



**Copenhagen
Business School**
HANDELSHØJSKOLEN

Real Economy - Real Weights

Forecasting GDP with fundamental
indexing

June

2012

Copenhagen Business School
- Applied Economics & Finance

Candidates

Tom Rune Bjanes
Magnus Vegem Dahle

Counsellor

Lisbeth Funding La Cour Department of Economics

Number of pages: 115
Number of characters: 178 456

EXECUTIVE SUMMARY

The lately observed distress in macroeconomic conditions around the world has shown that fiscal and monetary policymakers would benefit from an early and accurate prediction of changes in economic activity. Per today GDP is a commonly used proxy for this. This paper attempts to provide this through stock indices weighted on various fundamental data, rather than the usual market capitalization weighting. The intuition behind this is that an index weighted on fundamental data is more representative of the real economy than an index weighted on market capitalization. In general there are few leading indicators for the Norwegian GDP and our methodology has not been tested for the Norwegian GDP or internationally.

In this thesis we start by presenting fundamental indexing as pioneered by the research agency Research Affiliates. Following is the basics of Gross Domestic Product and its shortcomings, as well as an overview of leading indicators in Norway. Next we perform a thorough analysis of the developed fundamental indices (FI), and the different comparable indicators. We found that leading indicators based on fundamental metrics in general perform well compared to the peer indicators chosen in this thesis. Among all the 14 indicators examined, the fundamentally based indices are ranked 2-4 and 6-8 when ranked by both Root Mean Squared Errors and Root Mean Squared Forecasting Errors.

The thesis concludes that leading indicators based on the methods presented in this paper are well suited as single leading indicators for Norwegian GDP. However, only three out of six FI based indicators perform better than the indicator based on the market capitalization weighted OBX index. The FI based indicators that performed best includes averages of different fundamental data. We believe that for this methodology to perform better than a regular market capitalization based indicator, the Fundamental Indices must be weighted in such a way that it represents the companies in more than one fundamental metric.

Recognition

We would like to say a special thanks to our counsellor Lisbeth Funding La Cour who has been exceptional in her guidance, remarks, and discussions.

We would also like to thank everybody that has helped us with proofreading, discussions, and other practicalities, and also each other.

Table of Contents

1	Introduction.....	1
1.1	Purpose of the thesis.....	1
1.2	Problem Statement	2
1.2.1	Hypothesis:	3
1.3	Methodology overview.....	4
1.4	Limits of the thesis.....	5
1.4.1	Limitation of goals	5
1.4.2	Limitations in the econometrics	5
1.4.3	Limitations in data	5
1.5	Structure of the thesis	6
2	Background.....	7
2.1	Period description	7
2.2	Gross domestic Product.....	10
2.2.1	The Norwegian GDP.....	11
2.2.2	Seasonality in GDP	11
2.2.3	Seasonality adjustments.....	12
2.2.4	Shortcomings of GDP	14
2.2.5	Corrections to GDP	16
2.3	Why asset prices should have an important role in the Norwegian economy	17
2.3.1	Stocks' expectations of the real economy.....	22
2.4	The OBX Index	23
2.5	Fundamental indexing.....	24
3	Literature review	28
3.1	International research	29
3.1.1	Bloomberg Financial Conditions Index for the US.....	29
3.1.2	Citi Financial Conditions Index.....	29
3.1.3	Deutsche Bank Financial Conditions Index.....	30
3.1.4	Federal Reserve Bank of Kansas City Financial Stress Index.....	30

3.1.5	Macroeconomic Advisers Monetary and Financial Conditions Index	31
3.1.6	OECD Financial Conditions Index.....	31
3.1.7	The OECD Composite Leading Indicators	32
3.1.8	The National Bureau of Economic Research	33
3.2	Norwegian research	34
3.2.1	Business cycles in Norway	34
3.2.2	Historical indicators as predictors of banking crises	35
3.2.3	Financial figures and the real economy.....	37
4	Methodology	38
4.1	The OLS assumptions.....	38
4.2	Testing for Model specification and wrong functional form.....	38
4.2.1	Normality	38
4.2.2	Autocorrelation	42
4.2.3	Multicollinearity	43
4.2.4	Heteroscedasticity	44
4.2.5	Stationarity	45
4.2.6	Ramsey`s RESET test	46
4.3	Parameter Stability	46
4.3.1	The Chow test.....	46
4.4	Causality test	48
4.4.1	Granger Causality test	48
4.5	Box-Jenkins modeling	51
4.6	Dynamic regression models.....	51
4.6.1	Intermediate and long term effects in dynamic models	55
4.7	Model specification	56
4.8	Model misspecification.....	58
4.9	Errors of measurement	59
4.9.1	Dependent variable	59
4.9.2	Explanatory variables	60
4.9.3	Underfitting and overfitting	61
4.10	Model selection and comparison:	63

4.11	Constructing the fundamental indices	64
4.11.1	Special notes.....	65
4.12	Data	66
4.12.1	Data gathering	66
4.12.2	Data transformation	67
4.12.3	Flaws in the data.....	67
5	Comparable Leading indicators	68
5.1	Single indicators.....	68
5.1.1	OBX index	68
5.1.2	Interest rate spread	68
5.1.3	Money supply	69
5.1.4	Credit indicator	69
5.1.5	House Price Index	70
5.2	Comparable indices	70
5.2.1	Norwegian Purchasing Manager Index.....	70
5.2.2	OECD Composite Leading Indicators	71
6	Results	72
6.1	In sample predictions	72
6.1.1	Our fundamental indices	72
6.1.2	Model comparison.....	76
6.1.3	Comparable leading indicators.....	76
6.2	Comparison of our models to a Naïve model.....	84
6.3	Out of Sample	85
6.3.1	Our fundamental indices	86
6.3.2	Comparable Indices	90
6.4	Comparison of our models to a naïve model	96
6.5	Structural breaks	97
6.6	Granger causality tests	100
6.6.1	Fundamental Indices	100
6.6.2	Comparable leading Indices	101
7	Discussion	103

7.1	Method and models	103
7.2	Hypothesis 1	105
7.3	Hypothesis 2	106
8	Conclusion	108
8.1	Further research	110
9	Bibliography.....	I
	Appendix.....	V
A	In-sample autocorrelation.....	V
B	In-sample normality.....	VII
C	In-sample Ramsey's RESET test	IX

1 Introduction

The financial crisis of 2008, which later developed into the euro crisis, has not just significantly impacted many countries' national economies, but also shown that that government intervention is both utterly important and necessary. Changes in fiscal or monetary policy do not always have an immediate impact on the real economy, hence the earlier the policymakers can get an indication of future national economic situation the better. The gross domestic product (GDP) is an important and frequently used measure of a country's economic state. A timely and accurate indicator of future GDP is therefore highly desirable. There is a wide range of research on leading indicators trying to predict both GDP, and its turning points. However, the number of leading indicators for the Norwegian GDP is not comprehensive, and any serious attempt to predict it would be welcomed. Everybody, from the man on the street to high ranking government officials, look to stock markets, like the Oslo Stock Exchange (OSE), and their indices to get an impression of the current economic state. This because they are expected to give an indication of how the economy is performing, as it contains the total expectation of the future economy of the market participants (Nette, 2010). In many ways, the stock market's ability to immediately address for instance news and changes in the economic environment can make it a good indicator of what is to come, whereas GDP describes what actually happened.

We argue that a leading indicator based on a stock index weighted on companies' fundamental data will be a better predictor of GDP than one based on market capitalization weights. This is because we expect that fundamental data will reflect more of the real economic aspects.

1.1 Purpose of the thesis

The main purpose of this Master Thesis is to investigate whether or not a fundamentally weighted stock index (FI) is able to predict future changes in Norwegian GDP. Furthermore we will investigate whether a fundamentally weighted index is better at predicting changes in GDP than a market capitalization (MCAP) weighted index. We have used the OBX-index, which will be introduced shortly, as a representative for an MCAP weighted index.

The GDP is primarily used as an indicator for the health of a nation's economy. Policymakers use GDP to assess the current economic situation, and to adjust fiscal and monetary policies to ensure healthy future GDP growth. There is a general consensus that a yearly increase of 2.5-3.5% in real GDP is the most beneficiary. It is seen as high enough for healthy growth in corporate profits and employment, but low enough to prevent unwanted high inflation¹. A decrease in GDP can be viewed as a sign of decreasing corporate investments, employment, or household consumption, which can lead to a vicious circle with even less corporate investments, more unemployment and less household consumption. Policymakers adjust monetary and fiscal policy, like interest rate, taxes, and government consumption in order to maintain a healthy GDP growth. A good forecast of future GDP can help policymakers make early adjustments to the policies and prepare for larger changes in advance of changes in the economy.

As mentioned earlier, there are few leading indicators created for the Norwegian economy. This thesis attempts to fill some of this gap, and thereby give Norwegian policymakers a broader platform to base their decisions on.

1.2 Problem Statement

Through our studies financial markets, asset allocation, and their link to macroeconomic conditions have become a key interest. Our approach to a leading indicator has never been researched before, neither for Norway or internationally. Further, there are in general few leading indicators for the Norwegian economy, hence our problem statement is one of general interest;

Is a fundamentally weighted stock market index a good base for a leading indicator of the Norwegian economy, and does it outperform a regular market capitalization weighted index in this regard?

We wish to investigate fundamental indices, and their ability to indicate the welfare growth of Norway. Most indices are weighted by company market capitalization, and we argue that because of this they do not give a good picture of a country's economic prosperity. An index weighted on

¹ Ryan Barnes, <http://www.investopedia.com/university/releases/gdp.asp#axzz1uZKkiMcE>.

fundamental data, like number of employees or sales, would be better suited to reflect the real economy, represented by GDP. We will use the OBX index as a base of our fundamental indicators and analyze how the movement of the index will change when weighted on different fundamental data. We will then benchmark our indicators to other known indicators of GDP, how the original OBX perform, and a naïve model.

