INNOVATION AND ECONOMIC GROWTH IN NORWAY

BY

PIA KVAM SILFVENIUS

MENTOR:

LORAN CHOLLETE

HANDELSHØYSKOLEN VED UNIVERSITETET I STAVANGER

MASTEROPPGAVE, ØKONOMI OG ADMINISTRASJON, MØAHOV

DET SAMFUNNSVITENSKAPLIGE FAKULTET

Universitetet i Stavanger

DET SAMFUNNSVITENSKAPELIGE FAKULTET,

HANDELSHØGSKOLEN VED UIS

MASTEROPPGAVE

STUDIEPROGRAM: Master i økonomi og administrasjon med spesialisering innen anvendt finans	OPPGAVEN ER SKREVET INNEN FØLGENDE SPESIALISERINGSRETNING: Anvendt finans ER OPPGAVEN KONFIDENSIELL? (**NB**! Bruk rødt skjema ved konfidensiell oppgave)

TITTEL:

Innovasjon og økonomisk vekst i Norge

ENGELSK TITTEL:

Innovation and Economic Growth in Norway

FORFATTER(E) Pia Kvam Silfvenius		VEILEDER: Loran Chollete
Studentnummer: 895675	Navn: Pia Kvam Silfvenius	

OPPGAVEN ER MOTTATT I TO – 2 – INNBUNDNE EKSEMPLARER

Stavanger, ……/…… 2014 Underskrift administrasjon:……………………………

SENSURSKJEMA for masteroppgave
Handelshøgskolen ved UiS

Studieprogram:____

Spesialisering: ____

Opptaksår: ______

Er oppgaven konfidensiell? ☐ Nei ☐ Ja til: ______mnd _____ år (Maks 2 år)

Fylles ut av instituttet

Oppgaven innlevert i 2 eksemplarer:________(dato)

Signatur mottatt institutt________________________

Forfatter(e):

Studentnummer.	**Navn:**

Tittel

Norsk tittel

____

____

Engelsk tittel:

____

____

Oppgaven har fått karakter: _________________

bokstav

Sensor:

____

Sted/dato signatur

Veileder:

____

Sted/dato signatur

Sensur registrert i databasen:

____

Sted/dato signatur

SUMMARY

In this paper, I examine how investing in innovation affects economic growth in Norway. I attempt to conduct an empirical study using regression to examine the relationship between investing in innovation using R&D data and economic growth with GDP.

The analysis is based on time series data for the period 1970 to 2011, which is accumulated from two well known sources; Statistics Norway and the Norwegian Research Council. The results reveal that not all the assumptions for OLS are met; consequently the relationship between innovation and economic growth in Norway is not confirmed. Several explanations to these results are suggested, and even though the relationship was not empirically confirmed, one still believes in a positive relationship between innovation and economic growth.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

PREFACE

This thesis is written as part of the Master in Economics and Administration, with a main profile in Applied Finance at the University of Stavanger. The topic of the thesis is determined by my own desire and a strong interest in innovation and sustainable solutions.

Working on this thesis has at times been challenging but also very educational. The biggest challenges have been related to data collection and analysis where the acquirements of new knowledge in the field of statistics and econometrics have been necessary. Proving a relationship between innovation and economic growth has turned out to be much more difficult than expected. I have met a lot of hardship while writing this thesis; however, thanks to these problems I have become a better problem solver and learned to think in new directions. I hope the paper will be of interest to the reader.

Finally, I thank my mentor Loran Chollete for good advice and feedback.

Stavanger, 16.06.14

Pia Kvam Silfvenius

1.0 INTRODUCTION

1.1 MOTIVATION FOR CHOICE OF TOPIC

The idea behind this paper started with the question "How can money make the world more sustainable?" After reading Robert J. Shiller's book *"Finance and the good society"*, where he makes the case that finance is one of the most powerful tools we have in solving our common problems and increasing the general well-being. He offers financial innovation as one of the solutions (Shiller, 2012). Whether innovation leads to a good society is an empirical question, which is very difficult to answer. It's hard to find data to prove a good society. Therefore, I decided to look more into innovations and how they are influencing Norway's economic growth. I'm not saying that a good society is the same as a financially strong society, but it is interesting to see how much a motion towards a better society influence the economic growth. More precisely, does investing in innovation through research and development (R&D) improve the Norwegian gross domestic product (GDP)?

1.2 RESEARCH PURPOSE AND PROBLEM

Since the start of the second industrial revolution in the beginning of the 1870's, there has not been much doubt that innovation has played a significant role in the motivation of economic growth. However, the economic growth motivated by innovation can be difficult to repeat in recent decades. The value of innovations might have diminished compared with the past, since the era of elementary changes in the living standard may have gone (Wang, 2013).

Does investing in innovation have a positive effect on the economic growth in Norway?

1.3 RELEVANT RESEARCH WITHIN THE TOPIC

After the financial crisis began in 2007, many have expressed doubts about the goodness of the financial sector. These doubts are based on moral principles and traditions of a larger society (Shiller, 2013). In 2012 Robert J. Shiller released his book *"Finance and the good society"*, where he expresses the need to settle these doubts with financial practice. He explains the term good society as the kind of society in which we should aspire to live, it is a society in which all people respect and appreciate each other. At first glance, finance seems to be working against the achievement of such a good society (Shiller, 2012).

Shiller (2013) believes that we need to redesign finance to move towards a good society, to achieve this, a wide variety of factors need to be considered, both from theoretical finance and from psychology, history and culture. He also states that innovations (especially financial innovations) can and does contribute to the good society, and that innovations are important elements of the progress of our civilization.

Technological innovation is said to be one of the main sources of economic growth and development. There is an understanding that innovation is something one should invest in, and there aren't many who seriously doubt that innovators outperform non-innovators (Geroski, Machin, & Van Reenen, 1993). Recent theories on economic growth highlight technological change as the explanation of growth patterns in the economy. The pioneer behind endogenous growth models is Paul M. Romer, who said that technological innovation is created in the research and development (R&D) sectors using human capital and existing knowledge. These endogenous growth models notions that innovation facilitates sustainable economic growth, given that there are constant returns to innovation in terms of human capital engaged in the R&D sectors (Ulku, 2004).

A vast number of researchers have looked at the relationship between innovation and productivity, profitability and growth. These are studies mostly at firm level, but also in a bigger picture at country or cross-country level. However, these studies have met mixed results. This is not exactly essential here, but one can see that it may be difficult to establish a link between innovation and profits since there is a variety of factors that affect profits. (Cameron, 1996). Geroski et al. (1993) argue that that innovation has a positive effect on profit, but it is not possible to prove if it is greater than the cost of R&D.

Through the number of earlier research into the effect of innvation and growh there is a consensus that whether measured by R&D spending, patenting, or innovation counts, innovation has a significant effect on growth at the level of the firm, industry and country (Cameron, 1996). On the other hand, a study by Robert J. Gordon (2012) focuses on the concern that there has been a plunge in the value of inventions in the recent decades compared with the significant set of inventions during the second industrial revolution. He argues that new technologies often fail to improve people’s living standard in a cost effective way. Using the fact that the rate of life expectancy has declined since the 1950’s by two thirds compared with that of the earlier half century, he has support for his view. It is therefore reasonable to question whether there is still a positive relationship between innovation and economic growth.

1.4 STRUCTURE AND CONTENT

To enlighten the hypothesis, I will first in section 2 introduce a theoretical background on innovation and economic growth. Moreover I explain innovation and economic growth in depth and how they are functioning in Norway.

Section 3 is the design and methodology chapter where I describe the data collection and the variables I believe are essential for the research. Additionally I evaluate the data gathered. Further on the methods of regression, OLS and stationarity is explained.

In section 4 the results are presented and the analysis process is explained. In the fifth section the results from the analysis is discussed and compared to theory. The weaknesses of the analysis are also discussed. In section 6 the conclusion is presented.

2.0 THEORY

In this chapter I will first briefly introduce the theoretical background to innovation and growth. Further on I will talk about innovation, its influence on growth, and Norway's view on innovation. Then I'll introduce economic growth and gross domestic product, the weaknesses of gross domestic product as a measure of economic growth, and factors influencing economic growth.

2.1 INNOVATION AND ECONOMIC GROWTH: A THEORETICAL BACKGROUND

The good society is a human invention; made by laws, customs, procedures and organizations, to encourage all the complex patterns of human behavior into a congenial and effectual whole. The question is whether our society contributes to a system that is realistically better than alternatives in helping people to pursue their individual goals (Shiller, 2013).

It is said that past societies have changed much later than our society does today. Modern society has institutionalized investments in innovation that aims to "tear down to then build up”, i.e. to lay the foundations for sustainable economic activities in the future, even when the new products or processes are created at the expense of what already is established. The Egyptians built pyramids and Romans built aqueducts which show that humans have known for a long time that investing in innovations will ensure social and economic progress. Nevertheless, knowledge, technology and organization have over the last centuries constantly been developed and renewed, and the old discarded. This is how society has changed more during this time than in any previous historical period. Overall, the transformations has given substantial financial growth and as a result increased welfare (Ørstavik, 2001).

Researchers have ever since Adam Smith and Karl Marx pointed out the importance of exploiting new knowledge for economic growth. Schumpeter expanded the perspective when he argued that the foundation of the immense developments taking place in the modern economy is based upon the technological innovation through commercial exploitation of new knowledge and scientific analytical methods (Ørstavik, 2001).

2.2 INNOVATION

"Innovation is an application of knowledge to produce new knowledge."

(Cho & Pucik, 2005, p. 556)

2.2.1 DEFINITION

Innovations have often blurred contours and confusing course: what innovation is, who are the innovators, and who are the significant others, may provide both confusion and disagreement. In the reality, innovations can be transformed over time and the results in the end can be quite different from where it started. The term innovation is here explained as introduction of new or significantly improved products or processes (Ørstavik, 2001).

2.2.2 INNOVATION AND ITS UNCERTAINTIES

We generally say that innovation is important, but it is not easy to say anything more specific about when innovation pays off and how investing in innovation should be done in different situations. This is an important issue, especially because there is great risk associated with innovation: Innovation can often fail. Perhaps one can't make what was planned, it might be more expensive, and maybe it takes a lot longer than anticipated. An innovation can also lead to increased costs; result in counter-reactions from rivals, impact on established groups of cooperation in an industry - or in business internally (Ørstavik, 2001).

The most important thing with innovations is that it always will be about learning. However, this does not mean that science is the fundamental driving force to all economic development. Investment in scientific activities and the development of academic institutions doesn't directly and inevitably contribute to economic growth and development. For example, some academic and scientific researches are not aimed towards economic activities and new economic effects at all. Also, research that claims to be economically useful can be driven ineffective and give poor quality results. (Ørstavik, 2001).

Innovation in itself comes with a lot of uncertainty, and there is no way to avoid this uncertainty. Innovation will partly always entail seeking new paths in unknown terrain.

However, even if few reach the “goal” it is worth that many tries, because those who in the end succeed can have a great economic and social importance (Ørstavik, 2001).

2.2.3 INNOVATION IN NORWAY

The Norwegian Government's goal is to strengthen innovation capacity so that it helps maximize overall growth in the Norwegian economy; an economic growth that is sustainable so it can help achieve their welfare objectives. Innovation and creation are increasingly important roles in the various Norwegian sectors. The solid growth in the Norwegian business sector in recent years has been the result of local and regional adaptability. Innovation in the business sector has been a major driving force for development of robust business communities across the country (Nærings- og Handelsdepartement, 2008). This gives new challenges, but also new opportunities. Through an active innovation policy a company can exploit these opportunities, and develop what they are good at. That is why the Norwegian Government believes that good innovation policy is good economic policy (Nærings- og Fiskeridepartementet, 2014).

One of the problems with innovation in Norway, and probably the rest of the world, is the lack of courage to invest in something new and unfamiliar, for example new financial instruments are attractive only if they can be bought and sold easily; they have to be adopted widely before people want to adopt them widely. Shiller (2012) suggests that the solution to this problem is with government-supported tax incentives.

Facilitating innovation runs like a red thread through the Norwegian Government policies. The Government has a key role when it comes to adding framework that makes it possible for companies to innovate (Nærings- og Handelsdepartement, 2008). An important tool for particularly stimulating increased R&D investment is the system of tax credits for R&D projects, which is similar to Shiller’s solution. SkatteFUNN is a tax credit system where all Norwegian companies with research and/or development projects, or planning to start such a projects may apply for approval so that the company can use it rights to tax deduction (Forskningsrådet, 2013).

2.3 MODERN ECONOMIC GROWTH

Modern economic growth is an increase over a longer period in domestic product per capita. Economic growth is not the same as economic development. However, there is little evidence that a country, for example a developing country, can have an economic development without economic growth. There is therefore a reason to emphasize that a country's gross production always will be a central part of the description of economic growth and development (Munthe, 1992).

What the Government first and foremost wants to achieve by facilitating innovation, is to contribute to increase welfare through added value. Today Norway scores high on various measures of added value and living standards. When you look at the gross domestic product, Norway has gone from being a county among the average of the OECD countries, to a country that each year is at the top among the world's richest countries (Nærings- og Handelsdepartement, 2008). The Norwegian GDP might be 20 times bigger today than in the 1900, and over 30 times larger than in 1865. It is obvious that it has something to do with the increase in production capacity over the last 130 years. If we look at the long term change in GDP in Norway, we find some characteristics: (1) the long-term trend is rising; (2) there is acceleration in the growth until the mid-1970; (3) the yearly change in domestic product is generally positive; (4) the rise from year to year is not even (Munthe, 1992).

