X)^{-1}$$
$$= \sigma^2(X'X)^{-1}X'\Psi X(X'X)^{-1}$$

Proof 23: GLS transformation of regression model

$$\Psi^{-1} = P'P$$

$$\Psi = (P'P)^{-1} = P^{-1}(P')^{-1}$$

$$P\Psi P' = PP^{-1}(P')^{-1}P' = I$$

$$Py = \tilde{y}$$

$$PX\beta + P\epsilon = \tilde{X}\beta + \tilde{\epsilon}$$

$$\tilde{y} = \tilde{X}\beta + \tilde{\epsilon}$$

$$E[\tilde{\epsilon}|X] = E[P\epsilon|X] = PE[\epsilon|X] = 0$$

$$Var[\tilde{\epsilon}|X] = Var[P\epsilon|X] = PVar[\epsilon|X]P' = \sigma^2 P\Psi P' = \sigma^2 I$$

Proof 24: Extended growth-initial level regression

$$y^* = \left(\frac{s}{n+g+\delta}\right)^{\frac{\alpha}{1-\alpha}}$$

$$ln\, y^* = \frac{\alpha}{1-\alpha}ln\, s - \frac{\alpha}{1-\alpha}ln(n+g+\delta)$$

$$f(y(t)) = ln\, y(t)$$

$$f(y^*) = ln\, y^*$$

$$f'(y(t)) = -\lambda\left(f(y(t)) - f(y^*)\right)$$

$$\frac{\partial f(y(t))}{\partial t} = -\lambda\left(f(y(t)) - f(y^*)\right)$$

$$\int \frac{1}{f(y(t)) - f(y^*)}\partial f(y(t)) = \int -\lambda\, \partial t$$

$$\ln\left(f(y(t)) - f(y^*)\right) = -\lambda t + c$$

$$f(y(t)) - f(y^*) = e^{-\lambda t + c}$$

$$f(y(0)) - f(y^*) = e^{-\lambda * 0 + c} = e^c$$

$$f(y(t)) = f(y^*) + e^{-\lambda t}\left(f(y(0)) - f(y^*)\right) = (1 - e^{-\lambda t})f(y^*) + e^{-\lambda t}f(y(0))$$

$$f(y(t)) - f(y(0)) = (1 - e^{-\lambda t})f(y^*) + e^{-\lambda t}f(y(0)) - f(y(0))$$
$$= (1 - e^{-\lambda t})\left(f(y^*) - f(y(0))\right)$$

$$\ln y(t) - \ln y(0) = (1 - e^{-\lambda t})(\ln y^* - \ln y(0))$$

$$\ln\left(\frac{y(t)}{y(0)}\right) = (1 - e^{-\lambda t})\left(\frac{\alpha}{1 - \alpha}\ln s - \frac{\alpha}{1 - \alpha}\ln(n + g + \delta) - \ln y(0)\right)$$

$$\ln\left(\frac{y_t}{y_{t-1}}\right) = (1 - e^{-\lambda t})\frac{\alpha}{1 - \alpha}\ln s_t - (1 - e^{-\lambda t})\frac{\alpha}{1 - \alpha}\ln(n_t + g_t + \delta_t)$$
$$- (1 - e^{-\lambda t})\ln y_{t-1} + \epsilon_t$$

$$\ln y_t = (1 - e^{-\lambda t})\frac{\alpha}{1 - \alpha}\ln\left(\frac{s_t}{n_t + g_t + \delta_t}\right) + e^{-\lambda t}\ln y_{t-1} + \epsilon_t$$

Proof 25: Extended growth-initial level regression for the augmented Solow model

$$k^* = \left(\frac{s_k^{1-\beta} s_h^\beta}{n + g + \delta}\right)^{\frac{1}{1-\alpha-\beta}}$$

$$h^* = \left(\frac{s_k^\alpha s_h^{1-\alpha}}{n + g + \delta}\right)^{\frac{1}{1-\alpha-\beta}}$$

$$y^* = k^{*\alpha}h^{*\beta} = \left(\frac{s_k^{1-\beta} s_h^\beta}{n + g + \delta}\right)^{\frac{\alpha}{1-\alpha-\beta}}\left(\frac{s_k^\alpha s_h^{1-\alpha}}{n + g + \delta}\right)^{\frac{\beta}{1-\alpha-\beta}}$$

$$ln \, y^* = \frac{\alpha}{1-\alpha-\beta} \ln s_k^{1-\beta} s_h^{\beta} - \frac{\alpha}{1-\alpha-\beta} ln(n+g+\delta) + \frac{\beta}{1-\alpha-\beta} \ln s_k^{\alpha} s_h^{1-\alpha}$$

$$- \frac{\beta}{1-\alpha-\beta} ln(n+g+\delta)$$

$$= \frac{\alpha(1-\beta)+\alpha\beta}{1-\alpha-\beta} \ln s_k + \frac{\alpha\beta+(1-\alpha)\beta}{1-\alpha-\beta} \ln s_h - \frac{\alpha+\beta}{1-\alpha-\beta} ln(n+g+\delta)$$

$$= \frac{\alpha}{1-\alpha-\beta} \ln s_k + \frac{\beta}{1-\alpha-\beta} \ln s_h - \frac{\alpha+\beta}{1-\alpha-\beta} ln(n+g+\delta)$$