1.2.1 Hypothesis:

This leads to the following two hypotheses:

H_0^1 : *A fundamentally weighted stock index is better than a market capitalization weighted stock index as a leading indicator for Norwegian GDP*

H_A^1 : *A fundamentally weighted stock index is not better than a market capitalization weighted stock index as a leading indicator for Norwegian GDP*

and

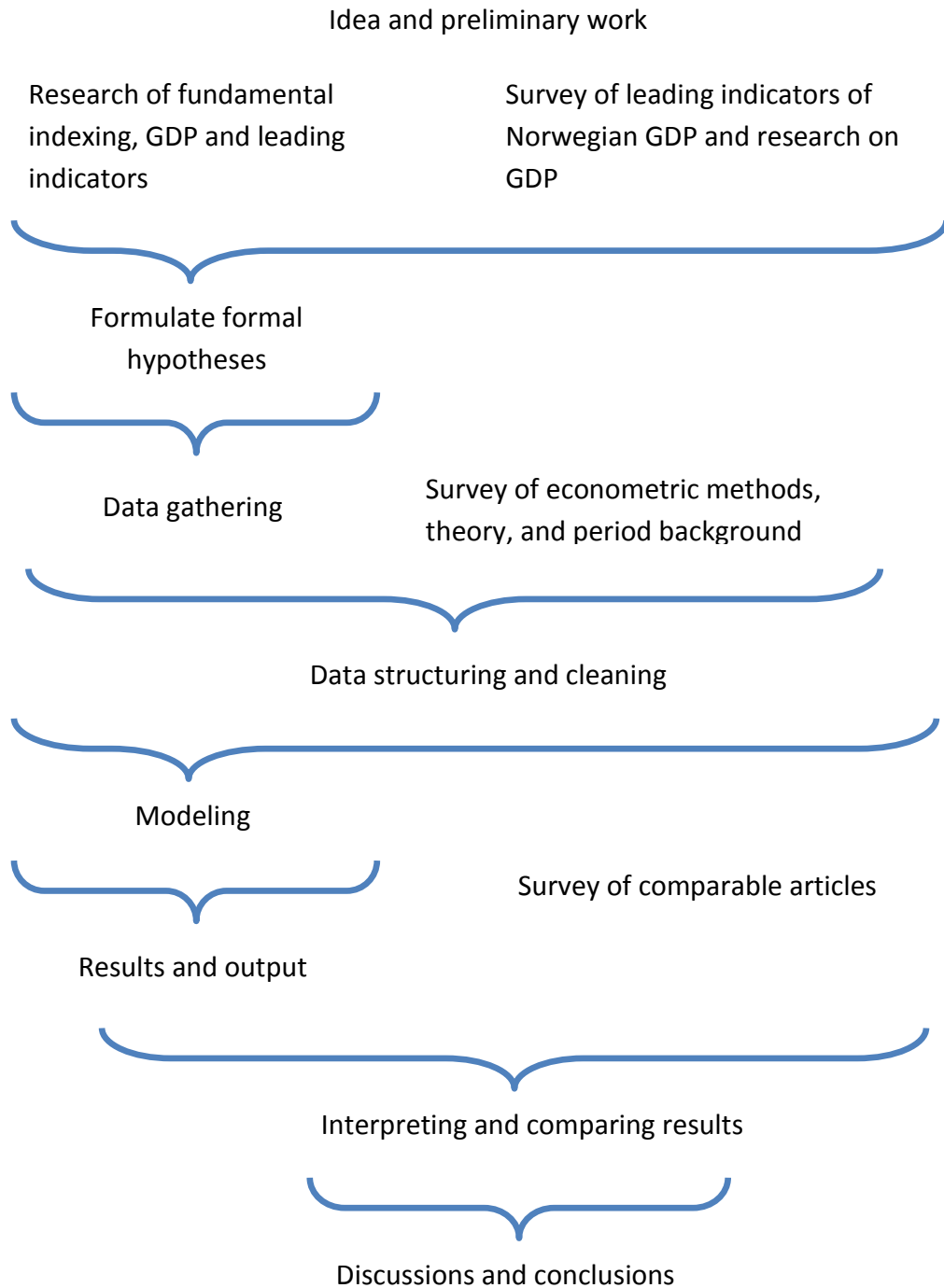
H_0^2 : *A fundamentally weighted stock index is a good indicator for Norwegian GDP*

H_A^2 : *A fundamentally weighted stock index is not a good indicator for Norwegian GDP*

1.3 Methodology overview

Here, we will give a quick overview of our working process:

Figure 1.3-1



1.4 Limits of the thesis

1.4.1 Limitation of goals

We focus on making the best possible leading indicator for Norwegian policymakers, in order for them to adjust policies as early as possible, not necessarily forecasting GDP per se. This focus has limited us in a number of ways; number of companies included, number of countries included, different frequencies, and so on.

1.4.2 Limitations in the econometrics

There are a number of econometric tests and theories not accounted for in this thesis. We have mentioned some, like the possible correction for outliers with dummy variables, but we will not use this in our main models. We want to see how our indicators perform when we let the events fold out unopposed. This is of course discussable as not correcting for outlier may give poorer results in general, but better in a new crisis. We chose not to correct for outlier as we wanted to avoid "reality mining." In the words of Peter Kennedy, *"Economists' search for 'truth' has over the years given rise to the view that economists are people searching in a dark room for a non-existing black cat; econometricians are regularly accused of finding one."*²

We will perform a Granger causality test, but not go into deep discussion of the possible exogeneity in this thesis. The possibility of a spurious correlation between GDP and other factors that involve FI index or market capitalization will not be investigated in depth.

1.4.3 Limitations in data

We also had to limit our thesis in regards to data gathering and quality checking. We have assumed the data from Datastream, and other sources, to be correct. If we could not find the appropriate data through searches in this database we tried to find annual reports, but as companies in Norway are not obligated by law to save these for more than ten year, information on some companies are lacking.

² Peter Kennedy, 1992, *A guide to Econometrics*. Third edition, The MIT Press, Cambridge, Mass, pp. 82.

1.5 Structure of the thesis

This thesis is divided into 8 sections, and the following are; (2) a background description of the period we are examining, GDP, fundamental indexing and the OBX index are also explained. Then in (3) we introduce the reader to research done on leading indicators, and other indicators used in practice. Section (4) explains the methodology used in this thesis. The focus is on modeling and econometrics, but we also address how our fundamental indices are created. In section (5) we present the peers, or comparable indicators, as we will refer to them, that is used and defined by this thesis. Then in section (6) we present the results of our analyses. In section (7) we discuss and elaborate on the results, and how they apply to our hypotheses. Section (8) contains our conclusion. Here we sum up our findings and what they imply. We round off by suggesting topics of further research.

2 Background

2.1 Period description

To better understand the financial condition during the period of 1996-2011 we will now present a short summary of the period derived from Statistics Norway's (SSB) semi-annual reports from the period called "Norsk Økonomi (Norwegian Economy)."³

From **1996** to **1997** Norway had a steady increase in GDP each quarter, but experienced some reduction in the petroleum industry's contribution. The school system was changed in Norway in this period, so children starts one year earlier, this change increased government spending, and thus increased GDP.

1998, The Asian financial crisis stagnates the export to Asia, and the industry production slows down. The crisis is clearly visible at the stock exchanges. Though the Asian crisis has an impact on the European economy, Norway increased their export to the area, hence Norwegian GDP suffered only a minor decrease and remained stable through the year. In **1999** the GDP growth rate is lower than the year before, and policymakers start to see the end of the economic upturn Norway had seen since 1993.

Graph 2.1-1



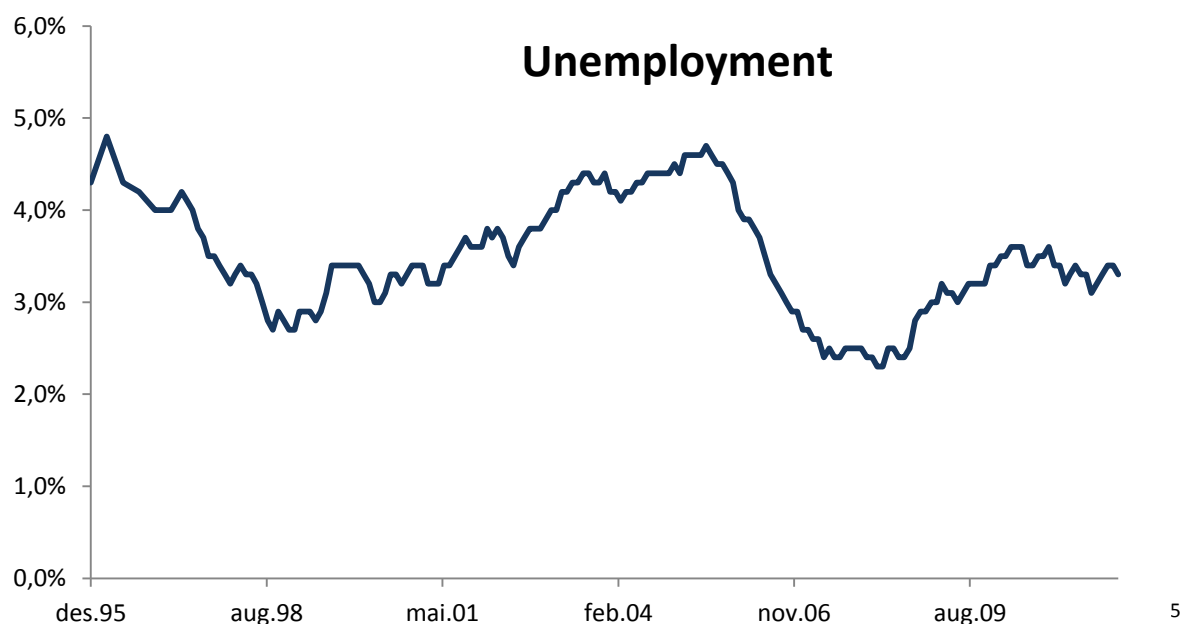
³ The reports are found here: http://www.ssb.no/oa/tidligere_utgaver.html

In **2001** the Norwegian parliament adopted a guideline regarding how to use the income from the petroleum industry. They also set a target inflation rate of 2.5 %. The IT-bubble burst in March 2001 started a downturn in the Norwegian economy that lasted until 2003.

2003 marked the end of one business cycle and the start of another. The turn in 2003 was mainly due to large investments in the petroleum industry. In **2004** SSB recognizes growth, but not at the same pace as in the US. In **2005** SSB expected the upturn to stop in 2006, but tax reductions during the year had a positive impact on the economy and they adjusted their expectation to 2007 or 2008.

In **2006** the consumer price index (CPI) went up, mainly because of high electricity prices, due to low water reserves in hydro plants. In **2007** there was a sharp increase in demand, the lowest unemployment rate in 20 years, and household consume increased due to a cut in car taxes.

Graph 2.1-2



⁴ Numbers from http://statbank.ssb.no/statistikkbanken/default_fr.asp?PLanguage=1 > 09 National economy and external trade > 09.01 National accounts/National accounts, quarterly/09177: Exports of goods and services. Unadjusted and seasonally adjusted figures (1978K1-2012K1)

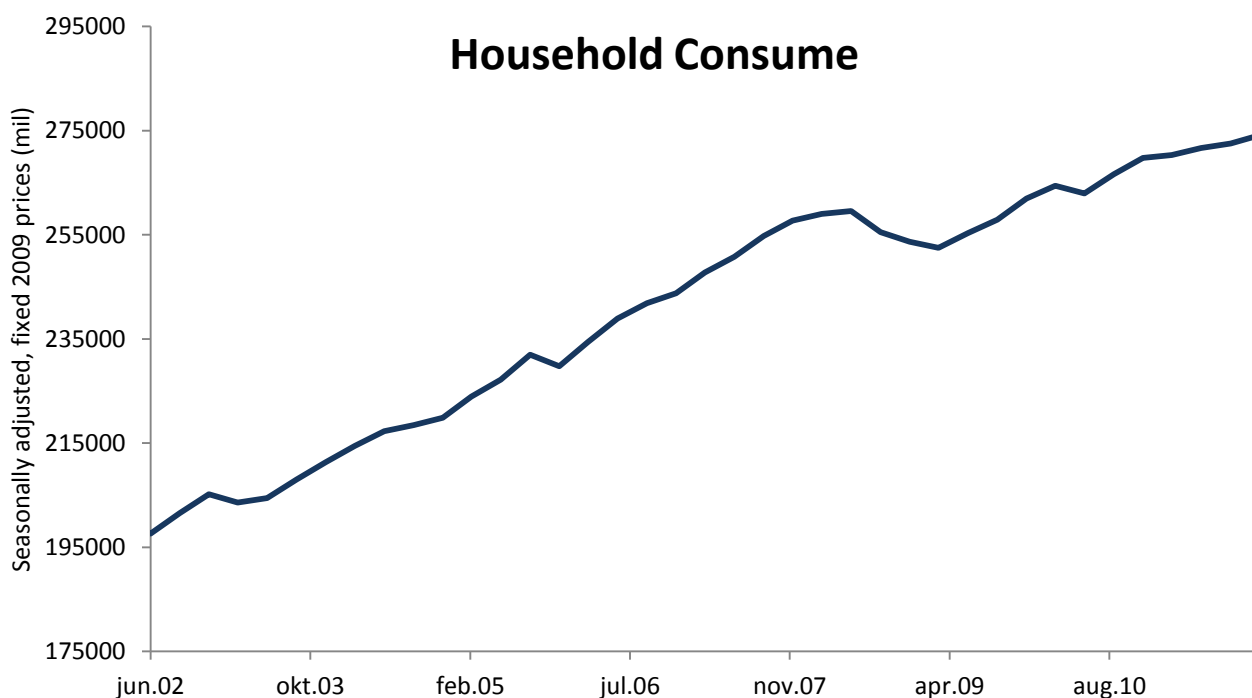
⁵ Numbers from:

http://statbank.ssb.no/statistikkbanken/Default_FR.asp?PXSid=0&nvl=true&PLanguage=0&tilside=selecttable/hovedtabellHjem.asp&KortnavnWeb=aku

As we know, in **2008** all of this turned. High interest rates and a high Norwegian Krone made it hard for exporters. But, due to investments in the petroleum industry, and increased government spending, the Norwegian economy got hit less severe by the financial crisis than most other countries.