2.3.1 GROSS DOMESTIC PRODUCT

Gross Domestic Product (GDP) is a measure wildly used to calculate an economy's performance and growth. It is a measure of the total economic activity in a country. It gives us the total monetary value of all final goods and services produced within the country's borders during a specific time period, usually set to a year. GDP represents also the earned income of those who contribute to the production in the country (Steigum, 2004).

In order to compare GDP from year to year, we need to determine how much of the change is due to changes in the price level (nominal change) and how much of the change that actually comes from a change in the number of goods and services produced (real change). By adjusting GDP for inflation and deflation we find real GDP, which is GDP in constant prices.

The growth in real GDP is what we are interested in, and when referring to GDP it is on the real GDP in question (Steigum, 2004).

2.3.2 WEAKNESSES WITH GDP

When it comes to measuring the total production and total economic welfare, GDP has some weaknesses. Firstly economic welfare is more related to consumption than production. The population of a country which exports a large part of production to build up foreign assets has low consumption and probably low economic welfare. That is why household consumption or income would be a better measure of welfare. Furthermore, the GDP contains a number of products that only helps to increase welfare because they help fix damages or failures that has occurred. Cleanup after environmental disasters are a commonly used example - the accident reduces welfare, clearing increases both GDP and welfare, but only because an accident has happened. Moreover there are goods and services that contribute to the welfare that are not included; these are activities such as taking care of one's own children, illegal- and black market activities, and unpaid volunteer work. In addition, neither GDP nor GDP per capita, say anything about how the income is actually distributed. It says nothing about health or education, only about how many resources it takes to produce services in these sectors. Last but not least - the value of the services that nature provides us does not generate income so it is not included, or the cost of the use of natural means. As a result, GDP undervalue the country's total production (Miljøverndepartementet, 2013).

2.3.3 FACTORS INFLUENCING ECONOMIC GROWTH

I will now explain some aspects that influence growth. The actual growth in a period of time results from a combination of many factors, both financial and non-financial. I will not describe all of the factors here, but instead discuss a few significant ones for this paper.

In studies on long term economic growth processes, it is natural to focus on production capacity in the economy. The capacity to produce goods and services is the limiting factor for national income. Particularly, it is the access to labor and physical capital that will be essential for the production capacity. An increasing access to labor and physical capital will therefore be fundamental for a higher GDP over time (Steigum, 2004).

According to Statistics Norway (2012) production is the value of goods and services from domestic production activities, i.e. market-oriented, production for own use and non-market operations in government and nonprofit organizations. Production of goods and services is not the same as the sale of goods and services. Production published in base value, i.e. subsidies on products is included, but not VAT or other taxes on products.

A country's private consumption includes all expenditure of households in a country. This is in connection with the purchase of consumer durables, semi-durables, non-durables and services. Private consumption or consume in households is known as the final delivery in the national accounts (Steigum, 2004). Durable consumer goods are goods that can be used repeatedly or continuously over a period of one year or longer, and include, among other appliances, furniture and vehicles. Semi-permanent consumer goods are goods like clothes and utensils, while non-durable consumer goods including food, beverages, etc. Expenses for services may include medical expenses, hairdressing and similar (Statistisk Sentralbyrå, 2012).

3.0 DESIGN AND METHODOLOGY

3.1 DATA COLLECTION

The purpose of this thesis is as previously explained to study the relationship between innovation and economic growth, and to see if I can prove that there is, in fact, still a positive relationship. To test this, I performed a regression analysis using STATA based on a sample of annual observations for the period 1970 to 2011. The data consist of gross R&D expenditure, R&D employment and other macroeconomic data. The material data are collected from two sources: Statistics Norway and the Norwegian Research Council.

3.1.1 DIFFICULTIES WITH MEASURING INNOVATION

The problems with confirming results of innovation may be a theoretical and conceptual problem. Basic mental images of innovation processes and innovations give us very poor tools for assessing successes and failures in innovation context. For example, the simple notion that businesses are discrete, permanent, unambiguous and rational participants is very often too simple: companies do not reflect such simple assumptions because innovation statistics and innovation analysis can be flawed. It is also difficult to specify and refine what is an innovation process and what is not. There are difficulties with following the innovation processes over time, and it can be difficult to find good indicators of the effects of them. The basic innovation model where one assumes that innovation is the product of a clear process in which an idea is transformed into a new product through research and development, are very often inadequate and misleading (Ørstavik, 2001).

EU ranks innovation activity in the member countries each year, with the so-called Community Innovation Survey (CIS) as the main source. The Norwegian innovation survey is compiled using the guidelines of the CIS, and included in the basis for Norway in these rankings (Nærings- og Handelsdepartement, 2008). However, since the Norwegian innovation survey hasn't been conducted for that long (since 1992), it does not provide enough data to do a yearly analysis. R&D data or patent statistics is therefore needed as innovation proxies (Nås & Leppälahti, 1997).

3.1.2 RESEARCH AND DEVELOPMENT DATA OR PATENT STATISTICS

Measuring innovation activity at a national level is generally believed to be complex. Since there is not a flawless innovation measure, a reliable indicator of innovation activity is needed. Research and development data and patent statistics are widely used in economic studies as innovation proxies, however both with support and criticism (Wang, 2013).

Research and Development data, either R&D expenditure or R&D-related employment, are the most commonly used innovation proxies. However, R&D data have several weaknesses (Wang, 2013). An important empirical objection is that R&D activity is a precondition for innovation. It is in fact a good deal of firms that have innovation activities, but do not perform

R&D. To assume that the path to innovation goes through research can potentially provide an error in the analysis (Cappelen, Raknerud, & Rybalka, 2007). That R&D can barely be considered as an exogenous variable is another problem related with estimating how much R&D affects economic growth. The amount invested in R&D often depends on the expected sales level. This makes knowing which direction the casual link is working a complicated task (Svensson, 2008).

While R&D measures innovation input, patent statistics provide innovation output measures. The benefit with using patents as innovation indicator is that patents represent successful innovations. Patent statistics have had a wide coverage in economics literature; still there are some potential issues when using patents as an innovation measure. Firstly, they are restricted by patent legislation, so only some types of inventions form a limited number of sectors can be patented. This leads to patents applications that are concentrated to the manufacturing and extractive industries. Furthermore, since patenting involves revealing an invention's technical details, many firms prefer secrecy over patenting. Because of the cost involved in patenting, patenting is unfeasible for small firms, which results in patent data being less representative in various firm sizes. Finally, patents represent inventions, and it is not certain that those inventions become innovations. Some patents are only used to prevent others from doing so as a purely anti-competitive strategy (Wang, 2013).

Even though patent data provide unique information for the analysis of technical change, the feasible data collected was not usable. My contacts at The Norwegian Patent and Trademark Office had some challenges with collecting cases before 1976, since the older data is not digitized as newer material. From Statistical Norway and the Norwegian Research Council's (NIFU) online database I collected R&D data back to the 1970's. Even if this data has some flaws, I decided to use R&D expenditure and R&D employment as innovation proxy in this analysis.

3.1.3 R&D AND SPILLOVER

Unless a company uses patenting, they may find it hard to prevent other companies from using the new knowledge they get from investing in R&D. Knowledge becomes “a public good”. It is also doubtful that a company will by themselves will be capable to utilize all the knowledge generated by the R&D. This explains how R&D can lead to spillovers to other companies (Svensson, 2008).

At an aggregate level, R&D investments, together with the production factors, are the aspects that determine economic growth. It can be difficult to demonstrate that there really are spillover effects even if a link is found between economic growth and external R&D, as these effects are always indirect. Earlier research differ greatly in terms of the aggregated level (company, industry or nation), model specification, data sources (countries, periods of time), and how key variables are calculated. It is however important to note that the indirect spillover effects take longer to act than the direct effects of a company’s own R&D (private return) (Svensson, 2008).

When estimating how R&D affects growth or productivity at the aggregated level it is may be essential to take spillover effects from other countries into account. Earlier studies at aggregated national level have shown that the R&D conducted in other countries can be more significant than the R&D conducted within the country for the growth of productivity in the country concerned. Researchers have also found that productivity in small countries is affected to a greater extent by the R&D carried out in other countries than productivity in large countries (Svensson, 2008). In this study only data from Norway is used.

3.1.4 VARIABLES

In order for the regression to give the highest possible explanation level it is necessary to include factors other than just R&D data that affects the economic growth. I have earlier in this study had a general review of important variables affecting economic growth, according to theory. Below I briefly describe the numbers and indices compiled and from which database the data is retrieved.

3.1.4.1 GDP

I have chosen to use an annual index of total gross GDP in Norway as the measure of economic growth. The observations since 1970 (inclusive) are available from Statistical Norway's online database and are measured in NOK. GDP series in Norway consistently follows a rather similar and upward linear trend, and growth is relatively stable.

3.1.4.2 R&D DATA

The R&D data is, as stated above, collected by Statistics Norway and Norwegian Research Council (NIFU) and aims to measure the R&D activity in three different sectors; Institute sector, Universities and college sector, and the business sector. For the Norwegian business sector the main data is collected by printed questionnaires. Additional information from the Central Register of Establishments and Enterprises is used. Enterprise websites and annual reports are also applied (Longva & Blekstad, 2004). Influenced from earlier research I chose to have two R&D variables; R&D expenditure and R&D- related employment (from now on known as R&D staff). Both variables are the total annual numbers, and not divided into sectors.

3.1.4.3 PRODUCTION AND VALUE

As mention earlier, when studying long term economic growth processes it is expected to focus on production in the economy since the capacity to produce goods and services in a country has a huge influence on its national income. The production and value data is also collected from Statistics Norway's online database, measured annually in NOK.

3.1.4.4 CONSUME IN HOUSEHOLDS AND NON-PROFIT ORGANIZATIONS

Consume in households are known as the final distribution in the national accounts and has a great effect on the GDP. As the other macroeconomic data, consume in households and non-profit organizations are collected from Statistical Norway's online database and measured annually in NOK.

3.2 EVALUATION OF DATA

3.2.1 MISSING DATA

Missing data or missing values arise in a variety of forms; it is a common occurrence and may have a significant effect on the conclusions drawn from the data (Wooldridge, 2009). This is a problem that occurred when collecting data for this thesis. The R&D survey where only conducted every other year. Consequently, there is only R&D statistics for each other year until 2001, with a gap between 1974 to1977.

Missing data creates difficulties in scientific research because most data analysis procedures where not designed for them. The data collected in this thesis, with its missing values, makes it difficult to run a standard multiple regression analysis. Missingness is an irritation, but managing it in a principled way raises theoretical difficulties and computational challenges. However, the lack of resources or even theoretical framework, have made earlier researchers, methodologists, and software developers resort to editing the data to lend an appearance of completeness (Schafer & Graham, 2002).

After discussing this with the Norwegian Research Council, the method of averaging where conducted for the years with missing values. This was computed by using the formula below.

(1) (Count for Year One + Count for Year Two) / 2

By averaging the missing data I gain annually data, thus a regression analysis can be carried out. Unfortunately, edits of data may do more harm than good, producing results that are biased, inefficient (lacking in power), and unreliable (Schafer & Graham, 2002).

3.2.1.1 PROBLEMS WITH SCIENTIFIC RESEARCH

Academic scientists acknowledge that they often get things wrong. However, they believe that these errors will get corrected over time when other scientists try to take the work further. There are in fact more scientific papers with errors being published than anyone would expect, or like to think (The Economist, 2013a).

There are rarely done replications when research has gone wrong, mainly because it is hard and thankless work. Most academic researchers would rather spend time on work that is more likely to enhance their careers (The Economist, 2013a). This is because only the most striking findings make it into the leading journals. Failures to prove a hypothesis is rarely even offered for publication or accepted. However, knowing that something is false can be just as important as knowing something is true. The failure to report deficiencies means that researchers waste time and money on exploring dead ends already explored by other scientists (The Economist, 2013b).

I can't find any previous research that has used the same data as I have collected. The data assembled for this thesis is just the summarized statistics from Statistical Norway and the Norwegian Research Council online databases, which they have gathered from surveys. This means that earlier research on R&D and economic growth in Norway is done with much more advanced data, than what I have access to.

3.2.2 RELIABILITY

My main concern with the collected data is the stability of the R&D statistics. The R&D surveys towards the business sector have been conducted each other year since 1963 to 2001. From 2001 there is statistics from every year. The Norwegian R&D survey has gradually been extended since the beginning in 1963. From 1970 the surveys were carried out in a more systematic way and the statistics have been extended gradually. The first survey covered only the manufacturing industries, but the service industries were included gradually and have been well covered from 1995 onwards. This means that long time series are only available for the manufacturing industries. The time series are also affected by the methodological change in the survey from 1995 (Longva & Blekstad, 2004).

Finding data to measure Norway's innovation development is not an easy task. Despite the fact that the surveys and R&D data vary in extent from year to year, I believe this data will be a good representation for the innovation development in Norway.