$$= \frac{\alpha}{1-\alpha-\beta} \ln\left(\frac{s_k}{n+g+\delta}\right) + \frac{\beta}{1-\alpha-\beta} \ln\left(\frac{s_h}{n+g+\delta}\right)$$

$$ln\left(\frac{y(t)}{y(0)}\right) = (1-e^{-\lambda t})\left(\frac{\alpha}{1-\alpha-\beta} \ln\left(\frac{s_k}{n+g+\delta}\right) + \frac{\beta}{1-\alpha-\beta} \ln\left(\frac{s_h}{n+g+\delta}\right) - ln \, y(0)\right)$$

$$ln\left(\frac{y_t}{y_{t-1}}\right) = (1-e^{-\lambda t})\left(\frac{\alpha}{1-\alpha-\beta} \ln\left(\frac{s_{k_t}}{n_t+g_t+\delta_t}\right) + \frac{\beta}{1-\alpha-\beta} \ln\left(\frac{s_{h_t}}{n_t+g_t+\delta_t}\right)\right)$$

$$- (1-e^{-\lambda t}) ln \, y_{t-1} + \epsilon_t$$

$$\ln y_t = (1-e^{-\lambda t})\frac{\alpha}{1-\alpha-\beta} \ln\left(\frac{s_{k_t}}{n_t+g_t+\delta_t}\right) + (1-e^{-\lambda t})\frac{\beta}{1-\alpha-\beta} \ln\left(\frac{s_{h_t}}{n_t+g_t+\delta_t}\right)$$