In **2009** the unemployment increased heavily, but from an already low level. There was still a strong downturn in the economy, but expansive financial policies helped this to turn in **2010**. The turn in 2010 is still not an accepted fact, as it still is high uncertainty of the future because of the ongoing crises in the euro area. In **2011** the Oslo Stock Exchange has recovered from the bottom in 2008/2009, and Norway sees a decrease in unemployment. However, a low increase in productivity gives mixed signals of whether Norway is back in an economic upturn or not.

Graph 2.1-3



6

⁶ Numbers from http://statbank.ssb.no/statistikkbanken/default_fr.asp?PLanguage=1 > 09 National economy and external trade > 09.01 National accounts/National accounts, quarterly/ 09173: Final consumption expenditure of households. Unadjusted and seasonally adjusted figures (1978K1-2012K1)

2.2 Gross domestic Product

Gross domestic product is both used as an indicator of total value added in a country, as well as an expression for gross income generated by domestic production. GDP is measured in market prices and can be compiled with three different approaches; the production approach, the expenditure approach, and the income approach.

UN System National Accounts (SNA) (2008)⁷ defines the three different approaches in the following way:

- *GDP is the sum of gross value added of all resident producer units plus that part (possibly the total) of taxes on products, less subsidies on products, that is not included in the valuation of output.*
- *GDP is equal to the sum of the final uses of goods and services (all uses except intermediate consumption) measured at purchasers' prices, less the value of imports of goods and services.*
- *GDP is equal to the sum of primary incomes distributed by resident producer units.*

What seems to be the most popular way to describe how GDP is calculated is the expenditure approach, and it goes as follows:

Equation 2.2-1

$$GPD = C + G + I + NX$$

Where:

C = Private domestic consumption

G = Domestic government spending

I = Domestic investments

NX = Net exports (exports – imports)

⁷ European Communities, International Monetary Fund, Organization for Economic Co-operation and Development(OECD), United Nations, and World Bank, 2008, *UN System National Accounts*.

However, Statistics Norway defines GDP as the following⁸:

GDP is measured in market value and is defined as the sum of the gross product across all industries measured in basic values, in addition to all products taxes and subtracted all product subsidies.

2.2.1 The Norwegian GDP

Norway separates between total GDP and Mainland GDP. Mainland GDP is a measure of value creation in Norway without the income from the petroleum and international shipping sectors. When policymakers analyze the economic development, they usually use Mainland GDP⁹. Despite this, our index includes companies in both the petroleum and shipping sector and we will use the total GDP measure in this thesis.

The quarterly GDP is published about 50 days after the end of the quarter, hence the first quarter numbers are published in May and so on. What makes the GDP so difficult to forecast is that the GDP numbers are revisited and revised if something changes in one of the variables and so on, meaning that old GDP numbers are corrected. Because some historic GDP numbers later on are corrected, and done so in a varying degree, the models are estimated on both corrected and non-corrected numbers. This raises the question; are we estimating using the revised value of future GDP or the GDP as it first is published.

2.2.2 Seasonality in GDP

In half of the indicators that we use to compare how well our FI based indicators perform we find two significant lags of GDP, the second and the fourth, when performing our regressions. Why the third lag is not significant puzzle us as we use seasonally adjusted GDP. It might be that the other explanatory variables contain some seasonality that makes the fourth lag of GDP significant, but we have not investigated this further, as the models pass our specification test.

Graph 2.2-1 shows both non-adjusted and seasonally adjusted GDP, where the adjusted GDP growth is the dependent variable in this thesis. As you can see, there is a clear difference in the two lines.

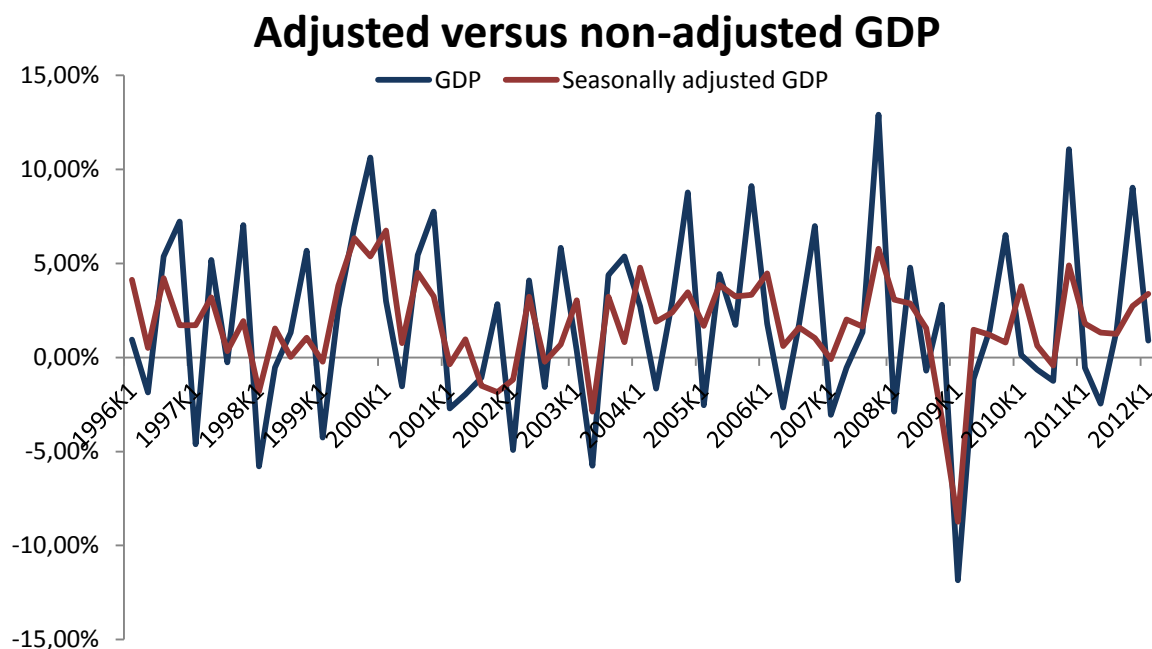
⁸ Statistics Norway, <http://www.ssb.no/metadata/conceptvariable/vardok/1743/en>

⁹ Finansdepartementet (Ministry of Finance)

http://www.ungokonomi.no/index.php?option=com_content&task=view&id=12&Itemid=16

Almost all the peaks in the non-adjusted GDP are fourth quarter growth, and we will now elaborate on how these seasonality adjustments are done, and why it is important.

Graph 2.2-1



10

2.2.3 Seasonality adjustments

In economic time series there can be large variations in the dataset because of seasonal variations. These types of season effects could be the increase in household consumption in November and December because of Christmas shopping, and the lower production in June and July because of the summer vacations. To say that the household consumption is much higher in November/December compared to the previous months is true, but a much sounder comparison would be to compare the consumption with the same quarter the previous year. That is why we have to seasonally adjust the data if we want to look at the development through the year.

When SSB seasonally adjust their data they take into account the number of working days in the period. An example of this is when Easter falls, if it is in April in one year and March the next this

¹⁰ Own calculations - Numbers from http://statbank.ssb.no//statistikbanken/default_fr.asp?PLanguage=1 > 09 National economy and external trade > 09.01 National accounts/National accounts, quarterly/09186: Final consumption expenditure. Unadjusted and seasonally adjusted figures (1978K1-2012K1)

needs to be accounted for. Also, the numbers needs to be adjusted for national holidays that fall on a weekend or a working day for different years.

When these moving events and other exceptional outliers, like labor conflicts or natural disasters, are taken care of, SSB uses a season adjustment program called X-12-ARIMA (Statistisk Sentral Byrå, 2012). This program was developed by the US Consensus Bureau, and is widely used as a seasonal adjustment tool¹¹. When correcting the data before the seasonality adjustments, SSB group the outliers into three categories;

- Extreme outlier; an extreme value that is apparent one period but gone the next.
- Shifts; in for example production. These shifts that make the time series permanently above or below the previous levels. Can be caused by for instance new technologies implemented in an industry.
- Temporary shifts. This is a cross-over of the points above. It is a shift caused by a shock, and the effects are slowly reverting towards its normal values.

When the pre-correcting of the time series is done, one can usually decompose the variation in the time series as the sum or product of three effects:

- Season effects component, S
- Trend effect component, T
- Irregular movement component, I

The S component is then removed from the time series and the seasonally adjusted series, A, will be

Equation 2.2-2

$$A = T + I.$$

SSB use a wide range of qualitative indicators to check if the adjustments are good enough. Among them is to measure the effect of changes in working days, the stability of the season and trend effect components, and the magnitude of variation in the data (Statistisk Sentral Byrå, 2012).

¹¹ United States Census Bureau, <http://www.census.gov/srd/www/x12a/>.

2.2.4 Shortcomings of GDP

GDP as a socio-economic indicator has been criticized for almost as long as the concept is old. It is obvious that reducing all social and economic data into one number is difficult, maybe even impossible.

From an economic welfare perspective there are four main limitations to how the GDP is measured:

- Many household activities that are productive in an economic sense are excluded.
This is for example; housekeeping, caring of the sick, transportation of household members, and so on. SNA argues that these activities mainly contribute within the household and does not have a large impact on the rest of the economy. Since it is not meant for the market it is difficult to price these contributions.
- The inclusion of non-market activities, like different government spending.
To include these values makes sense because they definitely create welfare, but they are also subject to some double counting of production. It is counted both at final consumption, and government production cost. Government goods and services are valued at its production costs, excluding consumer utility of the product.
- The GDP is an aggregate measure.
Research shows that wealth and financial inequality has an inverse relation with a nation's overall happiness and trust (Oishi, et al., 2011). Distribution of resources is crucial to welfare, but GDP does not take this into account.