3.2 REGRESSION ANALYSIS

In the analysis I use a multiple regression analysis. It explains the relationship between a dependent variable and several explanatory variables. For the dependent variable the notation Y is used and for the independent variable the notation X is used. Y can be expressed as a linear function of X with k explanatory variables as follows:

(2) $Y_t = \beta_0 + \beta_1 x_{1,\,t} + \beta_2 x_{2,\,t} + ... + \beta_k\,,\, x_k,\, t + \varepsilon_t$

β_0 = the intercept

Y_t = dependent variable at time t

X_t = explanatory variable at time t

β_t = the explanatory variable constant at time t

ε_t = the model error term at time t

Regression coefficient β indicates how much Y changes when X changes by one unit. In a simple regression model with only one independent variable coefficient can be seen as the slope of the regression line. β_0 is the intercept of the regression line, and informs what size of Y is at zero observations of the independent variable X (Wooldridge, 2009).

3.2.1 STATISTICAL SIGNIFICANCE

Statistical significance indicates that the observed effect of the variables in the analysis is plausibly due to chance. A statistical analysis is described as statistically significant if it is unlikely that the result occurred randomly. When we decide whether a hypothesis should be rejected or not, we must choose a level of how much rejection we are willing to accept. It is recommended to use a 5% significance level. If the null hypothesis is correct, we then accept a 5% chance of making a rejection error (Studenmund, 2006).

3.2.2 COEFFICIENT OF DETERMINATION R^2

A regression analysis is described with a variable called coefficient of determination, R^2. This variable explains how much variation in Y can be explained by X. This coefficient is appropriate to look at when you want to assess how well the model fits the observations. The problem with R^2 is that if it includes several independent variables it never decreases. This is because a variable cannot explain less than 0 % of the variation in the dependent variable. This problem can be avoided by using the adjusted R^2, which takes into account the degrees of freedom when adding more explanatory variables in the regression equation. Adjusted R^2 should be located as close to the R^2 as possible. This indicates that all the explanatory variables help to explain the dependent variable. If there is a big difference in adjusted R^2 and R^2 then one or more independent variables do not have explanatory power (Studenmund, 2006).

3.2.3 TIME SERIES DATA

Since my collected data consists of variables that are observed over time with a constant interval between each observation; the variables I will analyze are in the time series form and I can use regression "ordinary least squares" (OLS) (Wooldridge, 2009).

3.3 SELECTING NUMBER OF LAG

In many cases there is a possibility that time might elapse between a change in the independent variable and the resulting change in the dependent variable. A distributed lagged model explains the current value of Y as a function of the current and/or past values of X. There are several methods to determine the number of layers that are optimal to include for the different variables. General-to-Specific method starts to do the regression with the highest number of layers you think will make sense. If the result is not statistically significant eliminates one layer, and so it continues until the number of lag to income is significant. Using different information criterions to decide the correct number of lags to include in the model is another option. In this analysis Akaike's Information criterion (AIC), Schwartz's Baysian information criterion (SBIC) and Hannan Quinn criterion (HQIC) are included.

(3) $AIC = \ln\left(\widehat{\delta^2}\right) + \frac{2k}{T}$

(4) $SBIC = \ln\left(\widehat{\delta^2}\right) + \frac{k}{T}\ln T$

(5) $HQIC = \ln\left(\widehat{\delta^2}\right) + \frac{2k}{T}\ln\left(\ln(T)\right)$

$\widehat{\delta^2}$ is the variance of the residual, T is the number of observations and k = p + q + 1 is the total number of estimated parameters.

These methods weights the "Residual sum of squares" (RSS) to the degrees of freedom. Including an extra lag it will have two conflicting effects on the information criterion; RSS will fall as the extra lag will increase. Therefore, it is profitable to minimize the information criterion. Including an extra lag will only diminish the information criterion if the decrease in RSS is larger than the increase in the degrees of freedom. AIC punishes the number of degrees the least, then comes HQIC and in the end SBIC. As a result, the different information criterions may give different outcomes (Solbakken, 2011).

3.4 ORDINARY LEAST SQUARES

The observations in the data collected will never be exactly on a straight line, so a linear regression model will always be an estimation of reality. Since the accurate number of α and β are unknown, the estimates are made to form a straight line. This straight line will never precisely match the real regression so an error term, ε_t, is added. The difference between the real and estimated regression line are called residuals. If the estimated regression line has small residuals will be described as good. OLS tries to find the best estimated regression line that minimizes the sum of squared residuals. Mathematically minimization is expressed as follows:

(6) $Min \sum_{i=1}^{n} u_i^2 = Min \sum_{i=1}^{n} (\widehat{Y_i} - Y_i)^2 = Min \sum_{i=1}^{n} (y_i - \widehat{\beta_0} + \widehat{\beta_1}\ x_{i1} + \ldots + \widehat{\beta_k}\, x_{ik})$

Which gives k + 1 unknown $\beta_0, \beta_1, \ldots, \beta_k$

(Solbakken, 2011)

3.4.1 ASSUMPTIONS OF OLS

There are six assumptions that must be met for the OLS to be reliable. These assumptions are called the Gauss-Markov assumptions. When all conditions are met, the results of the regression are called BLUE, "Best Linear Unbiased Estimate" (Studenmund, 2006). I will briefly introduce these assumptions, the consequences of violation on these will have, and suggestions to how you can solve possible violations.

3.4.1.1 LINEARITY

The time series process should follow a model that is linear in the parameters. If the parameters are not linear, it means that you either have included regressors that you shouldn't, missing some important regressors or have unstable parameters. By looking at regression equation before making the actual analysis you can confirm whether the assumption of linearity is fulfilled (Wooldridge, 2009; Solbakken, 2011).

3.4.1.2 AVERAGE RESIDUALS HAVE EXPECTATION EQUAL $0, E(\varepsilon_t) = 0$

Factors that are not included in the model will not interfere with the dependent variable (Wooldridge, 2009; Solbakken, 2011). There is no need to explain this assumption any deeper, since it should not be an issue I this analysis.

3.4.1.3 NORMALLY DISTRIBUTED RESIDUALS

An important requirement for the standard errors and test values to provide proper interference in the analysis is that the residuals are normally distributed. There are several ways to test if the data set has a normally distributed error term. Bera-Jarque test is one of the most common tests for normality. It checks the distribution of skewness, which measures whether the distribution is symmetrical about the mean, and kurtosis, which measures how thick the tails of the distribution are (Solbakken, 2011).

3.4.1.4 NO AUTOCORRELATION FOR THE RESIDUALS

In time series analysis, autocorrelation or serial correlation is a common problem. It occurs when the errors associated with a given time period carry over into future time periods. If there is autocorrelation in the data, the estimated coefficients are no longer BLUE (Best Linear Unbiased Estimator) and the variance and standard error are no longer valid. More specific, the X_t increase, while the standard error will be underestimates of true values. This indicates that R^2 will be overestimated, and the t-statistics will look like they are more significant than they are. Hence, the consequences of ignoring autocorrelation are the same as those of ignoring heteroscedasticity, the OLS estimates and forecasts can still be unbiased and consistent, but inefficient. A solution to this problem, if not already done, is lagging the variables (Studenmund, 2006).

There are different ways to test whether the data series contains autocorrelation. However, a Durbin-Watson test cannot be used if the variables are lagged. Breusch-Godfrey test, on the other hand, is a test that takes into account any correlation between the explanatory variables and the lagged residual. This test also takes into account heteroscedasticity and serial correlation of higher order (Solbakken, 2011).

3.4.1.5 NO PERFECT MULTICOLLINEARITY

If two or more independent variables have high (put not perfect) correlation, then multicollinearity occurs. No perfect multicollinearity means that the coefficient of the independent variables do not change even if you add or remove a variable. A typical symptom of multicollinearity is that the t-values are not significant, while the F-test for the regression is significant and explanation level high (Wooldridge, 2009; Solbakken, 2011). One way to detect multicollinearity is to look at the correlation matrix between variables. You can expect the variables with the highest correlation are the variables that will cause problems with multicollinearity. If the assumption of no perfect multicollinearity is not met, then OLS will be unable to estimate the individual explanatory variables effect on the dependent variable (Studenmund, 2006; Solbakken, 2011).

3.4.1.6 HOMOSCEDATICITY

When the variances of the residuals are constant over time and independent of the explanatory variables, they are homoscedastic. If the variance of the residuals will be equal, indicating that the variance may change from observation to observation, then they are heteroscedastic (Wooldridge, 2009).

Breusch-Pagan test is one way to test for heteroscedasticity. This test examines whether the estimated residuals variances depends on the values of the independent variables. Where the null hypothesis is that the residuals have constant variance. The alternative hypothesis is then that the variables do not have constant variance. Heteroscedasticity is a problem if H_0 is rejected at either 5 % or 10 % significance level (Solbakken, 2011).

If the data are heteroscedastic, then it can be solved by adopting a so-called "weighted least square" regression. Observations with high residuals are either ignored or weighted so that they are less important. However, this method is best suited if there are a large number of observations. So another way to solve the problem is to use natural logarithms of the variables to reduce extreme observations (Wooldridge, 2009; Solbakken, 2011).

3.5 STATIONARITY

The difference between a stationary time series and a non-stationary time series is that stationary series has basic properties for example its mean and their variances do not change over time. Officially, a time series variable, X_t, is stationary if:

1. the mean of X_t is constant over time
2. the variance of X_t is constant over time
3. the simple correlation coefficient between X_t and X_{t-k} depends on the length of the lagged (k), but not on any of the other variables (for all k)

If one of more of these statements is not met, then X_t is non-stationary. In a non-stationary time series the relationship between Y and X will be behaving as though it were a “random walk”, where it won’t be possible to see how the independent variables affect the dependent variable. A random walk variable is non-stationary because it can wander up and down without an inherent symmetry and without approaching a long-term mean of any sort (Studenmund, 2006). Differentiation is one way to handle this problem. A time series variable that is differentiated *d* times to become stationary are defined as integrated of the order *d: I (d)* (Wooldridge, 2009; Solbakken, 2011).

Structural break or seasonal variation can also be reasons for non-stationarity. A structural break implies that the population function changes over the sample period so that the equilibrium value is affected. Unreliable seasonal patterns over time are what cause season variation. With this kind of developments one should differentiate seasons to achieve stationary time series (Solbakken, 2011).

A particular type of non-stationary time series that often occurs in financial data is unit root. The biggest consequence with unit root for regression analysis is that the regression results can be misleading and erroneous. This is called the spurious regression problem. A regression with variables that have spurious correlation will get statistical significant results, however this reflect a common trend and not an underling context. The significance of the estimated coefficients is then spurious, or invalid (Studenmund, 2006).

3.5.1 DICKEY-FULLER TEST

Testing for non-stationarity is important so we are sure that the equations we are estimating are not spurious. The base for a stationary analysis is the autoregressive model:

$$(7)\ y_t = \alpha + py_{t-1} + e_t$$

Where, t = 1, 2 …

If H_0: p = 1 then the Y is "unit root" and the time series is non-stationary, and if H_1: |p| < 1 the Y will be stationary. When using the Dickey-Fuller test it is important to know that the t-statistics don't have a normal distribution since y_{t-1} is I(1). This means that the standard t-distribution don't represent a reliable critical value for the Dickey-Fuller test. Many variables are autoregressive of a higher order than 1. If this is the case, one must use Adjusted Dickey-Fuller test. This test contains more lags to detect serial correlation in the variable. However, it is important not to include too many lags since one looses the degrees of freedom in the regression (Wooldridge, 2009; Solbakken, 2011).

$$(8)\ \Delta Y = \alpha + \theta y_{t-1} + \sum_{i-1}^{p} y_i \Delta Y_{t-1} + \varepsilon_t$$

Were, $\theta = (p - 1)$

3.5.2 LOW STRENGTH FOR DICKEY-FULLER TEST

The Dickey-Fuller test has low strength; this means it can find unity in the time series data even if this is not the case in reality. Therefore, one has to be careful with the interpretation of these results since there is a high likelihood of making a conclusion with the wrong result. To ensure a correct result one can for example use another unit root test. Philip-Perron test is an example of such a test. This one uses non-parametrical method to account for autocorrelation (Solbakken, 2011).

4.0 RESULTS

In this part of the paper I will present the results of the regression analysis I have done in STATA. I have completed a regression analysis where the effect on how the total R&D investments and R&D staff in Norway influences the Norwegian GDP.

4.1 CHOICE OF VARIABLES

As mentioned in the section data collection I decided that it was most expedient to start with four explanatory variables that I believe from theory have an influence on the dependent variable. The regression analysis starts with this model:

Model for R&D’s influence on GDP:

$$(9)\ GDP_t = RD\_expenditure_t + RD\_staff_t + Production\_Value_t + Consume\_Households_t$$

I started with a regression with unprocessed data to get an impression of the variables. The first regression gives an explanatory degree (R^2) of 0,999 which is extremely high, and the variable RD_staff has a non-significant p-value. Further on we will see if the model can be improved and become more robust.