$$+ e^{-\lambda t} \ln y_{t-1} + \epsilon_t$$

7.2 STATA DO-FILE

```
1    clear all
2    set more off
3    version 15.1
4    use "https://www.rug.nl/ggdc/docs/pwt90.dta"
5    keep country year rgdpna rconna rkna pop emp hc delta
6    describe
7    sort country
8    encode country, gen(country_n)
9    sort country_n year
10   xtset country_n year
11
12   //Sample selection
13   gen SS=(pop!=.&emp!=.&hc!=.&rgdpna!=.&rconna!=.&rkna!=.&delta!=.)
14   label variable SS "Sample selection"
15   gen obs=0
16   label variable obs "Observations per country"
17 ┌ forvalues i=1/182 {
18 │ quietly sum SS if SS!=0 & country_n==`i'
19 │ quietly replace obs=r(N) if SS!=0 & year==2014 & country_n==`i'
20 └ }
21 ┌ forvalues i=2013(-1)1950 {
22 │     quietly by country_n: replace obs=obs[_n+1]-1 if obs[_n+1]!=0 &
   ↳ │ year==`i'
23 └ }
24   histogram obs if year==2014, freq xlabel(0(5)65)
25   replace SS=(obs>=year-1969&year>=1970)
26   tab country if SS==1
27   drop obs
28
29   //Variables
30   gen rgdppc=rgdpna/pop
31   label variable rgdppc "Real GDP per capita (in 2011US$)"
32   gen y=rgdpna/emp
33   label variable y "Real GDP per worker (in 2011US$)"
34   gen k=rkna/emp
35   label variable k "Real capital stock per worker (in 2011US$)"
36   gen c=rconna/emp
37   label variable c "Consumption per worker (in 2011US$)"
38   gen s=(y-c)/y
39   label variable s "Savings rate (%)"
40   gen ln_L=ln(pop)
41   by country_n: gen n=ln_L-ln_L[_n-1]
42   label variable n "Population growth (%)"
43   gen ln_A=ln(emp/pop)
44   by country_n: gen g=ln_A-ln_A[_n-1]
45   label variable g "Technological progress (%)"
46   gen OECD=(country=="Australia"|country=="Austria"|country=="Belgium"|
   ↳ country=="Canada"|country=="Chile"|country=="Czech Republic"|country==
   ↳ "Denmark"|country=="Estonia"|country=="Finland"|country=="France"|
   ↳ country=="Germany"|country=="Greece"|country=="Hungary"|country==
   ↳ "Iceland"|country=="Ireland"|country=="Israel"|country=="Italy"|
   ↳ country=="Japan"|country=="Korea"|country=="Latvia"|country==
   ↳ "Luxembourg"|country=="Mexico"|country=="Netherlands"|country=="New
   ↳ Zealand"|country=="Norway"|country=="Poland"|country=="Portugal"|
```

```stata
        country=="Slovak Republic"|country=="Slovenia"|country=="Spain"|
        country=="Sweden"|country=="Switzerland"|country=="Turkey"|country==
        "United Kingdom"|country=="United States")
47      label variable OECD "OECD country (dummy)"
48
49      //Logarithmic transformations
50      gen ln_rgdppc=ln(rgdppc)
51      gen ln_y=ln(y)
52      gen ln_k=ln(k)
53      gen ln_y_ss=ln(s/(n+g+delta))
54      gen ln_hc=ln(hc)
55      gen ln_Y=ln(rgdpna)
56      gen ln_K=ln(rkna)
57      gen ln_AL=ln(emp)
58
59      //GDP per capita vs worker
60      bysort year: egen rgdppc_mean=mean(rgdppc) if SS==1
61      by year: egen y_mean=mean(y) if SS==1
62      label var rgdppc_mean "Average real GDP per capita"
63      label var y_mean "Average real GDP per worker"
64      tsline rgdppc_mean y_mean if year>=1970, xlabel(1970(10)2014)
65      by year: egen n_mean=mean(n) if SS==1
66      by year: egen g_mean=mean(g) if SS==1
67      label var n_mean "Average population growth"
68      label var g_mean "Average technological progress"
69      twoway (scatter n_mean year) (scatter g_mean year) if year>=1970,
        ylabel(0 "0%" .01 "1%" .02 "2%")
70
71      //Growth
72      sort country_n year
73      by country_n: gen g_y=((ln_y[65]-ln_y[_n])/(65-_n))
74      label variable g_y "Annual growth rate of real GDP per worker"
75
76      //Growth-initial level regression
77      twoway (scatter g_y ln_y) (lfit g_y ln_y) if SS==1 & year==1970, title
        ("Growth-initial level regression") ytitle("Annual growth of real GDP
        per worker 1970-2014") xtitle("Log real GDP per worker 1970") ylabel(-
        .05 "-5%" -.025 "-2.5%" 0 "0%" .025 "2.5%" .05 "5%") xlabel(7(2)13)
78      reg g_y ln_y if SS==1 & year==1970
79      predict r, resi
80      pnorm r
81      rvfplot
82      estat hettest r
83      imtest
84      drop r
85      reg g_y ln_y if SS==1 & year==1970, robust
86
87      twoway (scatter g_y ln_y, mlabel(country) msize(small)) (lfit g_y ln_y
        ) if SS==1 & year==1970 & OECD==1, title("Growth-initial level
        regression - OECD countries") ytitle("Annual growth of real GDP per
        worker 1970-2014") xtitle("Log real GDP per worker 1970") ylabel(0
        "0%" .01 "1%" .02 "2%" .03 "3%") xlabel(9.75(0.5)11.25)
88      reg g_y ln_y if SS==1 & year==1970 & OECD==1
89      predict r, resi
90      pnorm r
91      rvfplot
```

```stata
 92     estat hettest r
 93     imtest
 94     predict f, rstud
 95     predict h, hat
 96     gen f2=f^2
 97     label var f2 "Studentized residual squared"
 98     sum h f2 if SS==1 & year==1970 & OECD==1
 99     twoway scatter h f2 if SS==1 & year==1970 & OECD==1, mlabel(country)
  ↳    msize(tiny) yline(0.068955) xline(1.305577)
100     drop h f f2 r
101
102     //Sigma convergence
103     bysort year: egen y_sd=sd(y) if SS==1
104     by year: egen y_sd_oecd=sd(y) if SS==1 & OECD==1
105     label var y_sd "Standard deviation of real GDP per worker"
106     label var y_sd_oecd "Standard deviation of real GDP per worker - OECD
  ↳    countries"
107     tsline y_sd y_sd_oecd if year>=1970, xlabel(1970(10)2014)
108
109     //Sample selection 2
110     gen obs=0
111     forvalues i=1/182 {
112         forvalues j=1971/2014 {
113             quietly sum ln_y_ss if year<=`j' & country_n==`i' & year>=1971
114             quietly replace obs=r(N) if obs[_n-1]==r(N)-1 & year==`j' &
  ↳    country_n==`i'
115         }
116     } //Warning: takes time
117     gen Solow=0
118     by country_n: replace Solow=1 if (obs==44 & year==2014)
119     forvalues i=2013(-1)1971 {
120         quietly by country_n: replace Solow=1 if Solow[_n+1]==1 & year==`i'
121     }
122     drop obs
123     tab country if Solow==1 & SS==0
124     replace Solow=0 if country=="Oman"
125     tab country if Solow==1
126
127     //Solow model analysis
128     reg ln_Y ln_K ln_AL if SS==1
129     encode country if SS==1, gen(country_SS)
130     bysort country_n: egen ln_Y_imean=mean(ln_Y) if SS==1
131     twoway (scatter ln_Y country_SS, msymbol(circle_hollow)) (connected
  ↳    ln_Y_imean country_SS) if SS==1, xlabel(0(25)101)
132     xtreg ln_Y ln_K ln_AL if SS==1, fe
133     xttest3
134     xtreg ln_Y ln_K ln_AL if SS==1, fe vce(cluster country_n)
135     xtserial ln_Y ln_K ln_AL if SS==1
136     xtgls ln_Y ln_K ln_AL if SS==1, panels(hetero) corr(psar1)
137     test ln_K+ln_AL=1
138
139     reg ln_y ln_y_ss L.ln_y if Solow==1
140     encode country if Solow==1, gen(country_Solow)
141     bysort country_n: egen ln_y_imean=mean(ln_y) if Solow==1
142     twoway (scatter ln_y country_Solow, msymbol(circle_hollow)) (connected
  ↳     ln_y_imean country_Solow) if Solow==1, xlabel(0(13)53) ylabel(7(2)13)
```

```
143    xtreg ln_y ln_y_ss L.ln_y if Solow==1, fe
144    xttest3
145    xtreg ln_y ln_y_ss L.ln_y if Solow==1, fe vce(cluster country_n)
146    xtgls ln_y ln_y_ss L.ln_y if Solow==1, panels(hetero) corr(psar1)
147
148    xtreg ln_y ln_y_ss L.ln_y ln_hc if Solow==1, fe
149    xttest3
150    xtreg ln_y ln_y_ss L.ln_y ln_hc if Solow==1, fe vce(cluster country_n)
151    xtgls ln_y ln_y_ss L.ln_y ln_hc if Solow==1, panels(hetero) corr(psar1)
152
153    //New gowth theory
154    by country_n: gen gr_Y=(rgdpna-rgdpna[_n-1])/rgdpna[_n-1]
155    by country_n: gen gr_K=(rkna-rkna[_n-1])/rkna[_n-1]
156    reg gr_Y n if SS==1 & year>=1971
157    xtreg gr_Y n if SS==1 & year>=1971, fe
158    xttest3
159    estimates store fixed
160    xtreg gr_Y n if SS==1 & year>=1971, re
161    estimates store random
162    hausman fixed random
163    xttest0
164    xtreg gr_Y n if SS==1 & year>=1971, re cluster(country_n)
```