- GDP measure flows, not stock, of wealth in an economy

The stock of economic and natural resources have a large impact on GDP. Even though the stock does not change, the price of them can change considerably and thereby change the nation's ability to consume. Also, it does not take into account changes of stock due to e.g. natural disasters, wars, or finding of new natural resources. Here is a quote by Oliver Vaury too show how extreme it could be: "Burn Paris and you will make GDP grow!"¹²

(Commision on the Measurement of Economic Performance and social Progress, 2008)

As with most empirical data GDP are subject to measurement and estimating errors. When calculating quarterly GDP Statistics Norway separates these sources of error into four groups:

1. Missing or incomplete data

Parts of the economy is not covered by short term statistics or other updated information. Particularly uncertain data is investments in services and production in non-profit entities.

2. Partly relevant data

Short-term statistics and information sources may provide a variable that does not completely cover what the indicator is meant to represent. E.g. SSB may use a statistic for turnover as a proxy for production.

3. Uncertainty or errors in the underlying statistical data

This can be sampling errors, measurement errors, data handling errors, modeling errors, and so on

4. Errors when calculating GDP and the quarterly national accounts.

¹² Olivier Vaury, *Is GDP a good measure of economic progress?* *Post-autistic economics review*, issue no. 20, 3 June 2003, article 3, <http://www.paecon.net/PAEReview/issue20/Vaury20.htm>

Even though SSB have identified these sources of errors, they remain confident that the statistics provided by them are fairly accurate. This view seems to be shared by the International Monetary Fund (IMF) in their report on Norwegian macroeconomic statistics:

“In summary, Norway’s macroeconomic statistics are of generally high quality. They are adequate to conduct effective surveillance, although the mission identified some shortcomings that may detract from the accurate and timely analysis of economic and financial developments and the formulation of appropriate policy. These shortcomings include deficiencies in the scope, periodicity, and timeliness of government finance statistics and evolving weaknesses in the source data for balance of payments statistics.”¹³

2.2.5 Corrections to GDP

Despite its limitations as a general welfare measure GDP remains the most widely used proxy for a nation’s well being. It is however not the sole index proposed by SNA. Some incomes generated domestically might be sent abroad, and vice versa. The Gross National Income (GNI) measurement is proposed by the SNA to correct for this. It is defined as “...the sum of gross primary incomes receivable by resident institutional units or sectors.”¹⁴ This measure and others that SNA propose are all economic measures. As a proxy for welfare one should also take into account other welfare criteria like health and education. There have been several attempts, and some have indeed succeeded, to create a better figure or figures for measuring welfare. Legatums Prosperity Index, which uses eight different sub-indices which is identified as contributing to welfare and life satisfaction, and UNs Human Development Index, which uses three of sub-indices, is examples on welfare indices that take more social well-being proxies into account.

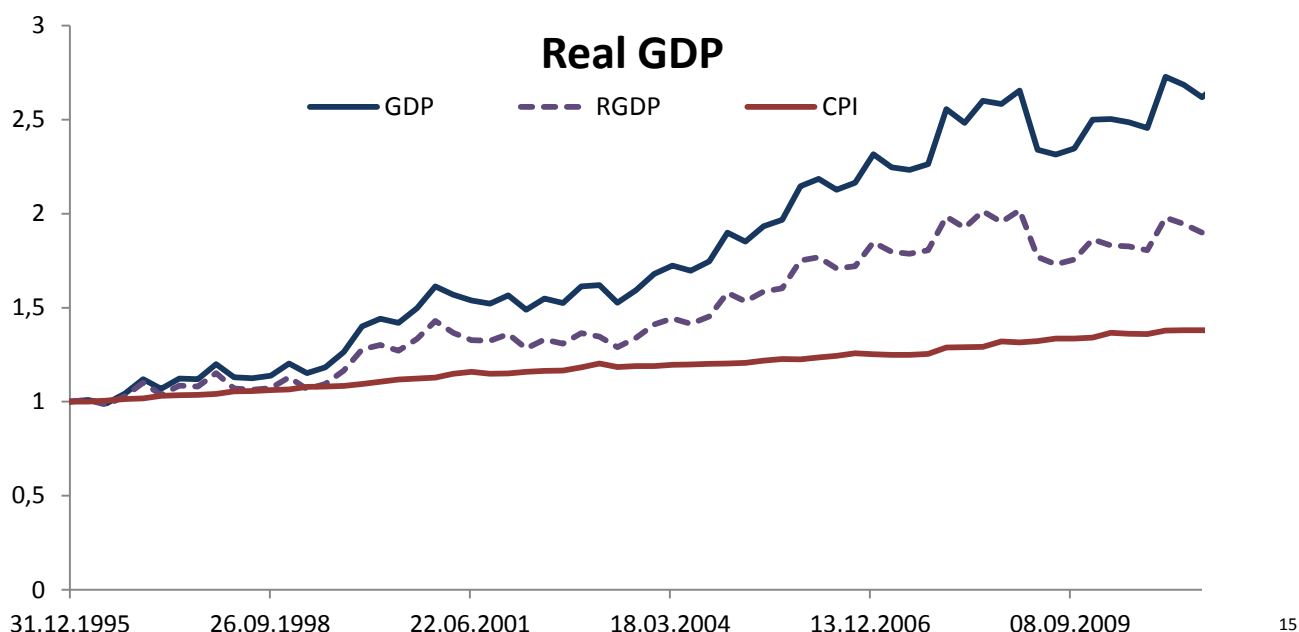
The most used correction of GDP is real gross domestic product (RGDP). RGDP is an inflation/deflation-adjusted version of GDP. The inflation is usually measured using the Consumer

¹³ International Monetary Fund, July 2003, *IMF Country Report No, 03/207*.

¹⁴ European Communities, International Monetary Fund, Organization for Economic Co-operation and Development(OECD), United Nations, and World Bank, 2008, *UN System National Accounts*.

Price Index (CPI). The graph below shows how the inflation slows the growth rate of RGDP relative to GDP.

Graph 2.2-2



Though GDP falls short in many areas, it is the most used measure for economic activity and welfare. The fact that neither the stock prices nor the regular GDP are inflation adjusted, and both are based on market values speak for our intuition; a market price based index could forecast GDP if weighted correctly.

2.3 Why asset prices should have an important role in the Norwegian economy

Aastveit and Trovik (2008) wrote an interesting paper on the role of asset prices in a small open economy like that of Norway. They find that the single most important block of data to improve estimates of current quarter GDP in Norway is asset prices. They also state that *“the strong impact from financial data is due to an ability of the market clearing process to impart information about the real activity in Norway in a timely manner.”* We find this a useful insight, even though Aastveit and

¹⁵ Own calculations – Numbers from <http://www.ssb.no/kpi/tab-01.html> and http://statbank.ssb.no/statistikkbanken/default_fr.asp?PLanguage=1 > 09 National economy and external trade > 09.01. National accounts/National accounts, quarterly/09186: Final consumption expenditure. Unadjusted and seasonally adjusted figures (1978K1-2012K1)

Trovik (2008) use this for a nowcasting purpose. A nowcast is a description of present conditions, and a forecast of those immediately expected, whereas we want to forecast one quarter ahead (Aastveit & Trovik, 2008). Some previous research, see for example Stock and Watson (2001)¹⁶, indicate that financial assets, when used as single indicators, are unstable when forecasting across different time periods, but usually they improve when combined.

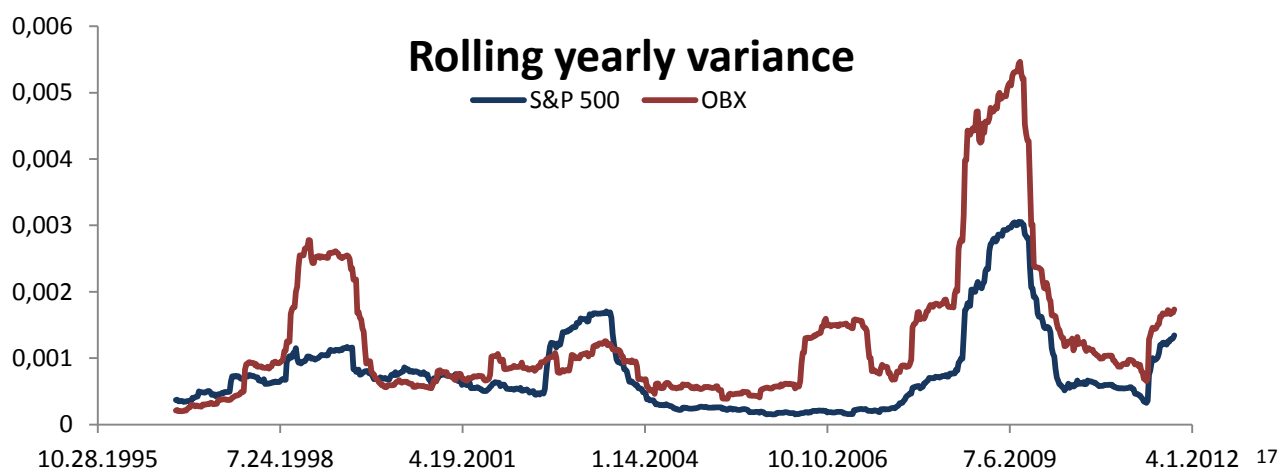
Aastveit and Trovik (2008) presents how changes in asset returns may have a signaling effect for changes in real activity in the same quarter, i.e., almost an immediate impact to real activity. They also mention an additional advantage with asset prices, namely their timeliness. Relevant news about current business activity is immediately reflected in the prices through the market clearing process. (Aastveit & Trovik, 2008)

The Norwegian economy has some obvious differences from the US: it is small and open, and Norway is a large petroleum exporter. The American economy is the opposite. One can assume that a small and open economy will experience an instant effect on the real economy if exposed to shocks like labor conflicts, large contracts won by a domestic company, or change in foreign labor supply. Therefore it is reasonable to assume that both trade shocks and other productivity/technology shocks play a more prominent role in the Norwegian economy. This is why we expect that timely news about such shocks, as reflected in exchange rates and equity returns, should be more important for Norway than the US. The fact that both GDP growth and returns on the Oslo Stock Exchange are more volatile than similar measures for the US is in line with such an assumption. Petroleum activities are capital intensive and carry substantial fixed costs; hence it is not surprising that the price of oil and the petroleum related content on the exchange do not have an immediate effect on real activity (Aastveit & Trovik, 2008). Because of this immediate reaction to economic changes in the Norwegian stock market, we use the current one quarter return of the FIs to predict next quarters change in GDP, which is somewhat in line with the nowcasts from Aastveit and Trovik (2008).

¹⁶ James H. Stock and Mark W. Watson, 2001. *Forecasting Output and Inflation: The Role of Asset Prices*.