4.2 TIME DEPENDENT VARIABLES

As explained above, a stationary time series stand out from other times series with the fact that it has a stable probability distribution over time. This could explain the high explanatory degree in the regression (R^2 = 0,999). We can easily get an overview with two way graphs:

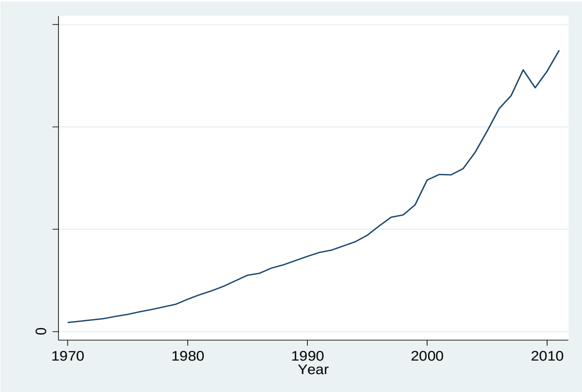


FIGURE 1: THE DEVELOPMENT OF GDP

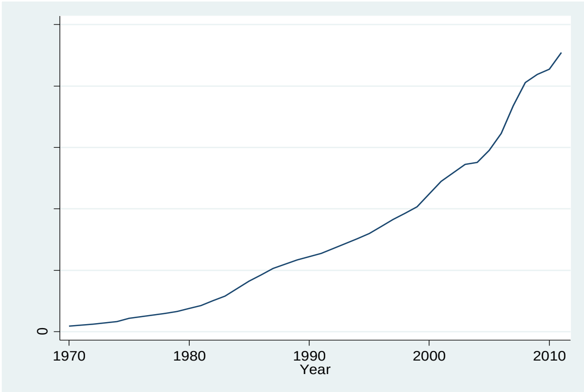


FIGURE 2: THE DEVELOPMENT OF R&D EXPENDITURE

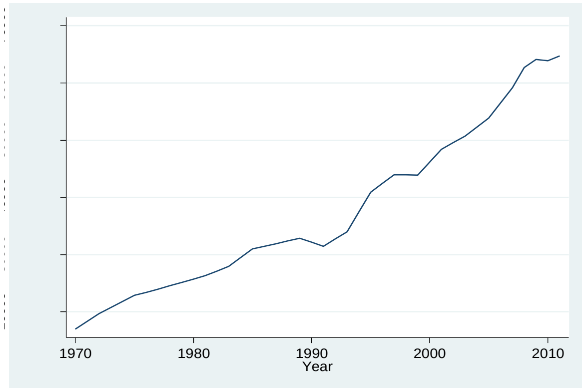


FIGURE 3: THE DEVELOPMENT OF R&D STAFF

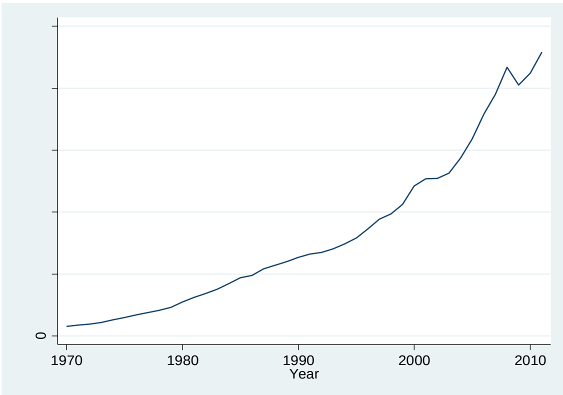


FIGURE 4: PRODUCTION AND VALUE DEVELOPMENT

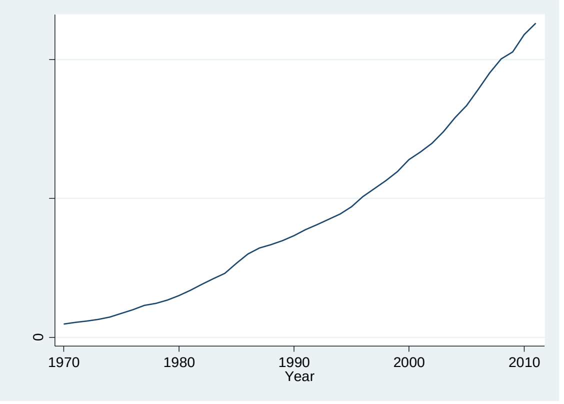


FIGURE 5: THE DEVELOPMENT OF CONSUME IN HOUSEHOLDS

The graphs reveal that the variables are non-stationary. When the variables are non-stationary, there is no point continuing the analysis since it indicates that we can't trust the results. However, there is different tactics to changing these results. I first try to convert the data to the natural logarithms so the extreme values will be modified and decreasing the difference. Unfortunately, this doesn't improve the data that much so I try differentiating the observations. I have to differentiate three times to get, what looks like, a fairly stationary result. I still can see some trace of trend, but I decide to continue the analysis to see what outcome I get.

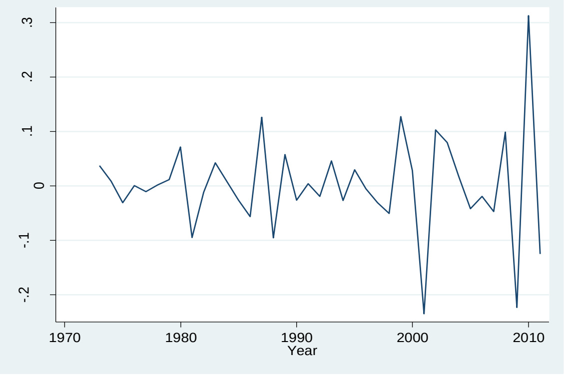


FIGURE 6: GDP AFTER BEING LOGGED AND DIFFERENCIATED THREE TIMES

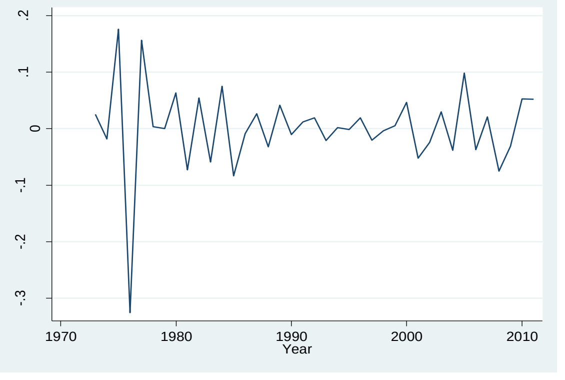


FIGURE 7: R&D EXPENDITURE AFTER BEING LOGGED AND DIFFRENCIATED THREE TIMES

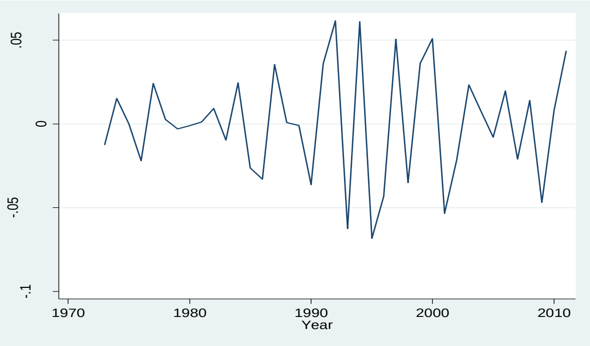


FIGURE 8: R&D STAFF AFTER BEING LOGGED AND DIFFERENCIATED THREE TIMES

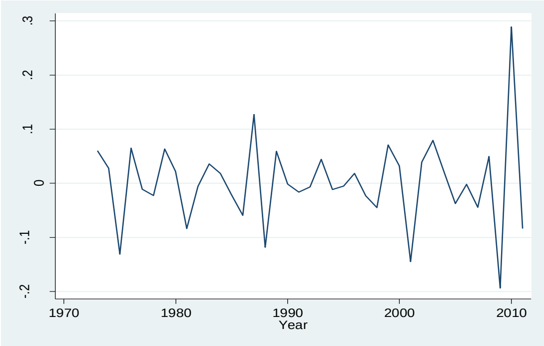


FIGURE 9: PRODUCTION AND VALUE AFTER BEING LOGGED AND DIFFERENCIATED THREE TIMES

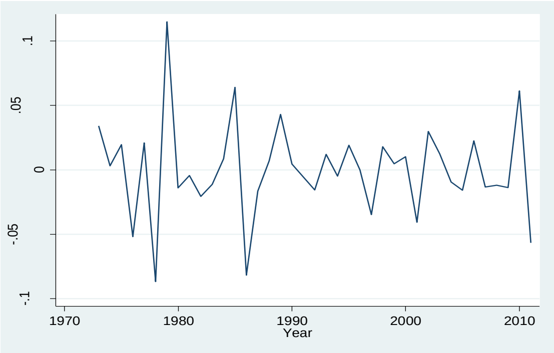


FIGURE 10: CONSUME IN HOUSEHOLDS AFTER BEING LOGGED AND DIFFERENCIATED THREE TIMES

4.3 NUMBER OF LAG

After manipulating the data, I still get a pretty high R^2 and the p-value for RD_staff and Consume_Household are too high. A reason for this might be that is takes some time before R&D expenditure, R&S Staff, Production and value and consume in households affect the GDP. That is why one of the first things I do is test the number of lags that is optimal to include in the analysis.

As mentioned in the method chapter, the following information criterions were used to decide the number of lags: Akaike's information criteria (AIC), Schwart's Baysian information criteria (SBIC) and Hannan Quinn criteria (HQIC). The information criterions are not always unanimous, as seen in the table:

Number of lag recommended			
Variable	AIC	HQIC	SBIC
R&D Expenditure	4	4	4
R&D Staff	4	4	3
Production and value	5	5	3
Consume in Households	5	5	5

TABLE 1: NUMBER OF LAG OF THE VARIABLE DESIDED BY THE IC

To reach the best model I used the information criteria as a starting point before I conducted many regressions with different number of lags of the different variables to test what combinations gave the best result. I also took the variables individually to test how they influenced the GDP with different lags, and with what lag the variables got the lowest p-value. The final result is illustrated in table 2:

Variable	**Number of lags**
R&D Expenditure	4
R&D Staff	1
Production and Value	1
Consume in Households	4

TABLE 2: NUMBER OF LAG TO USE IN THE REGRESSION

R^2 decreased to 0.52, but the p-values significantly improved. To see if the model could be improved further it necessary to test the assumptions for OLS.

4.4 ASSUMPTIONS FOR OLS

4.4.1 MULTICOLLINEARITY

To test this condition I used a correlation matrix that tests the correlation between all the different explanatory variables in the regression.

Variable	R&D Expenditure Lag 4	R&D Staff Lag 1	Productivity and Value Lag 1	Consume in Households Lag 4
R&D Expenditure	1			
R&D Staff	0,0845	1		
Productivity and Value	0,1193	0,2838	1	
Consume in Households	0,2025	0,0063	-0,1915	1

TABLE 3: CORRELATION MATRIX

We can see that none of the variables are considerably correlated. This means that all the variables should stay in the model.

4.4.2 HOMOSCEDASTICITY

For the results of the OLS to be robust, this assumption says that the residuals must have a constant variance. To test for heteroscedasticity in the data I chose to use a Breusch-Pagan test.

H_0 = Data is homoscedastic

H_1 = Data is heteroscedastic

The table shows us the results of the test:

Test:	Chi-2	P-value
Breusch-Pagan	0,00	0,9782

TABLE 4: RESULTS FROM THE BREUSCH-PAGAN TEST

As we see in the table, the test shows us a high p-value of 0.9782 which tells us that the null hypothesis, the data is homoscedastic, cannot be rejected.

4.4.3 NORMALLY DISTRIBUTED RESIDUALS

To test the assumption of normal distributed residuals I used Bera-Jarque test to see if "skewness" and "excess kurtosis" simultaneously is zero.

H_0 = the residuals are normally distributed

H_1 = the residuals are not normally distributed

Variable	Pr (skewness)	Pr (kurtosis)	Adj. Chi2 (2)	Prob > Chi2
Res (residuals)	0,3841	0,7810	0,88	0,6447

TABLE 5: RESULTS FROM THE BERA-JARQUE TEST

As we can see in the table, the p-value is higher than 0,05 which means that the null hypothesis can't be rejected. Therefore, the residual are normally distributed and the interference of the OLS tests where correct.

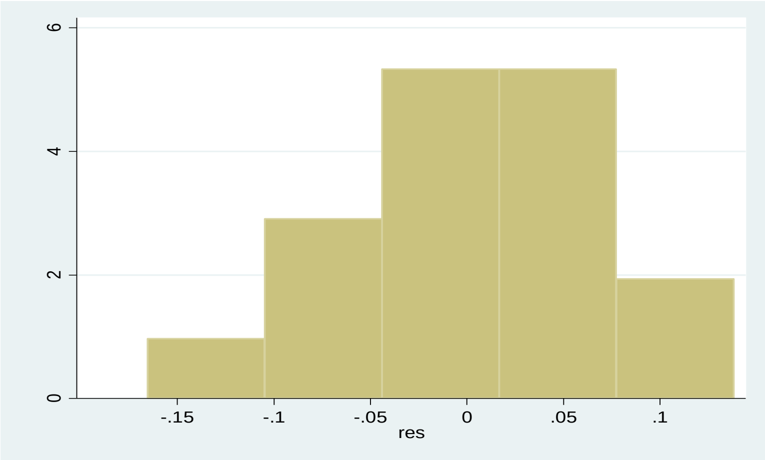


FIGURE 1: HISTOGRAM OF THE RESIDUALS

The figure shows a histogram of the residuals. When the residuals are normally distributed, it shows a bell shaped curve. This curve is not perfectly bell shaped, but still satisfactory.

4.4.4 AUTOCORRELATION

To test if the data contained autocorrelation I used Breusch-Godfrey test, because the revised model includes lagged values so the Durbin-Watson test is not applicable

H_0 = No serial correlation

H_1 = Serial correlation

Number of lags	Chi 2	df	Prob > Chi2
1	6,173	1	0,0130
2	11,674	2	0,0029
3	12,743	3	0,0052
4	16,345	4	0,0026
5	16,545	5	0,0054

TABLE 6: RESULTS FROM THE BREUSCH-GODFREY TEST

The test strongly rejects the null hypothesis of no serial correlation, even after the model has been refitted with lags. To support these results, I also did Durbin's alternative test which provides a formal test of the null hypothesis of serially uncorrelated disturbances against the alternative autocorrelation of order p.