7.3 REGRESSION OUTPUTS

Regression   output   1:

```
      Source |       SS           df       MS      Number of obs   =       101
-------------+----------------------------------   F(1, 99)        =     24.76
       Model |  .004803638         1  .004803638   Prob > F        =    0.0000
    Residual |   .01920985        99  .000194039   R-squared       =    0.2000
-------------+----------------------------------   Adj R-squared   =    0.1920
       Total |  .024013488       100  .000240135   Root MSE        =    .01393


         g_y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_y |  -.0056361   .0011328    -4.98   0.000    -.0078838   -.0033885
       _cons |   .0677053   .0111282     6.08   0.000     .0456246    .089786
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
        Ho: Constant variance
        Variables: r

        chi2(1)     =      6.51
        Prob > chi2 =    0.0107
```

```
Cameron & Trivedi's decomposition of IM-test
```

```
              Source |     chi2     df       p

  Heteroskedasticity |     5.76      2   0.0562
            Skewness |     2.54      1   0.1107
            Kurtosis |     0.00      1   0.9557

               Total |     8.30      4   0.0810
```

```
Linear regression                              Number of obs   =       101
                                               F(1, 99)        =     16.93
                                               Prob > F        =    0.0001
                                               R-squared       =    0.2000
                                               Root MSE        =    .01393
```

```
                             Robust
         g_y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ln_y |  -.0056361      .00137   -4.11   0.000    -.0083545   -.0029178
       _cons |   .0677053      .01358    4.99   0.000     .0407596    .094651
```

Regression output 2:

Source	SS	df	MS		
Model	.000400865	1	.000400865	Number of obs =	29
Residual	.00063277	27	.000023436	F(1, 27) =	17.10
				Prob > F =	0.0003
				R-squared =	0.3878
				Adj R-squared =	0.3651
Total	.001033635	28	.000036916	Root MSE =	.00484

g_y	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
ln_y	-.0103399	.0025001	-4.14	0.000	-.0154697	-.0052101
_cons	.1249485	.0263553	4.74	0.000	.070872	.179025

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
        Ho: Constant variance
        Variables: r

        chi2(1)   =   5.01
        Prob > chi2 =   0.0252

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	0.42	2	0.8099
Skewness	0.34	1	0.5605
Kurtosis	2.64	1	0.1042
Total	3.40	4	0.4933

Regression output 3:

Source	SS	df	MS		
Model	13105.331	2	6552.6655	Number of obs =	4,545
Residual	798.710009	4,542	.175849848	F(2, 4542) =	37262.84
				Prob > F =	0.0000
				R-squared =	0.9426
				Adj R-squared =	0.9425
Total	13904.041	4,544	3.05986818	Root MSE =	.41934

ln_Y	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
ln_K	.8016796	.0045705	175.40	0.000	.7927192	.81064
ln_AL	.1978789	.0052307	37.83	0.000	.1876242	.2081336
_cons	1.121765	.0531488	21.11	0.000	1.017568	1.225963

Regression output 4:

```
Fixed-effects (within) regression              Number of obs     =      4,545
Group variable: country_n                      Number of groups  =        101

R-sq:                                          Obs per group:
    within  = 0.8949                                         min =         45
    between = 0.9297                                         avg =       45.0
    overall = 0.9265                                         max =         45

                                               F(2,4442)         =   18912.83
corr(u_i, Xb)  = 0.2276                         Prob > F          =     0.0000
```

ln_Y	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
ln_K	.6231527	.0066148	94.21	0.000	.6101844	.636121
ln_AL	.3542276	.0109022	32.49	0.000	.3328539	.3756014
_cons	3.137903	.0712101	44.07	0.000	2.998295	3.27751
sigma_u	.4597528					
sigma_e	.16454165					
rho	.88645694	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(100, 4442) = 250.59             Prob > F = 0.0000


Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (101)  =   2.1e+05
Prob>chi2 =      0.0000


Fixed-effects (within) regression              Number of obs     =      4,545
Group variable: country_n                      Number of groups  =        101