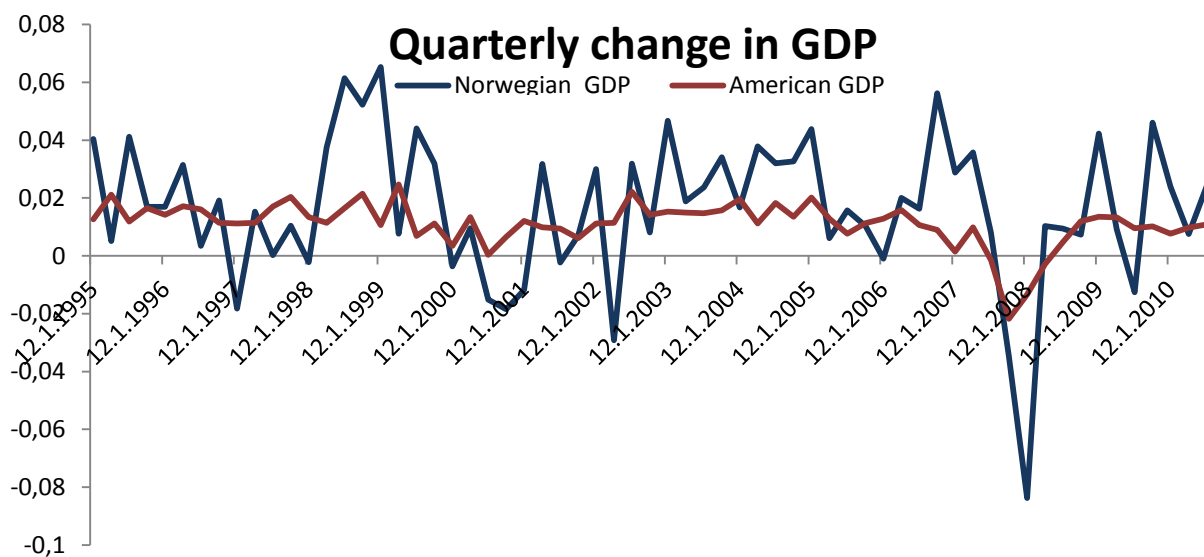
The volatility of the OBX is, as we can see in *Graph 2.3-1*, most of the time higher than the comparable variance for the American S&P 500 index which we chose to use as its peer. The S&P 500 are often used as a benchmark for US equity performance.

Graph 2.3-1



Graph 2.3-2 present the high volatility exhibited in the Norwegian GDP compared to the American GDP, it should be noted that the numbers are not exchange rate adjusted.

Graph 2.3-2

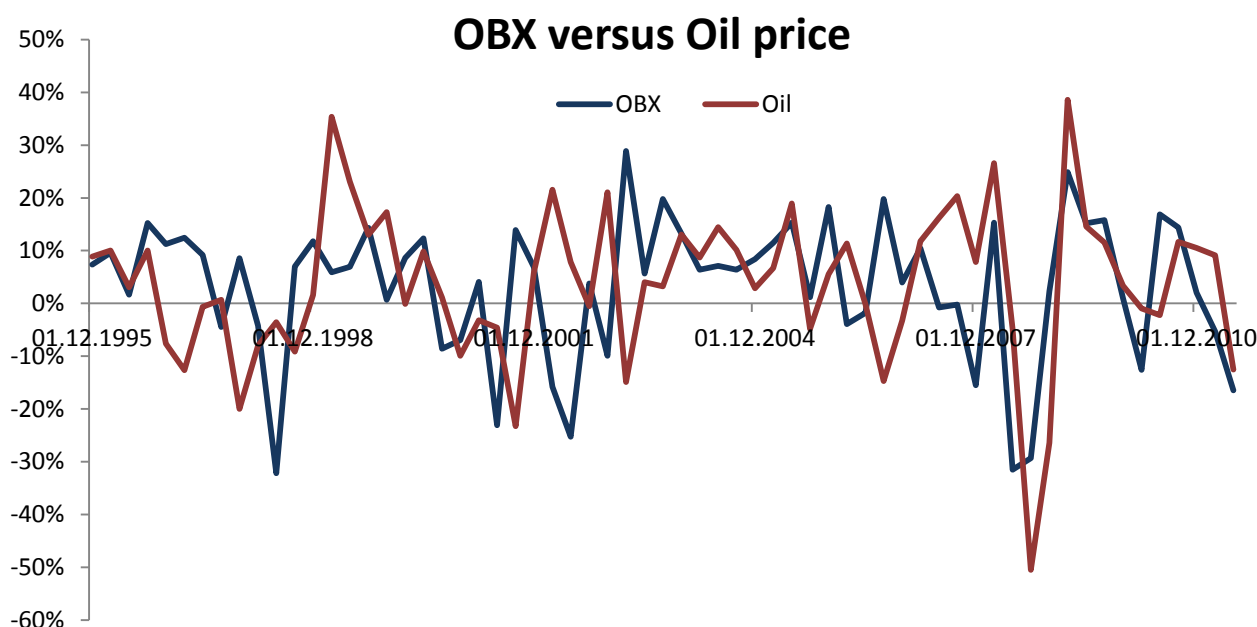


18

¹⁷ Own calculations – Numbers from; S&P 500: Datastream, OBX: provided by Oslo Stock Exchange

The Oslo Stock Exchange covered about 55 per cent of the real activity in the Norwegian economy in 2008 (Aastveit & Trovik, 2008). This coverage suggest, in accordance with our intuition, that the stock exchange reflects what market beliefs of large parts of the Norwegian economy. The OSE are often said to be an “oil-exchange.” Despite some periods of negative correlation, it is mostly positively correlated. You can see a graphical evidence below.

Graph 2.3-3



19

It can also be seen in the correlation coefficient between the percentage change in the OBX and in the oil price:

¹⁸ Own calculation – Numbers from; American GDP: <http://www.bea.gov/national/index.htm#gdp> , Norwegian GDP: http://statbank.ssb.no/statistikkbanken/default_fr.asp?PLanguage=1 > 09 National economy and external trade > 09.01 National accounts/National accounts, quarterly/09186: Final consumption expenditure. Unadjusted and seasonally adjusted figures (1978K1-2012K1).

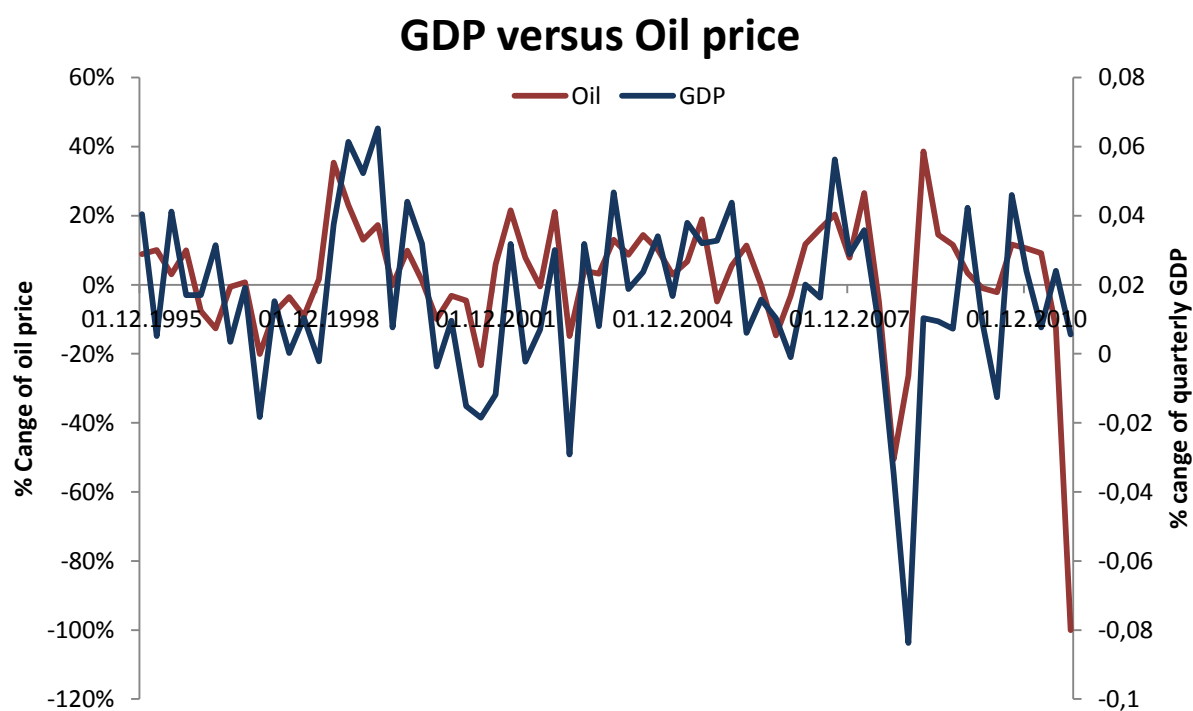
¹⁹ Own calculations – Source: OBX was provided by Oslo Stock Exchange and Oil prices were found using Datastream.

Table 2.3-1

	Oil	NOR GDP	US GDP	S&P 500	OBX
Oil	1				
NOR GDP	0.6349	1			
US GDP	0.3528	0.4901	1		
S&P 500	0.1607	0.1423	0.4054	1	
OBX	0.2393	0.1320	0.3824	0.7803	1

²⁰

Graph 2.3-4

²¹

It should be noted that in both the graphs above the oil price is in USD, so for the Norwegian economy there will be an exchange rate risk associated with fluctuations in the oil price.

²⁰ Own calculations – Numbers from: US GDP: <http://www.bea.gov/national/index.htm#gdp>, Oil Price and S&P 500: Datastream, OBX index provided by Oslo Stock Exchange, for Norwegian GDP: http://statbank.ssb.no/statistikkbanken/default_fr.asp?PLanguage=1 > 09 National economy and external trade > 09.01 National accounts/National accounts, quarterly/09186: Final consumption expenditure. Unadjusted and seasonally adjusted figures (1978K1-2012K1)

²¹ Own calculations – Numbers for Norwegian GDP: same as above, Oil prices were found using Datastream.

As we see from *Table 2.3-1* above, the correlation between Norwegian GDP and oil price, is much higher than between US GDP and oil price. The same is true for oil price and the different stock indices. This insight may question whether the correlation found between the stock markets and GDP may be due to the fact that both are correlated to fluctuations in oil price, hence raising the question of spurious causality in our models. We will comment more on this in section 6.6.

So why should equity-indices matter in Norway? The Oslo Stock Exchanges is positively correlated with the oil price, which means that the equity index will increase with the oil price, as will Norwegian GDP as we are a major oil exporter. This is in contrast to the American economy for example, who is one of the world's largest oil importers. The Norwegian economy is also characterized by being small, open and dependent on other natural resources like fish. While the macroeconomic statistics have lags, the effect in real activity could be rather immediate for such shocks, and so this could be an important reason for why equities are more informative in Norway. The OSE and the Norwegian GDP are more volatile than in many other countries, this is in line with this assumption (Aastveit & Trovik, 2008).

2.3.1 Stocks' expectations of the real economy

Stock prices are influenced by expectations about future interest rates and company earnings. Therefore they depend on the expected development in the real economy. Stock prices may also influence the economic development through other canals as well (Gerdrup, et al., 2006):

- **Wealth:** Stock prices can be important for the households total wealth, hence an increase in stock prices may motivate an increase in spending.
- **Credit:** Stock prices may influence the access and costs of financing due to information asymmetry. A reason for this is that a decrease in total wealth may reduce opportunities to obtain debt.
- **Investments:** Changes in the stock prices may give signals to a company's management to increase, or decrease, its real investments. A stock price increase may signalize that the corporations implemented real capital exceed the costs of financing. Hence, new real capital is worth more to the owners than the associated costs.

2.4 The OBX Index

The OBX Total Return Index is a free float market capitalization weighted total return index. The index is dividend adjusted (as of January 2nd 1996). It consists of the 25 most traded securities on the Oslo Stock Exchange Benchmark Index (OSEBX) the previous 6 months. The change in included companies and their weights in the index is implemented every third Friday of June and December. The companies are given a number of “OBX shares” according to their relative market capitalization, which the weights are based on. The number of shares is held constant through each period, with exceptions of special capping rules and various corporate events.