Number of lags	Chi 2	Df	Prob > Chi2
1	6,211	1	0,0127
2	14,119	2	0,0009
3	15,586	3	0,0014
4	23,146	4	0,0001
5	22,748	5	0,0004

TABLE 7: RESULTS FROM DURBIN'S ALTERNATIVE TEST

As expected, the null hypothesis of no serial correlation is still strongly rejected. This means that the estimated coefficients are no longer BLUE and the variance and standard error are no longer valid. As mentioned above, I cannot ignore this since the estimates are inefficient.

4.5 STATIONARY DATA

Even if the there is proven to be a serial correlation in the data, a decided to still perform a test to see if the data is stationary. I used Dickey Fuller test and an expanded Dickey Fuller test. The time series where tested with both trend and operation. If the t-value in the table underneath was under the critical value, the null hypothesis of non-stationary data at either 5% or 10% significance level is rejected.

H_0 = The variables are not stationary

H_1 = The variables are stationary

Tests/ Variables	**L4.rdexp _log3**	**L1.rdstaff_ log3**	**L1.prod _log3**	**L4.cons _log3**	**Critical Value (5%) L4/L1**	**Critical Value (10%) L4/L1**
DF with drift	-13,129	-9,261	-10,246	-9,383	-1,694/-1,690	-1,309/-1,306
DF with trend	-12,931	-9,181	-9,993	-9,235	-3,564/-3,552	-3,218/-3,211
ADF with trend and lag1	-7,474	-7,184	-7,857	-5,589	-3,568/-3,556	-3,221/-3,214
ADF with trend and lag2	-6,394	-4,116	-6,481	-7,153	-3,572/-3,560	-3,223/-3,216
ADF with drift and lag1	-7,581	-7,245	-8,102	-5,679	-1,697/-1,692	-1,310/-1,308
ADF with drift and lag2	-6,196	-4,144	-6,898	-7,226	-1,701/-1,696	-1,313/-1,309

TABLE 8: RESULTS FROM DICKEY-FULLER TEST AND ADJUSTED DICKEY-FULLER TEST

The test tells us that there is no unit root and we can reject H_0. However, the Dickey Fuller test describes what is referred to as low strength. This means that the test can find unit rot in the data even if it’s not the case. That is why I have also done the Philips-Perron test. The results of the test are illustrated in the table below:

H_0: The variables are non-stationary
H_A: The variables are stationary

Tests/ Variables	**L4.rdexp_log 3**	**L1.rdstaff_log 3**	**L1.prod_log 3**	**L2.cons_log 3**	**Critical value (5%)**
Z (rho)	-54,947	-50,145	-58,736	-48,255	-12,788/-12,884
Z (t)	-14,671	-9,614	-11,028	-9,610	-2,619/-2,966

TABLE 9: RESULTS FROM PHILIPS-PERRON TEST FOR UNIT ROOT

The test values are smaller than the critical values and H_0 is rejected. Still, because of the findings of autocorrelation I conclude that the variables are non-stationary.

5.0 DISCUSSION

5.1 HYPOTHESIS

Investing in innovation has a positive effect on the Norwegian economic growth.

As mentioned earlier, for the OLS regression to give reliable results, all six Gauss-Markov assumptions must be met. However, the fourth assumption of OLS was not met, since I discovered serial correlation and no solution to this problem was found. This means that the variance and standard error is no longer valid since the estimated coefficients are no longer BLUE. I can therefore not confirm a relationship between innovation and economic growth with the collected data and analysis performed. There can be several reasons why my analysis fails to prove the hypothesis. Those who I believe to have the greatest significance will be discussed in the next section.

5.1.6 REASONS BEHIND THE RESULT

There is probably more than one reason for why my analysis fails to prove the hypothesis of a relationship between innovation and economic growth. First of all I believe the editing of missing data in R&D can have had an influence. With my choice to edit by plotting average numbers for the years missing, might have done more harm than good and produced answers that are inefficient. However, this was my only solution if I wanted to do execute a regression analysis with the data available, and I could say with certainty there has been a continuous growth in R&D investment over the last decades. Supported by the Norwegian Research Council, using averaging was the most reasonable choice of fixing the missing data problem.

Another issue is the selection of data and its reliability. The weakness with R&D data is how the collection of statistics has changed over the years. Nevertheless, if I had chosen to use R&D data from after the biggest changes had occurred, I would only have statistics from 1995 an onward, which would have given me an even smaller data set, and the results would not be trustworthy at all. One can question how different the results would have been if I had managed to collect usable patent statistics. The discussion of what measures innovations best between R&D data and patent statistics is ongoing, and using both in an analysis might be better since R&D measures innovation input while patents measures innovation output.

GDP is the traditional measurement of economic growth and the only measurement I had access to. Nevertheless, since GDP has so many weaknesses when it comes to measuring total production and especially welfare, it might be time to break this tradition and find other measures or methods for economic growth.

That R&D can hardly be regarded as an exogenous variable can also have an influence on my results. The amount that is invested in R&D often depends on the expected sales level. This makes it difficult to know which direction the causal link is working. When estimating how R&D affects growth or productivity at the aggregated level it can also be important to take into account the spillover effects from other countries. However, estimating the elasticity of foreign R&D with domestic productivity is usually pretty low. Maybe if I had taken the spillover effect into consideration and used data from other countries I might have gotten a different result, since social return is bigger than private return.

More than a few researchers that has explored the relationship between innovation and growth and gotten mixed results. This may be the reason why most innovational research is done on firm level. The data might be easier to measure. However, earlier research shows us that also at firm level, the studies of innovations profitability are volatile. Since there are so many uncertainties with earlier studies in addition to looking at my results, I believe that the perfect way of measuring the relationship between innovation and economic growth is still not discovered and might never be. There is too much that can influence these types of innovation statistics.

Another thing to consider when it comes to unsatisfactory results is the science’s claim to objective truth. Where the foundation is based on the idea that the same experience always gets the same results, no matter who performs them. Even academic scientists admit that they often get things wrong. However, they believe that these errors get corrected over time as other scientists try to take the work further. For example, the next time someone wants to confirm the relationship between innovation and economic growth; they might choose to use other data, use another solution to the missing value problem, or use a totally different approach, and from there possibly get different results that might get us a step closer towards an easier way to look at the relationship between innovation and economic growth. It is therefore just as important to show the negative results as the positive results. This way, other researchers won't spend time and money trying to do the same as for example I have done here, but rather learn from my errors and advance from them.

Overall, looking at the theory presented and the increasing amounts invested in innovation and R&D, one would believe that there still is a positive relationship between innovation and economic growth. Every research and innovation might not be profitable, but in the bigger picture innovation increases the economic growth. Inventions today might not have the same massive impact as the Romans aqueducts or the inventions of the second industrial revolution, but they still matter. As Shiller believes; innovations can be the solution to a good society, even if its impact on economic growth decreases, it is still important for our society to grow towards an effective and congenial whole.

5.2 WEAKNESESS IN THE ANALYSIS

During the thesis I have encountered several problems that I have tried to solve in the best possible way. I see afterwards that some problems could have been solved differently and that my analysis contains certain weaknesses.

First and foremost, there is uncertainty associated with R&D data. Since there are no annual data, I had to calculate the missing values. This made the data less predictable, and I had to do some adjustment to make the regression analysis work. Only having yearly data for 42 years is not ideal and is not really enough to complete a good regression analysis. There is a high possibility that this is why autocorrelation occurred.

As discussed above there are suggestions to what could have been done differently. One solution might be that when looking at the aggregated level, it might be important to take into account the effects from other countries. The spillover effect can give more descriptive results. There might have been easier to look at private R&D investments in companies or one sector, such collected data from a number of companies could give a more "complete" data set.

There is a probability this hypothesis was to extensive for in such a small time frame. A lot of time went into searching and collecting data, which never became perfect. If I was to was to redo this paper, I would most likely either include data from other OECD countries and do a comparison between countries, or I would have chosen to do an analysis on company level with a selection of companies and looked at the effect research and development investment have on their financial growth.

6.0 CONCLUSION

The purpose of this paper is to identify the impact innovation has on economic growth in Norway. I wanted to prove that investing in innovation leads to economic growth, and is the basis of my hypothesis:

Investing in innovation has a positive effect on the Norwegian economic growth

The theoretical review of this paper shows that the hypothesis matches the basic theory. However, previous researchers have had difficulty proving this statement.

Not all the assumptions of OLS were met in the regression analysis, consequently the results were not valid, and the relation between innovation and economic growth was not proven. There can be several reasons why this occurred; unreliable collected data, missing value edit, or innovation and growth is just not possible to test by standard methods. Nevertheless, it is just as important to show the negative results as the positive results.

Despite the results in the analysis, one can still see from the theory and the increasing amount invested in innovation and R&D, that there is a positive relationship between innovation and economic growth. Perhaps the relationship is not as strong as it once was, but innovation is important. Not just for economic progress but for the society as a whole.

7.0 REFERENCES

Cameron, G. (1996). *Innovation and Economic Growth.* London: Center for Economic Performance, LSE.

Cappelen, Å., Raknerud, A., & Rybalka, M. (2007). *Resultater of SkatteFUNN - patentering og innovasjoner.* Oslo - Kongsvinger: Statistisk Sentralbyrå.

Cho, H.-J., & Pucik, V. (2005, April 11). Relationship between innovativeness, quality, growth, profitability, and market value. *Straegic Management Journal. Volume 26, Issue 6*, pp. 555-575.

Forskningsrådet. (2013, Juli 15). *Forskningsrådet*. Hentet fra http://www.forskningsradet.no/prognett-skattefunn/Artikkel/Hva_er_SkatteFUNN/1253988114414

Geroski, P., Machin, S., & Van Reenen, J. (1993, Summer). The pofitability of innovating firms. *The RAND Journal of Economics Vol 24, No. 2*, pp. 198-211. Retrieved from http://www.jstor.org/discover/10.2307/2555757?uid=3738744&uid=2129&uid=2&uid=70&uid=4&sid=21103651406617

Gordon, R. J. (2012, Desember 21). *The Wall Street Journal*. Retrieved from http://online.wsj.com/news/articles/SB10001424127887324461604578191781756437940

Longva, S., & Blekstad, B. (2004). *Forskning og utvikling i næringslivet 2001-2002.* Oslo-Kongsvinger: Statistisk Sentralbyrå.

Miljøverndepartementet. (2013). *Naturens goder - om verdier av økosystemtjenester.* NOU 2013:10 Oslo: Departementet.

Munthe, P. (1992). *Sirkulasjon, inntekt og økonomisk vekst.* Oslo: Universitetsforlaget.

Nærings- og Handelsdepartement. (2008). *Et Nyskapende og bærekraftig Norge.* St.meld.nr.7 (2008-2009). Oslo: Departementet.

Nærings- og Fiskeridepartementet. (2014, February 17). *Forskning og utvikling i næringslivet*. Hentet fra Nærings- og Fiskeridepartementet:

http://www.regjeringen.no/nb/dep/nfd/tema/forskning-og-utvikling-i-naringslivet/forskning-og-utvikling-i-naringslivet-.html?id=426434

Nås, S. O., & Leppälahti, A. (1997). *Innovation, firm profitability and growth.* Oslo: STEP Group.

Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods, Vol 7, No. 2,* pp. 147-177. Retrieved from http://isites.harvard.edu/fs/docs/icb.topic469678.files/missing_data_1.pdf

Shiller, R. J. (2012). *Finance and the Good Society.* Princeton, New Jersey: Princeton University Press.

Shiller, R. J. (2013, May). *Reflections on Finance and the Good Society.* Retrieved from Cowles Foundation for Research in Economics Yale University: http://cowles.econ.yale.edu/

Solbakken, K. (2011). *Oljemarkedets påvirkning på tankmarkedet*. (Masteroppgave) Bergen: Norges Handelhøyskole.

Statistisk Sentralbyrå. (2012, April 16). *Begreper i Nasjonalregnskapet*. Hentet fra Statistisk Sentralbyrå: http://www.ssb.no/nasjonalregnskap-og-konjunkturer/begreper-i-nasjonalregnskapet

Steigum, E. (2004). *Moderne Makroøkonomi.* Oslo: Gyldendal Norsk Forlag.

Studenmund, A. H. (2006). *Using Econometrics A Practical Guide, 5th Edition.* Boston: Pearson Education, Inc.

Svensson, R. (2008). *Growth through Research and Development - what does the research literature say?* Stockholm: VINNOVA - Swedish Governmental Agency for Innovation Systems.

The Economist. (2013a, October 19). *Unreliable research: Trouble in the lab.* Retrieved from The Economist: http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble

The Economist. (2013b, October 19). *Problems with scientific research: How science goes wrong.* Retrieved from The Economist:

http://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong

Ulku, H. (2004). *R&D, Innovation, and Economic Growth: An Empirical Analysis.* London: International Monetary Fund.

Wang, C. (2013). *The Long-run Effect of Innovation on Economic Growth.* (Thesis) Sydney, Australia: School of Economics.

Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach 4e.* Canada: South-Western CENGAGE Learning.