R-sq:                                          Obs per group:
    within  = 0.8949                                         min =         45
    between = 0.9297                                         avg =       45.0
    overall = 0.9265                                         max =         45

                                               F(2,100)          =     684.87
corr(u_i, Xb)  = 0.2276                         Prob > F          =     0.0000

                    (Std. Err. adjusted for 101 clusters in country_n)
```

ln_Y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
ln_K	.6231527	.0363589	17.14	0.000	.5510177	.6952877
ln_AL	.3542276	.0616164	5.75	0.000	.2319824	.4764729
_cons	3.137903	.3867716	8.11	0.000	2.370559	3.905247
sigma_u	.4597528					
sigma_e	.16454165					
rho	.88645694	(fraction of variance due to u_i)				

```
Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,     100) =    234.739
           Prob > F =      0.0000
```

Regression output 5:

```
Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   panel-specific AR(1)

Estimated covariances      =       101      Number of obs       =        4,545
Estimated autocorrelations =       101      Number of groups    =          101
Estimated coefficients     =         3      Time periods        =           45
                                            Wald chi2(2)        =     41338.98
                                            Prob > chi2         =       0.0000

─────────────┬────────────────────────────────────────────────────────────────
        ln_Y │     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
─────────────┼────────────────────────────────────────────────────────────────
        ln_K │  .7338368   .0062708   117.02   0.000     .7215462    .7461274
       ln_AL │  .2727956   .0078527    34.74   0.000     .2574046    .2881866
       _cons │    1.9312   .0747066    25.85   0.000     1.784778    2.077622
─────────────┴────────────────────────────────────────────────────────────────
```

Regression output 6:

```
      Source │       SS           df       MS      Number of obs   =      2,332
─────────────┼──────────────────────────────────   F(2, 2329)      >  99999.00
       Model │  3151.04132         2  1575.52066   Prob > F        =     0.0000
    Residual │  4.49204882     2,329  .001928746   R-squared       =     0.9986
─────────────┼──────────────────────────────────   Adj R-squared   =     0.9986
       Total │  3155.53337     2,331  1.35372517   Root MSE        =     .04392

─────────────┬────────────────────────────────────────────────────────────────
        ln_y │     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
─────────────┼────────────────────────────────────────────────────────────────
     ln_y_ss │   .006727   .0013655     4.93   0.000     .0040492    .0094048
             │
        ln_y │
         L1. │  .9943759   .0008184  1215.07   0.000     .9927711    .9959807
             │
       _cons │  .0610169   .0081397     7.50   0.000     .0450552    .0769787
─────────────┴────────────────────────────────────────────────────────────────
```

Regression output 7:

```
Fixed-effects (within) regression          Number of obs     =        2,332
Group variable: country_n                  Number of groups  =           53

R-sq:                                      Obs per group:
    within  = 0.9781                                    min =           44
    between = 0.9997                                    avg =         44.0
    overall = 0.9985                                    max =           44

                                           F(2,2277)         =     50854.64
corr(u_i, Xb)  = 0.7063                     Prob > F          =       0.0000
```

ln_y	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
ln_y_ss	.0186022	.001991	9.34	0.000	.0146979	.0225065
ln_y						
L1.	.9759862	.0031362	311.20	0.000	.9698362	.9821362
_cons	.2357188	.0320148	7.36	0.000	.1729375	.2985
sigma_u	.02692247					
sigma_e	.04092971					
rho	.30200043	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(52, 2277) = 7.78              Prob > F = 0.0000


Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (53)  =   47495.39
Prob>chi2 =      0.0000
```

```
Fixed-effects (within) regression              Number of obs     =      2,332
Group variable: country_n                      Number of groups  =         53

R-sq:                                          Obs per group:
    within  = 0.9782                                       min =         44
    between = 0.9998                                       avg =       44.0
    overall = 0.9986                                       max =         44

                                               F(3,52)           =    6605.64
corr(u_i, Xb)  = 0.7537                         Prob > F          =     0.0000

                            (Std. Err. adjusted for 53 clusters in country_n)
─────────────────────────────────────────────────────────────────────────────
                         Robust
      ln_y |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
───────────┼─────────────────────────────────────────────────────────────────
   ln_y_ss |   .0180847   .0040518     4.46   0.000     .0099542    .0262152
           │
      ln_y │
       L1. |   .9721523   .0072917   133.32   0.000     .9575205    .9867841
           │
     ln_hc |    .019578   .0110304     1.77   0.082    -.0025561    .0417122
     _cons |   .2605209   .0742457     3.51   0.001     .1115361    .4095057
───────────┼─────────────────────────────────────────────────────────────────
   sigma_u |  .02594951
   sigma_e |  .04085974
       rho |  .28741235   (fraction of variance due to u_i)
─────────────────────────────────────────────────────────────────────────────
```

Regression output 8:

```
Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   panel-specific AR(1)

Estimated covariances      =        53    Number of obs     =      2,332
Estimated autocorrelations =        53    Number of groups  =         53
Estimated coefficients     =         3    Time periods      =         44
                                          Wald chi2(2)      =    1130477
                                          Prob > chi2       =     0.0000

─────────────────────────────────────────────────────────────────────────────
      ln_y |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
───────────┼─────────────────────────────────────────────────────────────────
   ln_y_ss |   .0109273    .001233     8.86   0.000     .0085106     .013344
           │
      ln_y │
       L1. |   .9927044   .0010213   971.98   0.000     .9907027    .9947062
           │
     _cons |   .0752894   .0104402     7.21   0.000      .054827    .0957519
─────────────────────────────────────────────────────────────────────────────
```