The capping rules are; that the largest company cannot have a weight larger than 30 % of the index and no other company may have more than 15 % of the weight. Violations of these rules may also result in a change in OBX shares, and distribution of weights.

In 2012 a new capping rule was introduced to the OBX. This rule says that the aggregate weight of companies registered outside European Economic Area (EEA) could maximum be 10 %. The new rule changed the weights of firms like Subsea 7 S.A. This new rule resulted a minor change in the OBX, Oslo Stock Exchange own estimations showed a 99.995 % correlation between the “new” and “old” index in 2010 numbers (Oslo Børs, 2010).

When a firm is bought by a non-OBX company, or delisted, the company is simply removed from the OBX and its weight distributed amongst the remaining companies, this is an example of a corporate event. This happened for instance when Amersham PLC was taken of the exchange in 2004²² and when Awilco was bought in 2008²³. Another example of a corporate event was when the two OBX listed companies, Subsea 7 Inc. and Acergy S.A, combined to Subsea 7 S.A. In the merger, one Subsea 7 Inc. share entitled the shareholder to 1.065 shares in the new firm. This merger impacted the OBX in two ways, first of all there were only 24 companies left in the index, and second a new weighing was needed. The new firm got the number of shares equal to the number of OBX shares held by Acergy

²² Oslo Stock Exchange, Oslo, 18. mars 2004, *Oslo Børs Aksjederivatmelding 16-04*.

²³ Oslo Stock Exchange, Oslo, 7th July 2008, *Exchange Notice Derivatives 15-08*.

S.A. and 1.065 times the number of OBX shares held by Subsea 7 Inc²⁴. Similar cases are the merger between DnB and Gjensidige in 2003²⁵ and when Saga Petroleum decided to convert the b-shares into a-shares in 1998²⁶.

2.5 Fundamental indexing

Fundamental indexing was pioneered by the research agency Research Affiliates (RA), who specializes in investment strategies and asset allocations. Their Research Affiliates Fundamental Index (RAFI®) was introduced in 2005 by Robert D. Arnott, Jason Hsu, and Philip Moore in their book *Fundamental Indexation*. In developed markets, research has shown that using fundamental indices as exchange traded funds (ETF) gives 2 % to 4 % yearly added value from large company stocks compared to a market capitalization weighted index. The potential for added value in emerging markets and similar is even higher. In this section we will explain the rationale of fundamental indexing, and also provide two theories of why this strategy has excess return compared to regular market capitalization weighted indices. Then we will shed some light on why we think this is a better indicator for GDP than a regular MCAP weighted index.

MCAP weighted indices weigh the companies according to the companies' share of the total market value of the index. As a company's stock price increase or decrease, so will its weight in the index. Research Affiliates follows the assumptions of inefficient markets that cause mispricing, and that prices will revert back to their fair value (Research Affiliates, 2007-2012). They also assume that the prices across all the companies of an index will be correct on average. Following Research Affiliates assumptions the stock prices are not necessarily correct, and by construction an overpriced stock will be over weighted in a capitalization weighted index, and an underpriced stock will be underweighted. The basic idea of fundamental indexing is to remove the connection between market expectations and prices from index weights, and in doing so efficiently removing the double error of mispricing that can occur in a MCAP weighted index. This "double error" is called a performance drag. Instead of using the "Wall Street" term market capitalization to weigh the index, the fundamental index uses

²⁴ Oslo Stock Exchange, Oslo, 5th January 2011, *EXCHANGE NOTICE DERIVATIVES 01-11*.

²⁵ Oslo Stock Exchange, Oslo, 4th desember 2003 *OSLO BØRS AKSJEDERIVATMELDING 32-03*.

²⁶ Oslo Stock Exchange, Oslo, 23. April 1998, *OSLO BØRS AKSJEDERIVATMELDING 27-98*.

"Main Street" definitions of company size such as sales, revenue, profit, employees, and total assets. In contrast to a MCAP weighted index a fundamentally weighted index must be rebalanced to follow its definition. If this is done often the transaction costs will be high, done too seldom the difference between the wanted weights and the actual ones becomes large enough so that some of the negative attributes of a MCAP weighted index may be introduced (Arnott, et al., 2005). Research Affiliates rebalances once a year and argues that the transaction costs are low because the indices are concentrated in liquid stocks. The discussion of transaction cost is not relevant to us as we are not going to trade, only measure the change of an index.

Research Affiliates argues that most of the excess return comes from the elimination of the performance drag. Jason C Hsu (2006)²⁷ provides the following example of how this works;

Suppose there are only two stocks in the market, A and B, each with one share outstanding. Suppose the fair fundamental values (which investors do not observe) are \$10 per share for each stock. Further, suppose that market prices are noisy, and that there is a 50/50 chance that a stock can be overvalued or undervalued by \$2 (equivalent to assuming a 20% noise in price). Note that the expected "mispricing" in either of the two stocks is zero and we cannot know which stock is overvalued or undervalued. For simplicity, we also assume that the two stocks have the same systematic factor exposure (same market beta in the CAPM context), which leads to a 10% return on equity capital. Observe that the cap-weighted market portfolio has $\left(\frac{12}{12+8}\right) = 60\%$ in the overvalued stock and $\left(\frac{8}{12+8}\right) = 40\%$ in the undervalued stock. However, had prices reflected fundamentals, the portfolio weight would have been $\left(\frac{10}{10+10}\right) = 50\%$ in each. After one period, even assuming that the overvaluation and undervaluation does not dissipate, the cap-weighted portfolio return would be $60\% \times \frac{\$10 \times 10\%}{\$12} + 40\% \times \frac{\$10 \times 10\%}{\$8} = 10\%$ However, had the "fair-value-weight" been applied, the "fair-value-portfolio" would earn a return of $50\% \times \frac{\$10 \times 10\%}{\$12} + 50\% \times \frac{\$10 \times 10\%}{\$8} = 10.42\%$. The intuition for the cap-weighted portfolio's return drag is clear. The cap-weighted portfolio

²⁷ Jason C. Hsu, 2006. Cap-Weighted Portfolios are Sub-Optimal Portfolios. Journal of Investment Management, Vol. 4, No. 3, (2006), pp. 1–10

underperforms, because it puts more weight in the overvalued stock and less weight in the undervalued stocks. The return drag is clearly related to the over/undervaluation. Suppose in this example the mispricing was \$3 (30%), instead of \$2, the return drag on the cap portfolio relative to the fair-value weighted portfolio would be 0.99. At \$4 and \$5 mispricing the return drags are 1.90% and 3.33%, respectively!

Andersson (2009)²⁸ shows that fundamentally weighted indices, at least in a Swedish context, have a bias towards small-cap and value stocks. This kind of stocks have had a high return the last 30-40 years, but are often associated with higher risk. Andersson finds that fundamental indexing is an indirect way to pick small-cap and value stock and accordingly a high risk- high return strategy. This indicates that the abnormal return of the fundamental index is due to the Fama and French three-factor model. In their seminal paper, *The Cross-Section of Expected Stock Returns* (1992), Fama and French found that small-cap and value stocks had higher expected return than large-cap and growth stock. If one believes in the efficient market, this would imply higher risk. William J. Bernsteins research on the matter shows that slightly less than two thirds of the surplus return of fundamental indices, compared to MCAP weighted indices, is due to the value and size factors, and the remaining third should be credited Research Affiliates, as he does not find the source of this. Bernstein recommends further research into how this latter third occurs (Bernstein, 2006). The source of abnormally high return and then the value/size tilt is the most discussed topic in papers on fundamental indexing.

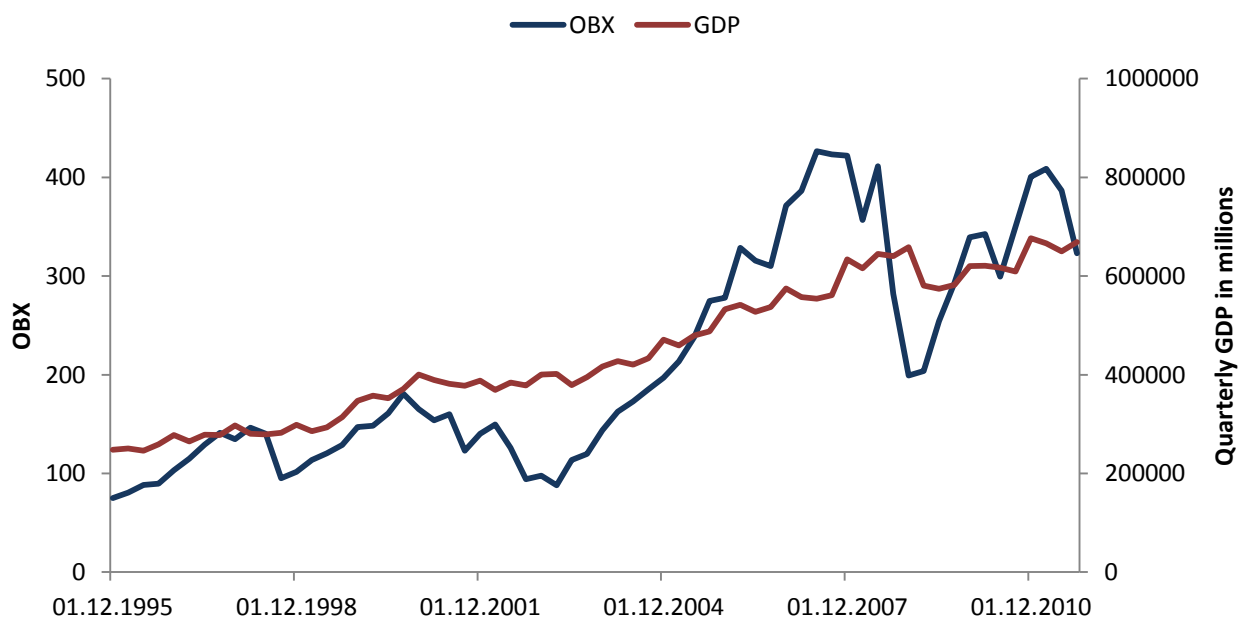
How important are these discussions on a value and size tilt for our leading indicator model? We want another way to weigh the index so its weights better represents the current company size, and therefore its importance for the nation's economy. The size and value tilts found in fundamental indices are tilted from a market capitalization point of view. And this is our point; we do not want market expectation to be part of the weighing as companies with very high price to book ratio may

²⁸ Jan O. Andersson, 2009, *Irrational Indexation*. Stockholm School of Economics, Department of Finance, Master Thesis in Finance.

have large market capitalization because of expected future growth and size, and not its current contribution to national economy.

Leaving the discussion of whether this is a better way to weigh an index behind us, there are more arguments for why a fundamental index is a better leading indicator than a regular MCAP weighted index. In Arnott et al. (2008)²⁹ it is shown that a fundamentally weighted index follows a MCAP weighted index benchmark, though underperforming during a bubble, but it significantly outperform the benchmark during a bust.

Graph 2.5-1



30

As we see from *Graph 2.5-1* showing the original OBX and GDP (Q4 1995 =1), the accumulated growth of GDP is much smoother than the OBX. If we get the same results as Research Affiliates, our fundamental index will not increase as rapidly during bubbles and not fall as heavily during the busts, giving a smoother line, closer resembling GDP growth.

²⁹ Research Affiliates, LLC, Robert D. Arnott, Jason C. Hsu, John M. West, 2008. *The fundamental index: a better way to invest*. John Wiley & Sons, Inc., Hoboken, New Jersey.