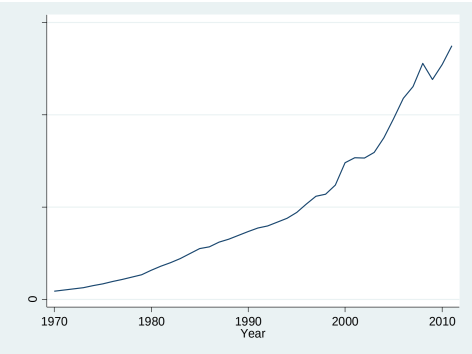
Ørstavik, F. (2001). *Innovasjoner - suksesser? Identifisere innovasjoner 3 år etter.* Oslo: STEP Group.

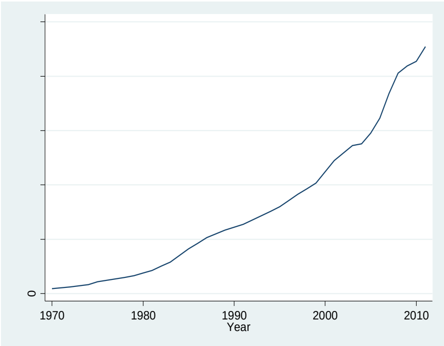
8.0 APPENDIX

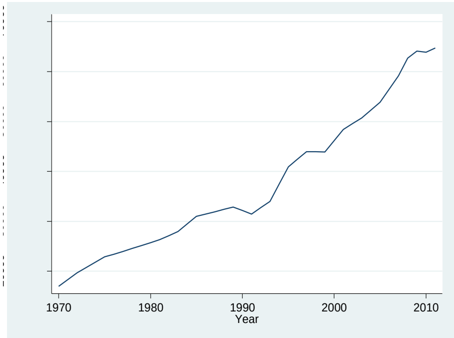
FIRST REGRESSION WITH UNPROCESSED DATA

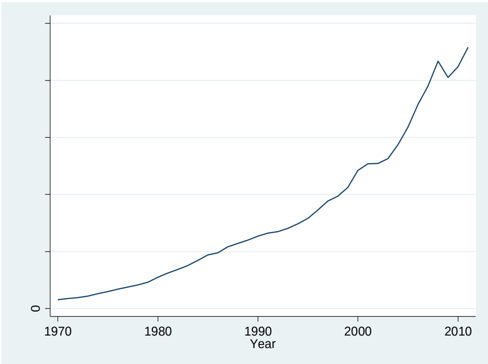
```
. regress gdpinnok rdexpenditure rdstaffinnorway productionandvalue consuminhou
> seholdsandidealo

      Source |       SS       df       MS              Number of obs =      42
-------------+------------------------------           F(  4,    37) =18130.66
       Model |  2.5288e+25     4  6.3220e+24           Prob > F      =  0.0000
    Residual |  1.2902e+22    37  3.4869e+20           R-squared     =  0.9995
-------------+------------------------------           Adj R-squared =  0.9994
       Total |  2.5301e+25    41  6.1709e+23           Root MSE      =  1.9e+10

------------------------------------------------------------------------------
    gdpinnok |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rdexpendit~e |  -10.85587   3.805608    -2.85   0.007    -18.56676   -3.144975
rdstaffinn~y |   -1912391    1950059    -0.98   0.333     -5863586     2038803
production~e |   .6386966   .0307649    20.76   0.000      .576361    .7010322
consuminho~o |   .3710703   .1475304     2.52   0.016     .0721452    .6699954
       _cons |   4.27e+09   3.56e+10     0.12   0.905    -6.78e+10    7.63e+10
------------------------------------------------------------------------------
```

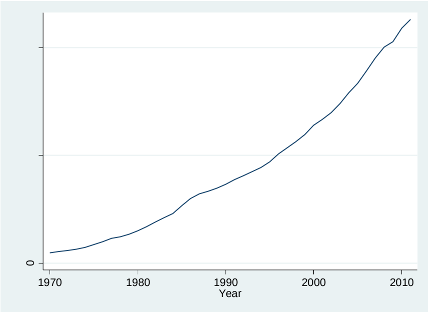
MEASURING TO SEE IF THE VARIABLES ARE STATIONARY







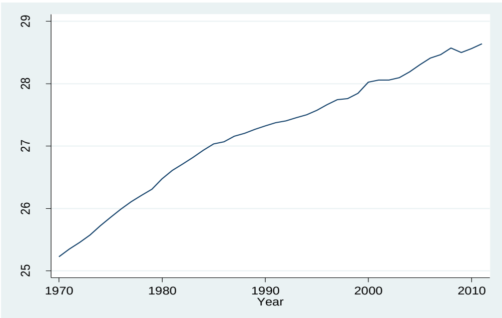


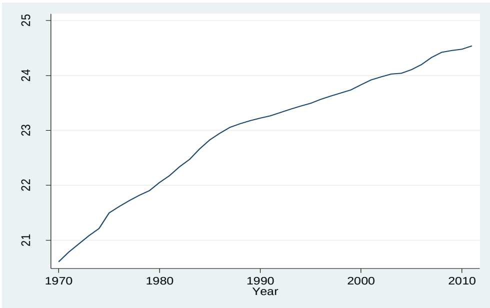


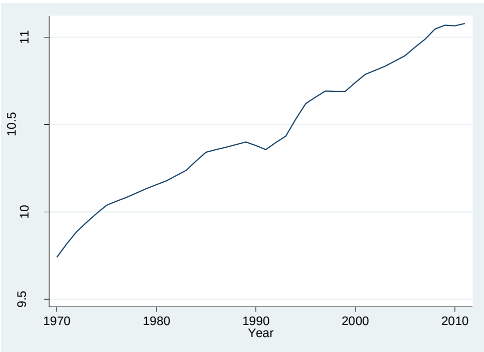
We can see clearly that the variables are not stationary.

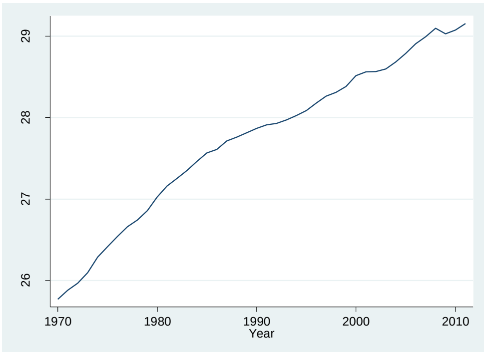
LOGGING THE VARIABLES

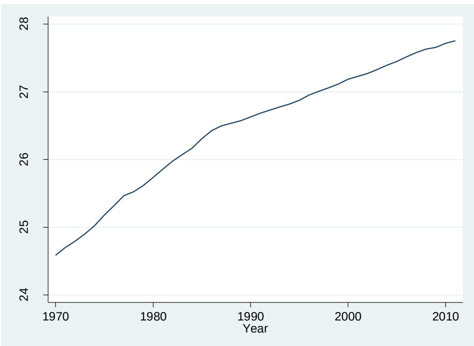
Gen gdpinnok_log = log (gdpinnok)





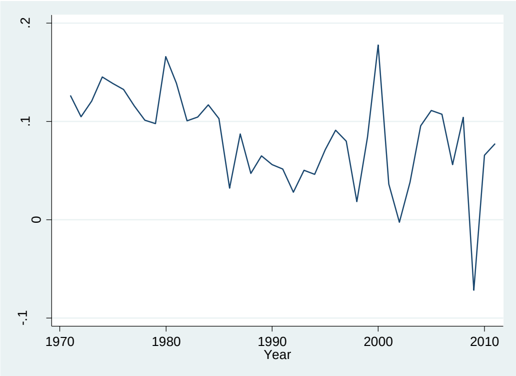


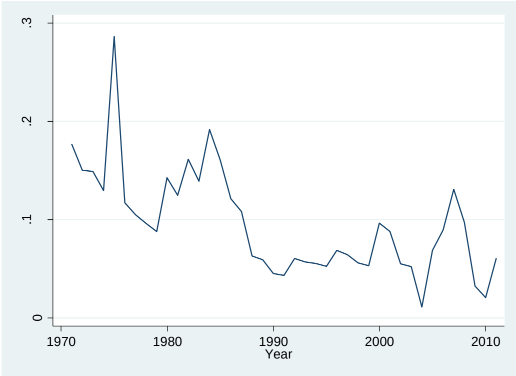


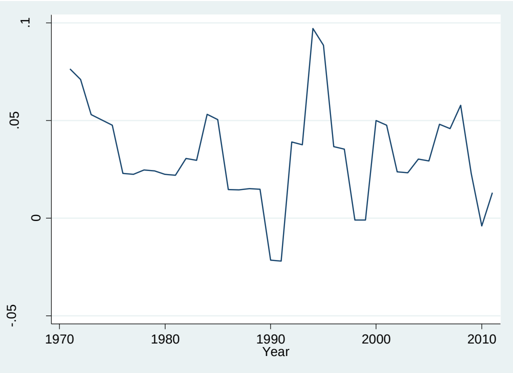


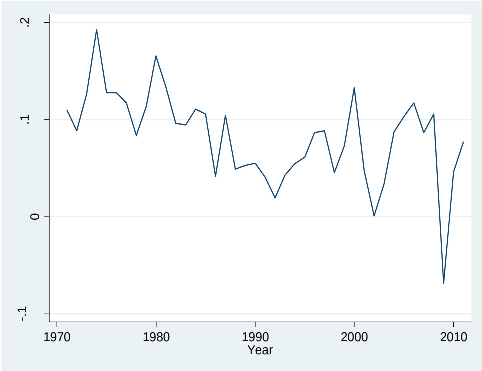
This doesn’t make that much difference.

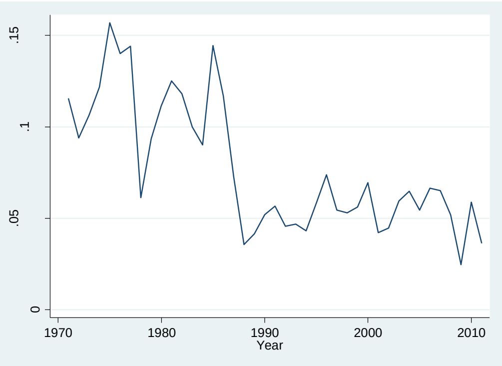
gen gdpinnok_log1= gdpinnok_log-L.gdpinnok_log







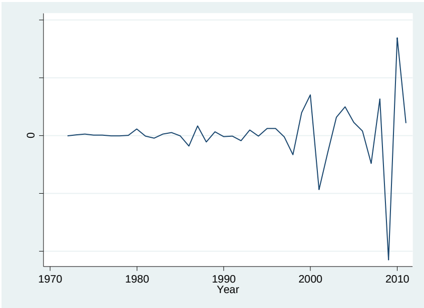




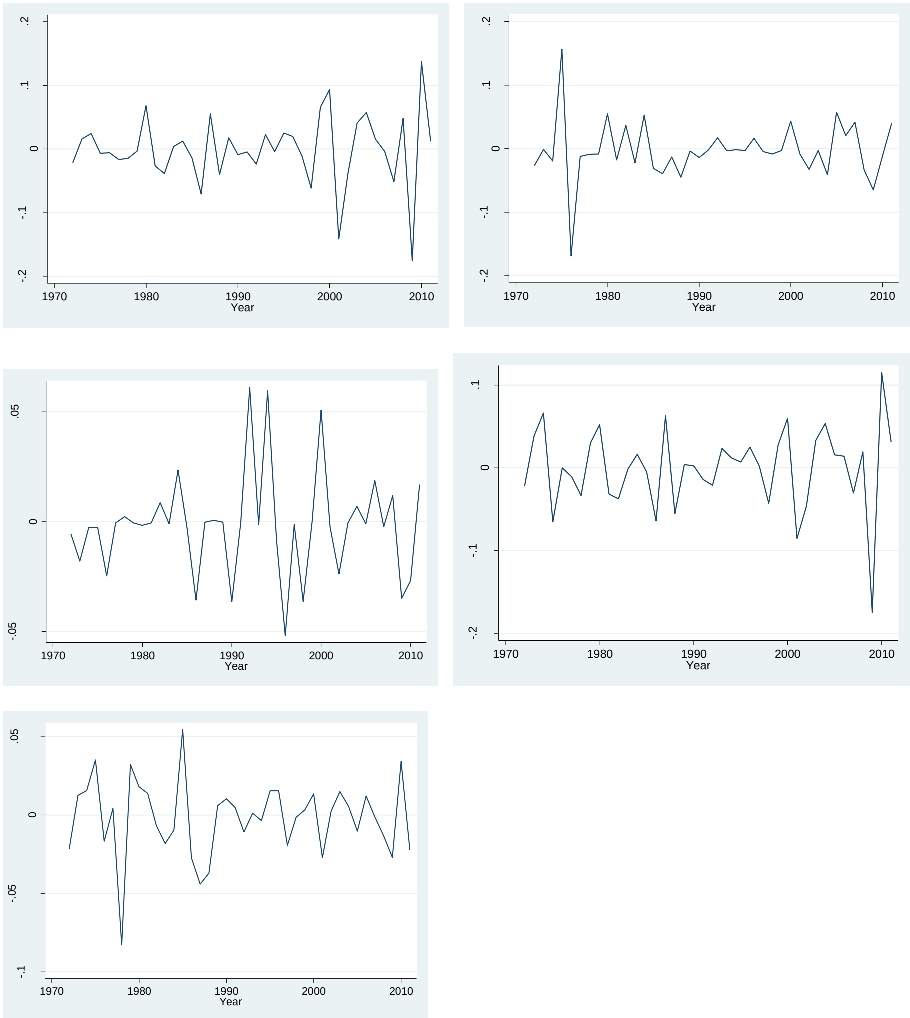
We can still see a trend.