Regression output 9:

```
Fixed-effects (within) regression            Number of obs     =      2,332
Group variable: country_n                    Number of groups  =         53

R-sq:                                        Obs per group:
    within  = 0.9782                                       min =         44
    between = 0.9998                                       avg =       44.0
    overall = 0.9986                                       max =         44

                                             F(3,2276)         =   34022.24
corr(u_i, Xb)  = 0.7537                      Prob > F          =     0.0000
```

ln_y	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
ln_y_ss	.0180847	.0019952	9.06	0.000	.0141721	.0219973
ln_y L1.	.9721523	.0033869	287.03	0.000	.9655105	.9787941
ln_hc	.019578	.0065978	2.97	0.003	.0066397	.0325164
_cons	.2605209	.033035	7.89	0.000	.1957391	.3253027
sigma_u	.02594951					
sigma_e	.04085974					
rho	.28741235	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(52, 2276) = 5.62            Prob > F = 0.0000


Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (53)  =   47323.60
Prob>chi2  =     0.0000
```

```
Fixed-effects (within) regression              Number of obs      =      2,332
Group variable: country_n                      Number of groups   =         53

R-sq:                                          Obs per group:
     within  = 0.9782                                       min =         44
     between = 0.9998                                       avg =       44.0
     overall = 0.9986                                       max =         44

                                               F(3,52)            =    6605.64
corr(u_i, Xb)  = 0.7537                         Prob > F           =     0.0000

                              (Std. Err. adjusted for 53 clusters in country_n)
─────────────────────────────────────────────────────────────────────────────
                              Robust
        ln_y │     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
─────────────┼───────────────────────────────────────────────────────────────
     ln_y_ss │  .0180847   .0040518     4.46   0.000     .0099542    .0262152

        ln_y │
         L1. │  .9721523   .0072917   133.32   0.000     .9575205    .9867841

       ln_hc │   .019578   .0110304     1.77   0.082    -.0025561    .0417122
       _cons │  .2605209   .0742457     3.51   0.001     .1115361    .4095057
─────────────┼───────────────────────────────────────────────────────────────
     sigma_u │  .02594951
     sigma_e │  .04085974
         rho │  .28741235   (fraction of variance due to u_i)
─────────────────────────────────────────────────────────────────────────────
```

Regression output 10:

```
Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   panel-specific AR(1)

Estimated covariances      =         53     Number of obs      =      2,332
Estimated autocorrelations =         53     Number of groups   =         53
Estimated coefficients     =          4     Time periods       =         44
                                            Wald chi2(3)       =    1297245
                                            Prob > chi2        =     0.0000

─────────────────────────────────────────────────────────────────────────────
        ln_y │     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
─────────────┼───────────────────────────────────────────────────────────────
     ln_y_ss │  .0096391    .001194     8.07   0.000     .0072988    .0119794

        ln_y │
         L1. │  .9875006   .0013946   708.11   0.000     .9847673    .9902339

       ln_hc │  .0264094   .0040883     6.46   0.000     .0183965    .0344223
       _cons │  .1086098   .0119475     9.09   0.000     .0851931    .1320265
─────────────────────────────────────────────────────────────────────────────
```

Regression output 11:

```
Fixed-effects (within) regression              Number of obs     =      4,444
Group variable: country_n                      Number of groups  =        101

R-sq:                                          Obs per group:
    within  = 0.0076                                        min =         44
    between = 0.1675                                        avg =       44.0
    overall = 0.0197                                        max =         44

                                               F(1,4342)         =      33.40
corr(u_i, Xb)  = 0.0034                         Prob > F          =     0.0000

─────────────────────────────────────────────────────────────────────────────
        gr_Y │      Coef.    Std. Err.       t    P>|t|    [95% Conf. Interval]
─────────────┼───────────────────────────────────────────────────────────────
           n │   .4587507    .0793738     5.78    0.000    .3031374    .6143639
       _cons │   .0295828    .0016489    17.94    0.000    .0263501    .0328155
─────────────┼───────────────────────────────────────────────────────────────
     sigma_u │   .01351428
     sigma_e │   .05199174
         rho │   .06328828    (fraction of variance due to u_i)
─────────────────────────────────────────────────────────────────────────────
F test that all u_i=0: F(100, 4342) = 2.97                 Prob > F = 0.0000


Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (101)  =  13226.77
Prob>chi2 =     0.0000


Random-effects GLS regression                  Number of obs     =      4,444
Group variable: country_n                      Number of groups  =        101

R-sq:                                          Obs per group:
    within  = 0.0076                                        min =         44
    between = 0.1675                                        avg =       44.0
    overall = 0.0197                                        max =         44

                                               Wald chi2(1)      =      53.34
corr(u_i, X)  = 0 (assumed)                     Prob > chi2       =     0.0000