³⁰ Own calculations - Numbers from http://statbank.ssb.no/statistikkbanken/default_fr.asp?PLanguage=1 > 09 National economy and external trade > 09.01 National accounts/National accounts, quarterly/09186: Final consumption expenditure. Unadjusted and seasonally adjusted figures (1978K1-2012K1), OBX values provided by Oslo Stock Exchange

3 Literature review

Even though there are a wide number of methodologies available for constructing FCIs, they tend to fall into two broad categories: the weighted-sum approach and the principal-components approach. Below we will provide a short description of the two approaches, as presented in Hatzius et al. (2010):

“In the weighted-sum approach, the weights on each financial variable are generally assigned based on estimates of the relative impacts of changes in the variables on real GDP. These estimates or weights have been generated in a variety of ways, including simulations with large-scale macroeconomic models, vector autoregression (VAR) models, or reduced-form demand equations.”

“The second broad approach is a principal components methodology, which extracts a common factor from a group of several financial variables. This common factor captures the greatest common variation in the variables and is either used as the FCI or is added to the central bank policy rate to make up the FCI (this latter method is a combination of the weighted-sum approach and the principal-components approach).”

There are some commonalities and differences often seen within the FCI methodologies. Financial condition indexes are mostly based on the current value of financial variables, but some take into account lagged financial variables as well. Some FCIs summarize the impact of financial conditions on growth, others measuring whether financial conditions have tightened or loosened. Though the specific variables included in various FCIs may differ considerably, there are some commonalities. Most of the FCIs include some measure of short-term interest rates, long-term interest rates, risk premium, equity market performance, and exchange rates. One can for instance use the levels of each of the variables, often seen in the weighted-average approach, or you could standardize the variables. Only a few cases include a variable of stock market wealth or market capitalization, there are mostly rates or financial prices which are used in the different FCIs (Hatzius, et al., 2010).

3.1 International research

3.1.1 Bloomberg Financial Conditions Index for the US

The Bloomberg Financial Conditions Index for the US (BFCIUS) is readily accessible to those in the financial markets and it is updated daily, thus making it a convenient measure to track financial conditions. The index is an equally weighted sum of three major sub-indexes: Money Markets, Equity Markets, and Bond Markets. Each of these major sub-indexes is then again made up of a number of underlying indicators. These are also equally weighted within the major sub-indexes (Hatzius, et al., 2010). The values of this index are Z-scores, which represent the number of standard deviations that the current financial conditions lay above or below the average of the 1994 – June 2008 period. If the index is negative it indicates stress within one or all of the sub-markets or potentially stock market weakness³¹.

In 2009 they introduced a new index as the BFCIUS did not capture the recent sub-prime crises in the US. The new index was called BFCIUS+ and includes real estate prices and several yields. The two indices are mostly correlated, but the new index gave signals of the financial crises in 2007, and was generally a better measure of the overall economic performance (Rosenberg, 2009).

3.1.2 Citi Financial Conditions Index

The Citi Financial Conditions Index (CFCI) contains six financial variables and utilizes a weighted sum approach. The weights are determined according to reduced-form forecasting equations of the Conference Board's index of coincident indicators (the six-month percent change in the coincident index). It was created to summarize the financial variables effects on economic activity in a simple and transparent fashion. The index includes variables of corporate spreads, money supply, equity values, mortgage rates, the trade-weighted dollar, and energy prices. All nominal values are deflated, and real measures are transformed in order to eliminate trends. The CFCI is measured in number of standard deviations from the norm, meaning that with a CFCI-value of zero one expects a normal pace of expansion, positive values indicate that the financial variables jointly exert an expansionary force on the economy (D'Antonio, 2008).

³¹ Bloomberg, <http://www.bloomberg.com/quote/BFCIUS:IND>.

3.1.3 Deutsche Bank Financial Conditions Index

Deutsche Bank utilizes a combination of the principal component and weighted sum approach in their financial conditions index (DB FCI). First they extract the first principal component from a total of seven different standardized financial variables. Then they set the FCI to the weighted sum of this principal component and the target federal funds rate using a regression of real GDP growth on the financial variables and lagged GDP growth to determine the weights that is to be assigned to each of them. One can interpret the index as percentage drag, or boost, to GDP from financial conditions at a point in time given by the level of the index, depending on whether the index is negative or positive, respectively (Hatzius, et al., 2010).

In 2010 Deutsche Bank presented an updated version of DB FCI which included as many as 45 financial indicators in their global Economic Perspectives report. This new FCI was called the Monetary Policy Forum Financial Conditions Index (MPF FCI). The index is constructed using unbalanced panel estimation techniques as some of the new variables has limited history and are released with a time lag (Deutsche Bank Securities inc., 2010)

3.1.4 Federal Reserve Bank of Kansas City Financial Stress Index

This index contains 11 different, standardized, financial indicators, and utilizes a principle-component approach. It was developed in early 2009 by Hakkio and Keeton. The index is designed to capture five key aspects of financial distress in which each of the 11 indicators are a part of. The variables chosen can be divided into two categories: yield spreads and asset price behavior. They were chosen so that they satisfy these three criteria: 1) be available monthly, and have a history extending back to at least 1990; 2) be either market prices or yields; and 3) represent at least one of five financial stress features identified by the Kansas City Federal Reserve (Hatzius, et al., 2010).

Positive values of the Federal Reserve Bank of Kansas City Financial Stress Index (KCFSI) indicate that financial stress is above the long term average and vice versa. Historically it is found that high values of KCFSI often coincided with known periods of financial stress. Hakkio and Keeton (2009) also show that the KCFSI provides valuable information about future economic growth. Results indicate that financial stress can lead to a decline in the economy through three possible channels; (1) uncertainty

about prices or other investors actions, (2) decreasing business and households financial spending and (3) tighter credit standards in banks either by raising (a) the interest rate or (b) minimum standards. Unfortunately, the article does not provide a critical level at which financial stress is a serious concern (Hakkio & Keeton, 2009).

3.1.5 Macroeconomic Advisers Monetary and Financial Conditions Index

Macroeconomic Advisers (MA) constructed its monetary and financial conditions index to take into account the dynamic effects of financial variables on GDP over time in the late 1990s (Macroeconomic Advisers, 1998). They developed a “surface impulse response” methodology in aggregating the five different financial variables into a Financial Condition Index (MA FCI). These are; real short rate, real long rate, dividend ratio, real exchange rate, and real stock market capitalization. MA generate the response functions by estimating the partial effects of changes the financial variable on real GDP growth over time using simulations with MA’s large-scale macroeconomic model. Next they inverted and aggregated the response functions so that the MA FCI shows the combined effects of current and past changes in each of the financial variables on real GDP growth in the current period at any point in time (Hatzius, et al., 2010).

3.1.6 OECD Financial Conditions Index

The OECD Financial Conditions Index (FCI) was constructed by Guichard and Turner (2008) on behalf of the OECD.

It utilizes the weighted sum approach including six financial variables. The index weights are found by using a regression of the output gap on a distributed lag model of the financial indicators. They have normalized the weights relative to the change in interest rates, to make it so that a one unit increase in the FCI is equivalent to the GDP effects of a one-percentage-point increase in the real long-term interest rate (Hatzius, et al., 2010). They estimate using two models with overlapping data material: first a reduced form equation model, then a Vector Auto-Regression model (VAR) to account for any type of correlation between the variables (Guichard & Turner, 2008).

One new feature in this FCI is the inclusion of a variable for tightness in non-price credit bank lending standard, namely: the Federal Reserve Senior Loan Officer Survey. The survey is conducted by the Federal Reserve Board (FED). The FCI have been calibrated so that a unit decline in the index implies a 1 % reduction in the level of GDP after 4-6 quarters. This horizon is chosen because of its relevance for the monetary policy makers (Guichard & Turner, 2008).

This work was extended further too also include the Euro Area, Japan, United Kingdom by Guichard et al. (2009)³².

3.1.7 The OECD Composite Leading Indicators

In the 1970s the OECD developed their system of composite leading indicators (CLI) with the purpose of providing early signals of turning points in economic activity. OECD CLI uses a reference series as a proxy for economic activity. They measure fluctuations of economic activity as the variation in economic output relative to its long term long term trend, which is set equal to 100 in their model. In March 2012 the OECD investigated whether methods could be applied to generate monthly estimates of GDP based on the official quarterly estimates. This investigation demonstrated that it is feasible to do so, whilst also continuing to provide high quality results. Therefore from April 2012 the OECD switched to using GDP as the proxy, ceasing to rely on the previously used index of industrial production (IIP) as an intermediate target (Gyomai & Guidetti, 2012).

The OECD CLI is constructed by finding economic time series that show similar cyclical fluctuations, or preferably precedes, the fluctuations of the business cycle. The different steps used for pre-selecting reference and component series, filtering, evaluation, aggregation, and presentation of the CLI model are well explained in Gyomai and Guidetti (2012). The OECD has different measures for different countries, and for Norway the measures is as follows³³:

³² Stéphanie Guichard, David Haugh, David Turner, 2009, *Quantifying the Effect of Financial Conditions in the Euro Area, Japan, United Kingdom and United States*. Economics Department Working Papers No. 677

³³ OECD, http://www.oecd.org/document/12/0,3746,en_2649_34249_1891148_1_1_1_1,00.html.

- *Exports to UK (USD)*
- *Stocks of orders for exports (manuf., mining, quarrying) (% balance)*
- *Production (manuf.): tendency (% balance)*
- *General judgment of the outlook for the enterprise in next quarter (manuf., mining, quarrying) (% balance)*
- *CPI All items (2005=100)*
- *Share price index (industrials) Oslo (2005=100)*

It is worth noting that they use equal weights for the different sub-indices used in the index. They have three different presentation forms, they are: (1) Amplitude adjusted, which is the average of the de-trended and smoothed component and reference series. This amplitude adjusted CLI rescale the averaged CLI to match the amplitudes of the de-trended reference series. It is this presentation form that we use to compare with our FIs in this paper, (2) Trend restored, this reflects the product of the reference series trend and the amplitude adjusted CLI, and (3) 12-month rate of change, calculated from the trend restored CLI (Gyomai & Guidetti, 2012).

3.1.8 The National Bureau of Economic Research

The National Bureau of Economic Research (NBER) use a standard Principal Component approach including 45 indicators, each analyzed quarterly. They try to improve the predictive capabilities of FCIs by expanding the data history and coverage, while at the same time disentangling macroeconomic and policy influences from pure financial shocks. Compared to other FCIs additional indicators are included and they cover a wider range of both quantitative and survey based indicators. In addition the use of panel estimation techniques allow for unbalanced time series, resulting in longer time series, and also by controlling for past GDP and inflation which further confer the predictive power of the FCI (Hatzius, et al., 2010).

This index includes a broader range of survey-based measures compared to the OECD FCI for the US. A total of seven surveys covering bank lending, consumer and business credit conditions in the US are included. They analyze the survey's performance using prediction tests, two and four months ahead,

and the results show that the relative root mean square error (RRMSE) of the group of surveys were lowest when tracking growth four months ahead, at least in the final period (2005 – End of Sample). In other words, when looking four months ahead there was least noise in the surveys' forecasting performance (Ørbeck & Torvanger, 2011).