Try to differentiate more than once

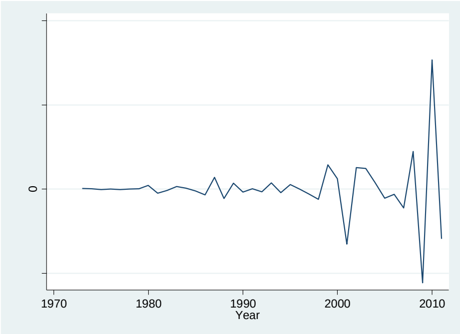
gen gdpinnok2= gdpinnok1-L.gdpinnok1



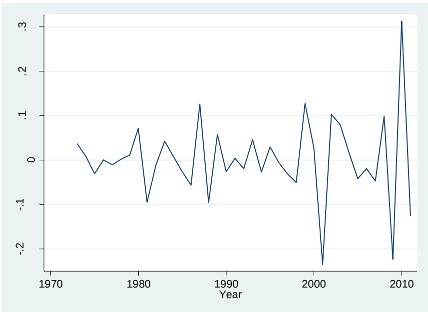
gen gdpinnok_log2= gdpinnok_log1-L.gdpinnok_log1

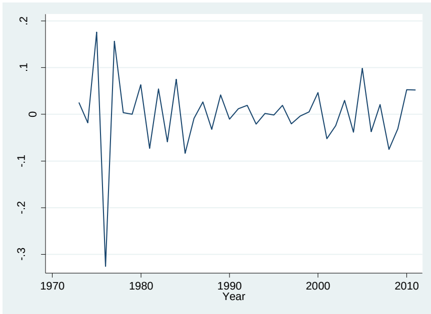


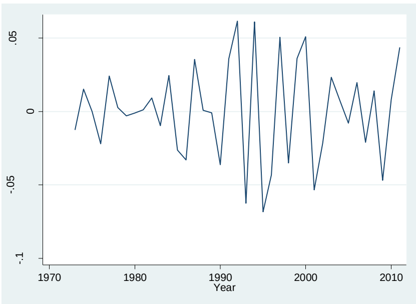
gen gdpinnok3= gdpinnok2-L.gdpinnok2

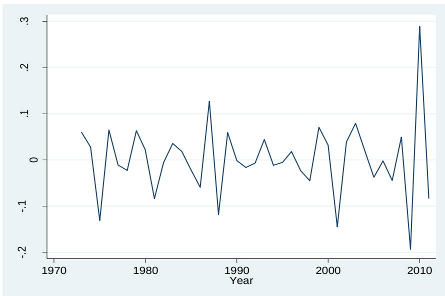


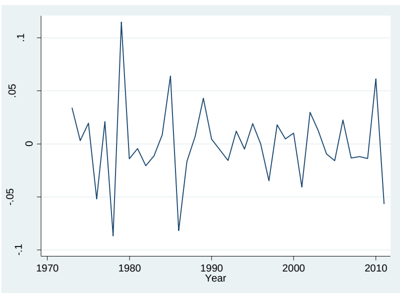
gen gdpinnok_log3= gdpinnok_log2-L.gdpinnok_log2











After logging the variables and differentiate three times, the variables are looking more stationary. But it needs to be tested.

_log1

```
. regress gdpinnok_log1 rdexpenditure_log1 rdstaffinnorway_log1 productionandva
> lue_log1 consuminhouseholdsandidealo_log1

      Source |       SS       df       MS              Number of obs =      41
-------------+------------------------------           F(  4,    36) =   79.11
       Model |  .08108126     4  .020270315           Prob > F      =  0.0000
    Residual | .009223856    36  .000256218           R-squared     =  0.8979
-------------+------------------------------           Adj R-squared =  0.8865
       Total | .090305116    40  .002257628           Root MSE      =  .01601

------------------------------------------------------------------------------
gdpinnok_l~1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rdexpendit~1 |   .0115774   .0812244     0.14   0.887    -.1531533    .1763081
rdstaffinn~1 |   .0117228   .1109247     0.11   0.916     -.213243    .2366886
production~1 |   .9550505   .0750671    12.72   0.000     .8028073    1.107294
consuminho~1 |   .0059535   .1300701     0.05   0.964    -.2578408    .2697479
       _cons |   .0026254    .006576     0.40   0.692    -.0107114    .0159622
------------------------------------------------------------------------------
```

_log2

```
. regress gdpinnok_log2 rdexpenditure_log2 rdstaffinnorway_log2 productionandva
> lue_log2 consuminhouseholdsandidealo_log2

      Source |       SS       df       MS              Number of obs =      40
-------------+------------------------------           F(  4,    35) =   54.09
       Model | .103030502     4  .025757626           Prob > F      =  0.0000
    Residual | .016666383    35  .000476182           R-squared     =  0.8608
-------------+------------------------------           Adj R-squared =  0.8448
       Total | .119696886    39  .003069151           Root MSE      =  .02182

------------------------------------------------------------------------------
gdpinnok_l~2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rdexpendit~2 |     .11318   .0828945     1.37   0.181    -.0551048    .2814649
rdstaffinn~2 |    .001237   .1650898     0.01   0.994    -.3339131    .3363871
production~2 |   1.010492   .0777866    12.99   0.000     .8525766    1.168407
consuminho~2 |   .0139159   .1626042     0.09   0.932    -.3161881    .3440198
       _cons |  -.0000415   .0034736    -0.01   0.991    -.0070932    .0070103
------------------------------------------------------------------------------
```

_log3

```
. regress gdpinnok_log3 rdexpenditure_log3 rdstaffinnorway_log3 productionandva
> lue_log3 consuminhouseholdsandidealo_log3

      Source |       SS       df       MS              Number of obs =      39
-------------+------------------------------           F(  4,    34) =   64.27
       Model | .288433178     4  .072108295           Prob > F      =  0.0000
    Residual |  .03814731    34   .00112198           R-squared     =  0.8832
-------------+------------------------------           Adj R-squared =  0.8694
       Total | .326580488    38  .008594223           Root MSE      =   .0335

------------------------------------------------------------------------------
gdpinnok_l~3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rdexpendit~3 |    .187194   .0761907     2.46   0.019     .0323559     .342032
rdstaffinn~3 |  -.0349194   .1780343    -0.20   0.846    -.3967286    .3268898
production~3 |    1.07221   .0762738    14.06   0.000     .9172033    1.227217
consuminho~3 |  -.0370403   .1664479    -0.22   0.825    -.3753031    .3012225
       _cons |  -.0009011   .0053664    -0.17   0.868    -.0118069    .0100047
------------------------------------------------------------------------------
```

High p-values RD_staff and Consume_households, also still a pretty high R^2

CHECKING FOR LAGS FOR THE NEW VARIABLES _LOG3

```
. varsoc  rdexpenditure_log3, maxlag(5) exog( rdstaffinnorway_log3 productionan
> dvalue_log3 consuminhouseholdsandidealo_log3)

   Selection-order criteria
   Sample:  1978 - 2011                            Number of obs      =        34
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	61.1775				.002029	-3.36338	-3.30214	-3.18381
1	66.1355	9.9161	1	0.002	.001609	-3.59621	-3.51966	-3.37174
2	66.194	.11692	1	0.732	.001704	-3.54082	-3.44896	-3.27146
3	66.7259	1.064	1	0.302	.001755	-3.51329	-3.40612	-3.19904
4	73.114	12.776*	1	0.000	.001282*	-3.83024*	-3.70776*	-3.47109*
5	73.3073	.38659	1	0.534	.00135	-3.78278	-3.645	-3.37875

```
   Endogenous:  rdexpenditure_log3
    Exogenous:  rdstaffinnorway_log3 productionandvalue_log3
                consuminhouseholdsandidealo_log3  _cons
```

```
. varsoc rdstaffinnorway_log3, maxlag(5) exog( rdexpenditure_log3 productionand
> value_log3 consuminhouseholdsandidealo_log3)

   Selection-order criteria
   Sample:  1978 - 2011                            Number of obs      =        34
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	68.7886				.001297	-3.8111	-3.74986	-3.63152
1	70.4283	3.2792	1	0.070	.00125	-3.84872	-3.77217	-3.62426
2	73.3278	5.7991	1	0.016	.00112	-3.96046	-3.8686	-3.6911*
3	73.3594	.06322	1	0.801	.001188	-3.9035	-3.79633	-3.58924
4	76.7998	6.8807*	1	0.009	.001032*	-4.04705*	-3.92457*	-3.6879
5	77.3723	1.1451	1	0.285	.001063	-4.0219	-3.88411	-3.61787

```
   Endogenous:  rdstaffinnorway_log3
    Exogenous:  rdexpenditure_log3 productionandvalue_log3
                consuminhouseholdsandidealo_log3  _cons

. . varsoc productionandvalue_log3, maxlag(5) exog( rdexpenditure_log3 rdstaffi
> nnorway_log3 consuminhouseholdsandidealo_log3)

   Selection-order criteria
   Sample:  1978 - 2011                            Number of obs      =        34
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	44.5045				.005411	-2.38262	-2.32138	-2.20305
1	51.7867	14.564	1	0.000	.003743	-2.75216	-2.67561	-2.5277
2	55.0103	6.4472	1	0.011	.003289	-2.88296	-2.7911	-2.6136
3	58.859	7.6974	1	0.006	.002788	-3.05053	-2.94336	-2.73628*
4	58.8591	9.6e-05	1	0.992	.002966	-2.99171	-2.86923	-2.63257
5	61.4151	5.1121*	1	0.024	.002717*	-3.08324*	-2.94546*	-2.67921

```
   Endogenous:  productionandvalue_log3
    Exogenous:  rdexpenditure_log3 rdstaffinnorway_log3
                consuminhouseholdsandidealo_log3  _cons

. varsoc consuminhouseholdsandidealo_log3, maxlag(5) exog( rdexpenditure_log3 r
> dstaffinnorway_log3 productionandvalue_log3)

   Selection-order criteria
   Sample:  1978 - 2011                            Number of obs      =        34
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	68.7917				.001297	-3.81128	-3.75004	-3.63171
1	72.1966	6.8096	1	0.009	.001127	-3.95274	-3.87619	-3.72827
2	72.947	1.5009	1	0.221	.001145	-3.93806	-3.8462	-3.6687
3	76.9233	7.9525	1	0.005	.000963	-4.11313	-4.00596	-3.79888
4	79.2162	4.5858	1	0.032	.000896	-4.18919	-4.06671	-3.83004
5	82.2074	5.9824*	1	0.014	.0008*	-4.30632*	-4.16853*	-3.90228*

```
   Endogenous:  consuminhouseholdsandidealo_log3
    Exogenous:  rdexpenditure_log3 rdstaffinnorway_log3
                productionandvalue_log3  _cons
```

RUN REGRESSIONS WITH DIFFERENT LAGS TO FIND THE BEST NUMBER OF LAGS FOR THE MODEL

```
. regress  gdpinnok_log3 L4.rdexpenditure_log3 L1.rdstaffinnorway_log3 L1.produ
> ctionandvalue_log3 L4.consuminhouseholdsandidealo_log3

      Source |       SS       df       MS              Number of obs =      35
-------------+------------------------------           F(  4,    30) =    8.13
       Model |  .16862973     4  .042157433           Prob > F      =  0.0001
    Residual |  .15553325    30  .005184442           R-squared     =  0.5202
-------------+------------------------------           Adj R-squared =  0.4562
       Total |  .32416298    34  .009534205           Root MSE      =    .072

------------------------------------------------------------------------------
gdpinnok_l~3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rdexpendit~3 |
         L4. |  -.3635198   .1605916    -2.26   0.031    -.6914916    -.035548
             |
rdstaffinn~3 |
         L1. |  -.7076267   .3749453    -1.89   0.069    -1.473367    .0581137
             |
production~3 |
         L1. |  -.5368502   .1643594    -3.27   0.003    -.8725169   -.2011835
             |
consuminho~3 |
         L4. |   .6228628   .3545355     1.76   0.089    -.1011954    1.346921
             |
       _cons |   .0031249   .0122129     0.26   0.800    -.0218172     .028067
------------------------------------------------------------------------------
```

TESTING FOR MULTICOLLINEARITY: MAKING A CORRELATION DIAGRAM

```
. corr gdpinnok_log3 L4.rdexpenditure_log3 L1.rdstaffinnorway_log3 L1.productio
> nandvalue_log3 L4.consuminhouseholdsandidealo_log3
(obs=35)

             |                L4.        L.        L.       L4.
             | gdpin~g3 rdexpe~3 rdstaf~3 produc~3 consum~3
-------------+---------------------------------------------
gdpinnok_l~3 |   1.0000
rdexpendit~3 |
         L4. |  -0.3235   1.0000
rdstaffinn~3 |
         L1. |  -0.3994   0.0845   1.0000
production~3 |
         L1. |  -0.5955   0.1193   0.2838   1.0000
consuminho~3 |
         L4. |   0.2565   0.2025   0.0063  -0.1915   1.0000
```

```
. corr L4.rdexpenditure_log3 L1.rdstaffinnorway_log3 L1.productionandvalue_log3
>  L4.consuminhouseholdsandidealo_log3
(obs=35)

             |       L4.        L.        L.       L4.
             | rdexpe~3 rdstaf~3 produc~3 consum~3
-------------+------------------------------------
rdexpendit~3 |
         L4. |   1.0000
rdstaffinn~3 |
         L1. |   0.0845   1.0000
production~3 |
         L1. |   0.1193   0.2838   1.0000
consuminho~3 |
         L4. |   0.2025   0.0063  -0.1915   1.0000
```

```
. hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of gdpinnok_log3

         chi2(1)      =     0.00
         Prob > chi2  =   0.9782
```