─────────────────────────────────────────────────────────────────────────────
        gr_Y │      Coef.    Std. Err.       z    P>|z|    [95% Conf. Interval]
─────────────┼───────────────────────────────────────────────────────────────
           n │   .4603912     .063039     7.30    0.000     .336837    .5839454
       _cons │   .0295528    .0017769    16.63    0.000    .0260701    .0330355
─────────────┼───────────────────────────────────────────────────────────────
     sigma_u │   .01109244
     sigma_e │   .05199174
         rho │   .04353654    (fraction of variance due to u_i)
─────────────────────────────────────────────────────────────────────────────
```

```
.                 ──── Coefficients ────
              │      (b)          (B)          (b-B)      sqrt(diag(V_b-V_B))
              │    fixed         random       Difference          S.E.
──────────────┼─────────────────────────────────────────────────────────────
            n │  .4587507      .4603912      -.0016405          .0482316
──────────────┴─────────────────────────────────────────────────────────────

                      b = consistent under Ho and Ha; obtained from xtreg
          B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

              chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =         0.00
            Prob>chi2 =       0.9729


Breusch and Pagan Lagrangian multiplier test for random effects

        gr_Y[country_n,t] = Xb + u[country_n] + e[country_n,t]

        Estimated results:
                          │        Var      sd = sqrt(Var)
                     ─────┼──────────────────────────────
                    gr_Y  │     .0028793        .0536589
                       e  │     .0027031        .0519917
                       u  │      .000123        .0110924

        Test:   Var(u) = 0
                              chibar2(01) =    171.06
                            Prob > chibar2 =   0.0000


Random-effects GLS regression              Number of obs     =      4,444
Group variable: country_n                  Number of groups  =        101

R-sq:                                       Obs per group:
    within  = 0.0076                                    min =         44
    between = 0.1675                                    avg =       44.0
    overall = 0.0197                                    max =         44

                                            Wald chi2(1)      =       5.80
corr(u_i, X)   = 0 (assumed)               Prob > chi2       =     0.0160