A large chi-square would indicate that heteroscedasticity was present. In this example, the chi-square value was small, indicating heteroscedasticity was probably not a problem (or at least that if it was a problem, it wasn't a multiplicative function of the predicted values).

```
. estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(14)     =    24.06
         Prob > chi2  =    0.0450

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

Source	chi2	df	p
Heteroskedasticity	24.06	14	0.0450
Skewness	10.35	4	0.0349
Kurtosis	0.08	1	0.7720
Total	34.50	19	0.0160

The White test on the other hand is more generic. It relies on the intuition that if there is no heteroscedasticity the classical error variance estimator should gives you standard error estimates close enough to those estimated by the robust estimator. Therefore, it is able to detect more general form of heteroscedasticity than the Breusch-Pagan test.

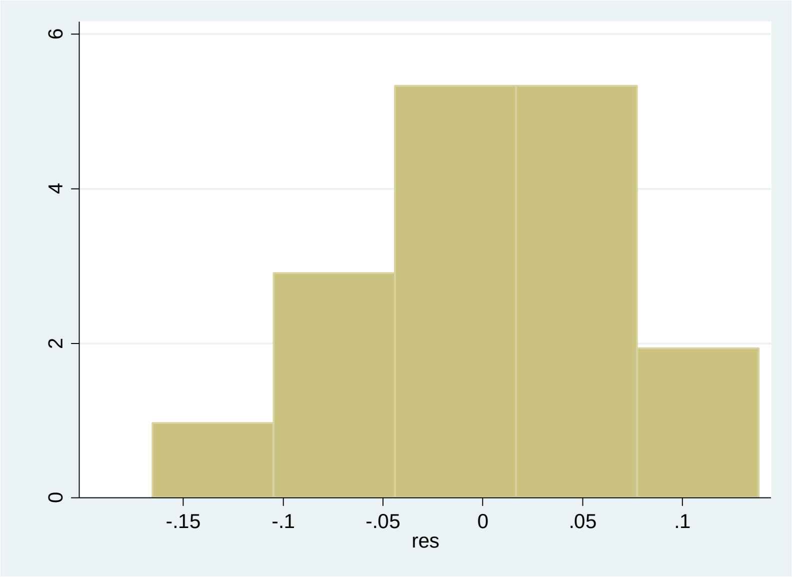
TESTING FOR NORMALLY DISTRIBUTED RESIDUALS

```
. predict res, r
(8 missing values generated)

. sktest res

                    Skewness/Kurtosis tests for Normality
```

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2) (joint)	Prob>chi2 (joint)
res	34	0.3841	0.7810	0.88	0.6447

.

```
. histogram res,  xtitle(res)
(bin=5, start=-.16548645, width=.06069104)
```



Almost bell shaped, approved.

Breuch-Godfrey test for different number of lags

```
. bgodfrey
```

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	6.173	1	0.0130

H0: no serial correlation

```
. bgodfrey, lags(1)
```

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	6.173	1	0.0130

H0: no serial correlation

```
. bgodfrey, lags(2)
```

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
2	11.674	2	0.0029

H0: no serial correlation

```
. bgodfrey, lags(3)
```

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
3	12.743	3	0.0052

H0: no serial correlation

```
. bgodfrey, lags(4)
```

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
4	16.345	4	0.0026

H0: no serial correlation

```
. bgodfrey, lags(5)
```

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
5	16.545	5	0.0054

H0: no serial correlation

```
. estat durbinalt
```

Durbin's alternative test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	6.211	1	0.0127

H0: no serial correlation

```
. estat durbinalt, lags (1)
```

Durbin's alternative test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	6.211	1	0.0127

H0: no serial correlation

```
. estat durbinalt, lags (2)
```

Durbin's alternative test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
2	14.119	2	0.0009

H0: no serial correlation

```
. estat durbinalt, lags (3)
```

Durbin's alternative test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
3	15.586	3	0.0014

H0: no serial correlation

```
. estat durbinalt, lags (4)
```

Durbin's alternative test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
4	23.146	4	0.0001

H0: no serial correlation

```
. estat durbinalt, lags (5)
```

Durbin's alternative test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
5	22.748	5	0.0004

H0: no serial correlation

.

ADF m/ trend og lags

```
. dfuller L4.rdexpenditure_log3, trend lags (1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        33

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -7.474            -4.306            -3.568            -3.221
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller L4.rdexpenditure_log3, trend lags (2)

Augmented Dickey-Fuller test for unit root         Number of obs   =        32

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -6.394            -4.316            -3.572            -3.223
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller L1.rdstaffinnorway_log3, trend lags (1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        36

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -7.184            -4.279            -3.556            -3.214
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller L1.rdstaffinnorway_log3, trend lags (2)

Augmented Dickey-Fuller test for unit root         Number of obs   =        35

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -4.116            -4.288            -3.560            -3.216
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0060

. dfuller L1. productionandvalue_log3, trend lags (1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        37

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -7.857            -4.270            -3.552            -3.211
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller L1. productionandvalue_log3, trend lags (2)

Augmented Dickey-Fuller test for unit root         Number of obs   =        36

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -6.481            -4.279            -3.556            -3.214
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

```
. dfuller L4.consuminhouseholdsandidealo_log3, trend lags (1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        33

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -5.589            -4.306            -3.568            -3.221
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller L4.consuminhouseholdsandidealo_log3, trend lags (2)

Augmented Dickey-Fuller test for unit root         Number of obs   =        32

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -7.153            -4.316            -3.572            -3.223
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

ADF m/drift og lags

```
. dfuller L4.rdexpenditure_log3, drift lags(1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        33

                               ----------- Z(t) has t-distribution ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -7.581            -2.457            -1.697            -1.310
------------------------------------------------------------------------------
p-value for Z(t) = 0.0000

.
. dfuller L4.rdexpenditure_log3, drift lags(2)

Augmented Dickey-Fuller test for unit root         Number of obs   =        32

                               ----------- Z(t) has t-distribution ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -6.196            -2.467            -1.701            -1.313
------------------------------------------------------------------------------
p-value for Z(t) = 0.0000

. dfuller L1.rdstaffinnorway_log3, drift lags(1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        36

                               ----------- Z(t) has t-distribution ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -7.245            -2.445            -1.692            -1.308
------------------------------------------------------------------------------
p-value for Z(t) = 0.0000

.
. dfuller L1.rdstaffinnorway_log3, drift lags(2)

Augmented Dickey-Fuller test for unit root         Number of obs   =        35

                               ----------- Z(t) has t-distribution ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -4.144            -2.453            -1.696            -1.309
------------------------------------------------------------------------------
p-value for Z(t) = 0.0001
```

```
. dfuller L1.productionandvalue_log3, drift lags(1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        36

                               ─────────── Z(t) has t-distribution ───────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
──────────────────────────────────────────────────────────────────────────────
 Z(t)             -8.102            -2.445            -1.692            -1.308
──────────────────────────────────────────────────────────────────────────────
p-value for Z(t) = 0.0000

.
. dfuller L1.productionandvalue_log3, drift lags(2)

Augmented Dickey-Fuller test for unit root         Number of obs   =        35

                               ─────────── Z(t) has t-distribution ───────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
──────────────────────────────────────────────────────────────────────────────
 Z(t)             -6.898            -2.453            -1.696            -1.309
──────────────────────────────────────────────────────────────────────────────
p-value for Z(t) = 0.0000

.
. dfuller L4.consuminhouseholdsandidealo_log3, drift lags(1)

Augmented Dickey-Fuller test for unit root         Number of obs   =        33

                               ─────────── Z(t) has t-distribution ───────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
──────────────────────────────────────────────────────────────────────────────
 Z(t)             -5.679            -2.457            -1.697            -1.310
──────────────────────────────────────────────────────────────────────────────
p-value for Z(t) = 0.0000

.
. dfuller L4.consuminhouseholdsandidealo_log3, drift lags(2)

Augmented Dickey-Fuller test for unit root         Number of obs   =        32

                               ─────────── Z(t) has t-distribution ───────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
──────────────────────────────────────────────────────────────────────────────
 Z(t)             -7.226            -2.467            -1.701            -1.313
──────────────────────────────────────────────────────────────────────────────
p-value for Z(t) = 0.0000
```

DF m/ trend

```
. dfuller L4.rdexpenditure_log3, trend

Dickey-Fuller test for unit root                   Number of obs   =        34

                               ────────── Interpolated Dickey-Fuller ─────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
──────────────────────────────────────────────────────────────────────────────
 Z(t)            -12.931            -4.297            -3.564            -3.218
──────────────────────────────────────────────────────────────────────────────
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller L1.rdstaffinnorway_log3, trend

Dickey-Fuller test for unit root                   Number of obs   =        37

                               ────────── Interpolated Dickey-Fuller ─────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
──────────────────────────────────────────────────────────────────────────────
 Z(t)             -9.181            -4.270            -3.552            -3.211
──────────────────────────────────────────────────────────────────────────────
MacKinnon approximate p-value for Z(t) = 0.0000
```

```
. dfuller L1.productionandvalue_log3, trend

Dickey-Fuller test for unit root                   Number of obs   =        37

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -9.993            -4.270            -3.552            -3.211
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. dfuller L4.consuminhouseholdsandidealo_log3, trend

Dickey-Fuller test for unit root                   Number of obs   =        34

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -9.235            -4.297            -3.564            -3.218
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

DF m/drift

```
. dfuller L4.rdexpenditure_log3, drift

Dickey-Fuller test for unit root                   Number of obs   =        34

                               ----------- Z(t) has t-distribution -----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -13.129            -2.449            -1.694            -1.309
------------------------------------------------------------------------------
p-value for Z(t) = 0.0000
. dfuller L1.rdstaffinnorway_log3, drift

Dickey-Fuller test for unit root                   Number of obs   =        37

                               ----------- Z(t) has t-distribution -----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -9.261            -2.438            -1.690            -1.306
------------------------------------------------------------------------------
p-value for Z(t) = 0.0000
. dfuller L1.productionandvalue_log3, drift

Dickey-Fuller test for unit root                   Number of obs   =        37

                               ----------- Z(t) has t-distribution -----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -10.246            -2.438            -1.690            -1.306
------------------------------------------------------------------------------
p-value for Z(t) = 0.0000
. dfuller L4.consuminhouseholdsandidealo_log3, drift

Dickey-Fuller test for unit root                   Number of obs   =        34

                               ----------- Z(t) has t-distribution -----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -9.383            -2.449            -1.694            -1.309
------------------------------------------------------------------------------
p-value for Z(t) = 0.0000

.
```

All the test statistics are below the critical value of 5%, therefore we can say that the variables are stationary.

```
. pperron L4.rdexpenditure_log3, lags (1)

Phillips-Perron test for unit root                Number of obs   =        34
                                                  Newey-West lags =         1

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(rho)          -54.947           -17.812           -12.788           -10.380
 Z(t)            -14.671            -3.689            -2.975            -2.619
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. pperron L1.rdstaffinnorway_log3, lags (1)

Phillips-Perron test for unit root                Number of obs   =        37
                                                  Newey-West lags =         1

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(rho)          -50.145           -18.016           -12.884           -10.440
 Z(t)             -9.614            -3.668            -2.966            -2.616
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. pperron L1.productionandvalue_log3, lags (1)

Phillips-Perron test for unit root                Number of obs   =        37
                                                  Newey-West lags =         1

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(rho)          -58.736           -18.016           -12.884           -10.440
 Z(t)            -11.025            -3.668            -2.966            -2.616
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

. pperron L4.consuminhouseholdsandidealo_log3, lags (1)

Phillips-Perron test for unit root                Number of obs   =        34
                                                  Newey-West lags =         1

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value            Value
------------------------------------------------------------------------------
 Z(rho)          -48.255           -17.812           -12.788           -10.380
 Z(t)             -9.610            -3.689            -2.975            -2.619
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

```
. regress  gdpinnok_log3 L4.rdexpenditure_log3 L1.rdstaffinnorway_log3 L1.produ
> ctionandvalue_log3 L4.consuminhouseholdsandidealo_log3

      Source |       SS       df       MS              Number of obs =      35
-------------+------------------------------           F(  4,    30) =    8.13
       Model |  .16862973     4  .042157433           Prob > F      =  0.0001
    Residual |  .15553325    30  .005184442           R-squared     =  0.5202
-------------+------------------------------           Adj R-squared =  0.4562
       Total |  .32416298    34  .009534205           Root MSE      =    .072

------------------------------------------------------------------------------
gdpinnok_l~3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rdexpendit~3 |
         L4. |  -.3635198   .1605916    -2.26   0.031    -.6914916    -.035548
             |
rdstaffinn~3 |
         L1. |  -.7076267   .3749453    -1.89   0.069    -1.473367    .0581137
             |
production~3 |
         L1. |  -.5368502   .1643594    -3.27   0.003    -.8725169   -.2011835
             |
consuminho~3 |
         L4. |   .6228628   .3545355     1.76   0.089    -.1011954    1.346921
             |
       _cons |   .0031249   .0122129     0.26   0.800    -.0218172     .028067
------------------------------------------------------------------------------
```