                          (Std. Err. adjusted for 101 clusters in country_n)
─────────────────────────────────────────────────────────────────────────────
              │             Robust
         gr_Y │      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
──────────────┼──────────────────────────────────────────────────────────────
            n │   .4603912   .1910892     2.41   0.016     .0858634    .8349191
        _cons │   .0295528   .0035916     8.23   0.000     .0225133    .0365922
──────────────┼──────────────────────────────────────────────────────────────
      sigma_u │   .01109244
      sigma_e │   .05199174
          rho │   .04353654   (fraction of variance due to u_i)
─────────────────────────────────────────────────────────────────────────────
```

7.4 REFLECTION NOTE

In my master thesis, I have applied empirical econometric methods to the study of macroeconomics and economic growth. Relevant equations, mostly from macroeconomic theory, have been derived and proven mathematically. Data has been collected from The Penn World tables which is a famous and well-maintained database. Tests and analysis have been performed on the sample data using methods of linear regression, time series and panel data analysis using the statistical software Stata.

As a basis prior to starting the thesis, I benefited from knowledge of advanced macroeconomic theory that I gained during my exchange period at the University of Economics in Prague. This knowledge made it possible for me to efficiently conduct preliminary research and understand the motivation of the debate on convergence. I have gained personal interest in the topic of convergence and have found the process of writing the thesis to be both academically challenging and rewarding. My master thesis is a highly representative pinnacle of both my bachelor and master programs. During the bachelor program of Mathematical Finance, I was provided with a comprehensive set of tools to approach and understand mathematical and statistical aspects that are essential in econometrics as well as an in economic and financial theory.

The results of my research reveal significant tendencies of convergence between countries. The convergence however, is not persistent and is greatly affected by unobserved characteristics that are unexplained by the neoclassical growth theory. The augmented Solow model includes the factor of human capital in the model which is proved to help with consistency between the neoclassical growth theory and the empirical results. The results motivate for further research that includes other characteristics.

Studying economic growth is important for the understanding of movements in the world income distribution and the welfare of individuals. The goal of economic growth research is to better understand the economic dynamics so as to enable pursuit of policies that increases standards of living and decreases world poverty. These are among the goals of international organizations such as The Organization for Economic Co-operation and Development (OECD) and The United Nations (UN). With drastically increasing globalization, countries become more interdependent and increasingly similar to each other in many ways. Therefore, the question of convergence is tightly connected to globalization, international markets and trade as well as international policies and agreements.

Macroeconomic theory aims to explain as much as possible of the economic behavior of economies through common characteristics. One characteristic is technology and how technological progress takes place. Technology, in many cases, has spillover effects such as when countries succeed in acquiring new technology that is created or realized by other countries through international trade or through the exchange of knowledge. Technology and knowledge in this thesis are the same and is defined as the employment rate. This implies that the increase in employment rate is driven by technological progress, also called innovative ideas. Innovative ideas being defined as only those which contribute to creating new jobs and increasing the employment rate. In real life cases, this is not always true but innovative ideas and entrepreneurship are nevertheless important drivers of creating new jobs.

Innovation in economic growth research is much needed. As my research shows, there are significant unobserved characteristics of economic growth that explain country specific differences. Innovation in economic growth can be achieved through identifying and measuring these characteristics. Observing and maintaining observations for as many countries done for The Penn World Tables requires significant effort. The Penn World Tables have included a measure for human capital only in recent versions. This shows the magnitude of work behind introducing an idea of a factor to measuring and collecting data for the quantity of countries in the world. Filling these data gaps increases the knowledge base for understanding aspects such as prosperity.

Policies that increase standards of living and decrease world poverty are of interest to the general public and considered to be a globally shared responsibility. However, there are policies that have the opposite effect on global welfare such as anti-competition, tax wars and protectionism. These policies are often strongly connected to political beliefs such as nationalism without regards to actual knowledge about economic dynamics. Motivations behind different political ideologies and philosophies are important to understand when predicting the dynamics of international prosperity. Consequently, I believe that this should also be taught in business schools in a larger extent, specifically the background for international policy making and how seemingly unethical policies and trades affect world income distribution and the welfare of individuals.

In conclusion, I am grateful for the opportunity of studying Mathematical Finance as my bachelor program before the program was unfortunately discontinued. I am also grateful for the exchange period which gave me new insights as well as a new perspective of international academia. I am genuinely convinced that the knowledge and understanding I have acquired

through the master program at the University of Agder will make a significant difference for me at the onset of my professional career, and to my ability to successfully contribute constructively in our quest to better our common globe.

8 REFERENCES

Aghion, P., & Howitt, P. (1992). A Model of Growth through Creative Destruction. *Econometrica 60 (March)*, 323-351.

Barro, R. J., & Sala-i-Martin, X. (2004). *Economic growth* (2nd ed.). Cambridge, Mass: The MIT Press.

Bartlett, M. S. (1946). On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series. *Supplement to the Journal of the Royal Statistical Society, 8*(1), 27-41.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed ed.). New York: New York : John Wiley & Sons, Incorporated.

Box, G. E. P., & Pierce, D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association, 65*(332), 1509-1526.

Breusch, T. S. (1978). Testing for Autocorrelation in Dnamic Linear Models. *Australian Economic Papers, 17*(31), 334-355.

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies, 47*(1), 239-253.

Cobb, C. W., & Douglas, P. H. (1928). A Theory of Production. *The American Economic Review, 18*(1), 139-165.

Cochrane, D., & Orcutt, G. H. (1949). Application of Least Squares Regression to Relationships Containing Auto- Correlated Error Terms. *Journal of the American Statistical Association, 44*(245), 32-61.

D'Agostino, R. B., & Belanger, A. (1990). A Suggestion for Using Powerful and Informative Tests of Normality. *The American Statistician, 44*(4), 316-321.

Devore, J. L., & Berk, K. N. (2012). *Modern Mathematical Statistics with Applications*. New York, NY: Springer New York, New York, NY.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association, 74*(366), 427-431.

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *Stata Journal, 3*(2), 168-177.

Durbin, J., & Watson, G. S. (1971). Testing for Serial Correlation in Least Squares Regression. III. *Biometrika, 58*(1), 1-19.

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The Next Generation of the Penn World Table. *American Economic Review, 105*(10), 3150-3182.

Galton, F. (1888). Co-Relations and Their Measurement, Chiefly from Anthropometric Data. *Proceedings of the Royal Society of London, 45*, 135-145.

Godfrey, L. G. (1978). Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica, 46*(6), 1293-1301.

Greene, W. H. (2012). *Econometric analysis* (7th ed., International ed. ed.). Boston: Pearson.

Grossman, G. M., & Helpman, E. (1991). *Innovation and growth in the global economy*. Cambridge, Mass: MIT Press.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica, 46*(6), 1251-1271.

Heij, C., De Boer, P., Franses, P. H., Kloek, T., & Van Dijk, H. K. (2004). *Econometric methods with applications in business and economics*. Oxford: Oxford University Press.

Inada, K.-I. (1963). On a Two-Sector Model of Economic Growth: Comments and a Generalization. *The Review of Economic Studies, 30*(2), 119-127.

Islam, N. (2003). What have We Learnt from the Convergence Debate? *Journal of Economic Surveys, 17*(3), 309-362.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, 6*(3), 255-259.

Keynes, J. M. (1924). Alfred Marshall, 1842-1924. *The Economic Journal, 34*(135), 311-372.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*(2), 297-303.

Lorentzen, L., Hole, A., & Lindstrøm, T. L. (2010). *Kalkulus med én og flere variable* (4. Opplag ed.). Oslo: Universitetsforlaget.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics, 107*(2), 407-437.

Romer, D. (2012). *Advanced macroeconomics* (4th ed.). New York: McGraw-Hill/Irwin.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy, 98*(5), S71-S102.

Sala-i-Martin, X. (1996). The Classical Approach to Convergence Analysis. *The Economic Journal, 106*(437), 1019-1036.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics, 70*(1), 65-94.

Stock, J. H., & Watson, M. W. (2012). *Introduction to econometrics* (3rd ed., global ed. ed.). Boston, Mass: Pearson.

Verbeek, M. (2012). *A guide to modern econometrics* (4th ed.). Chichester: Wiley.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica, 48*(4), 817-838. doi:10.2307/1912934

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed. ed.). Cambridge, Mass: MIT Press.