In their concluding remarks, Hatzius et al. (2010) state that this FCI outperformed many other FCIs in recent years, but did not do so during earlier periods. They find that their index is unstable over time, but it seems to do especially well in times of unusual financial stress originating from within asset markets. The authors found that the purging of macroeconomic influences had the most positive contribution in outperforming the other FCIs, and helping their FCI to work well in times with unusual financial stress. The overall index performed noticeably better in recent years than any of its major subcomponents (Hatzius, et al., 2010).

One major weakness of this FCI is its size which makes estimations and updates more complex. However, the authors also say that it is possible to translate the FCI into monthly intervals using monthly and quarterly data if desirable (Hatzius, et al., 2010).

3.2 Norwegian research

3.2.1 Business cycles in Norway

Husebø and Wilhelmsens (2005) analyze stylized facts regarding the Norwegian business cycles. They study the empirical relationships between the aggregate business cycle and the cyclical components of individual macroeconomic time series. The stylized facts methods are said to involve fewer assumptions about the structure of the economy, which is highly uncertain and difficult to model, than common econometric models. This allows the data to speak with fewer restraints. This is why stylized facts methods are a popular alternative to the econometric models. This paper was an extension of a paper by Bjørnland (2000)³⁴ and increases the number of macroeconomic variables to a total of 30, while Bjørnland used 10. The results presented by Husebø and Wilhelmsen (2005) suggests that the stylized facts about the Norwegian business cycle are fairly similar to those for the

³⁴ Bjørnland, Hilde Christiane, 2000, *Detrending methods and stylized facts of business cycles in Norway - an international comparison*. Empirical Economics, **25**, 2000, pp. 369-392.

US and euro area. The main discrepancy from the standard results found in the literature is that they find real wages to be procyclical and productivity to be acyclical in Norway during the period under review (Husebø, et al., 2005).

3.2.2 Historical indicators as predictors of banking crises

Riiser (2005) examine whether historical indicators can be used to predict banking crises through the last 150 years. The gap between actual observations and trend for real house prices, real equity prices, gross fixed investment and credit on the basis of Norwegian data are calculated back to 1819 by using a Hodrick-Prescott filter(HP filter)³⁵. To calculate the trend they use the HP filter, and a recursive method, meaning that data until the beginning of the year in question is included in the calculation of the trend. This is done so that one has the same information available as the decision makers would have had at any given time. They use a percentage deviation from the trend to calculate the gaps, with the exception of the credit gap, which is measured in difference in percentage points from the trend. She found that all of the gap indicators are useful for signaling the build-up of imbalances and banking crises in Norway. They show a common pattern, with few exceptions: the gaps widen from one to six years prior to the banking crises. She also finds that the threshold values may be higher in Norway than in comparable international studies. Another possibility is that the critical values of the gap indicators are not constant. For instance, the gap values may depend on the number of indicators that react. Second, they may depend on the financial strength of the banking sector. Narrow gaps can lead to banking crises if the banking system that is examined is not very sound, and vice versa; wide gaps may lead to banking crises when the banking system is more robust (Riiser, 2005).

The conclusion of this article is conditioned by the uncertainty associated with long historical time series and the lack of data for some gap indicators in certain periods. It is also mentioned that the gap indicator analysis should be supplemented by an analysis of the robustness of the banking sector as

³⁵ For a presentation of the Hodrick-Prescott filter, see: Robert J. Hodrick and Edward C. Prescott, *Postwar U.S. Business Cycles: An Empirical Investigation*. Journal of Money, Credit and Banking Volume 29, No. 1 (February 1997), pp. 1-16. Published by Ohio State University Press.

the financial systems ability to withstand pressure from disturbance depends both on the quality of banks' portfolio and on the banks' capital adequacy, among other things (Riiser, 2005).

Riiser (2008) presents updated figures from her 2005 article. One major difference was a decreased critical value for the credit gap. The new analysis also revealed high values for all of the gap indicators in 2007, indicating vulnerability in the financial system in 2007. Riiser explains in her article that the gap indicators cannot signal imbalances in other markets than those included. Here she also analyzes the gap indicators ability to predict future crises. Riiser (2008) discuss the Norwegian banks increased dependency of market funding as the banks' lending growth has experienced a much steeper increase than deposits in the same period. This leads to liquidity challenges for the banks in the event of turbulence in international money markets, this we saw a good example of during the last financial crises in 2008. Liquidity risks, and other types of risks, are typically underestimated by the financial institutions when there are excessive optimism that leads to a surge in asset prices and credit. Meaning that the gap indicators can be able to signal increased financial vulnerability in the future if there are a relationship between debts financed imbalances and liquidity risk (Riiser, 2008).

Riiser (2010) provides updated figures from the 2008 version up to 2009. The critical value for the credit gap is reduced due to a lower credit gap ahead of the crisis in 1988-1993 in the updated calculations compared with the analysis in Riiser (2008).

The housing price, investment, and credit gap were all higher than their critical values in 2007. In 2009 she found that all the gaps had diminished, despite the fact that the credit gap was still higher than its critical value again in 2009. In the words of Riiser: *A correction in financial imbalances can take a long time. The historical analysis indicates that financial stress has often arisen from one to six years after the indicators have reached their critical values.* Riiser (2010) concludes by suggesting that the Norwegian financial system may still encounter some challenging periods in the coming years (Riiser, 2010).

3.2.3 Financial figures and the real economy

Gerdrup et al. (2006) examine whether financial figures can provide information about future development in the production gap and GDP for mainland Norway. They found that house prices, and stock prices, among others might be used as leading indicators for growth in GDP and production gap using a simple bivariate correlation analysis.

Further they developed a simultaneous multi-equation model. This was done to overcome some of the limitations associated with simple correlation analysis and to account for the fact that many variables could affect GDP with different time lags. Their preferred model included lagged variables for real asset prices, several variables for domestic real credit to corporations, and GDP for mainland Norway. This model could forecast the GDP development fairly good for as many as eight quarters ahead (Gerdrup, et al., 2006).

Gerdrup et al. (2006) argued that the correlations between the economy and financial sizes may change over time. Therefore information from previous periods will contain less relevant information about the future. They also detected a constant long-term relationship between real credit and real share prices. Growth in real stock price was affected positively by an increased GDP growth in the same and previous quarter. Shocks in GDP growth could affect real stock growth, which in turn could affect real credit growth to companies and again affect GDP growth.

Gerdrup et al. (2006) found evidence saying that stock prices, credit growth, money supply growth, real exchange rate, and the spread between long- and short term interest rate all were good leading indicators for change in GDP and the production gap.

4 Methodology

4.1 The OLS assumptions

For an Ordinary Least Squares regression to be the Best Linear Unbiased Estimator (BLUE), it has to fulfill certain assumptions (Gujarati & Porter, 2009). For a simple regression model, that is for instance: $Y_t = \alpha_t + \beta X_t + u_t$, the assumptions are:

- No autocorrelation
 $\text{Cov}(u_i, u_j) = 0$
- The Regression model is linear in the parameters and in the disturbance term
- Homoscedasticity
 $\text{Var}(u_i) = \sigma^2$
- The average value of the disturbance term is zero
 $E(u_i) = 0$
- None of the explanatory variables is in perfect linear relation to one of the other explanatory variables.
- The disturbance terms are normally distributed
 $u_i \sim N(0, \sigma^2)$

4.2 Testing for Model specification and wrong functional form

4.2.1 Normality

After running an OLS regression one needs to test whether the normality of the disturbance term holds. This is important for us because our sample size is somewhat limited. The normality assumptions help determine the exact OLS estimators, and whether or not we can use the t-, F-, and chi-square distributions. If the error term is normally distributed, the OLS and Maximum Likelihood regressions coefficients will be identical. There are many different tests for normality, and among them are the Anderson-Darling test, and the Pearson's Chi-squared test. Hain (2010) found in his study that the extension of the Royston test, which he just called the zW_R test to be the best, but that no single test for normality is the best for every situation. The Jarque-Bera (JB) test is very often used

in practice, and results are easily obtained. That is why we chose to use this in our thesis. The null hypothesis of the JB test is that the residuals are normally distributed, and its statistics follows the Chi-square distribution (Hain, 2010).

"There is no excuse for failing to plot and look" - John W. Tukey³⁶

As there are no single preferred statistical test for normality in the empirical research, we will also use Q-Q plots to investigate further and compare results. For a short presentation review of the JB-test, see in Gujarati and Porter (2010)³⁷. For a more detailed presentation of the Jarque-Bera test, see Jarque and Bera (1987)³⁸

4.2.1.1 Q-Q plots

Interpreting the Q-Q plots can be difficult, and experience is the key. How does one detect skewness or heavy tails with a Q-Q plot? At Murdoch University they say the following about interpreting these plots:

A sufficiently trained statistician can read the vagaries of a Q-Q plot like a shaman can read a chicken's entrails, with a similar recourse to scientific principles. (Murdoch University, 2009)

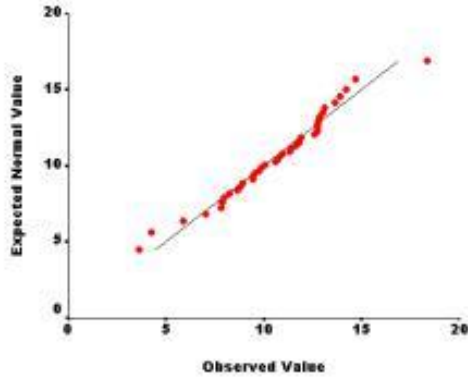
We wish the residuals to form a straight line in the Q-Q plots to know that they are normally distributed. Below is examples provided by Murdoch University of normally distributed residuals. The plots represent a population size of 40 and one of 10 which is close to the sample sizes when we perform our tests for structural breaks in *section 4.3*.

³⁶ John W. Tukey, 1977, *Exploratory Data Analysis*. First Edition. Addison Wesley.

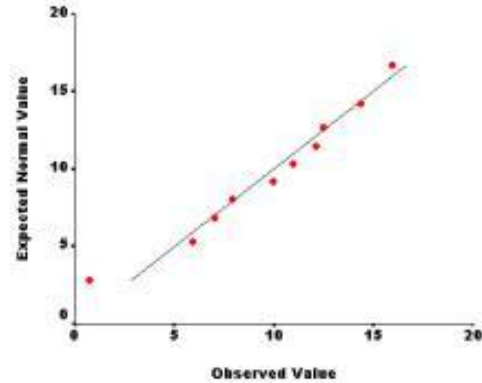
³⁷ Gujarati, D. N. & Porter, D. C., 2009. *Basic Econometrics*. Fifth Edition. New York: McGraw-Hill page 131-132

³⁸ C.M. Jarque and A.K. Bera, 1987. *A Test for Normality of Observations and Regression Residuals*. The International Statistical Review volume 55, pages 163-172.

Figure 4.2-1



Sample size = 40

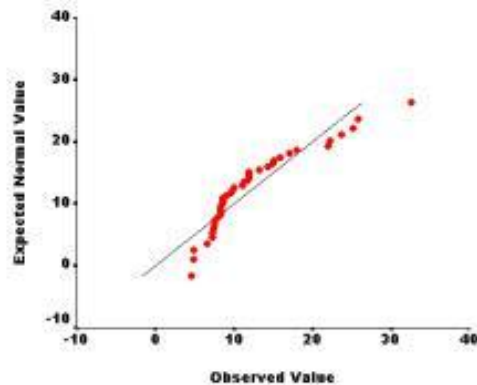


Sample size = 10

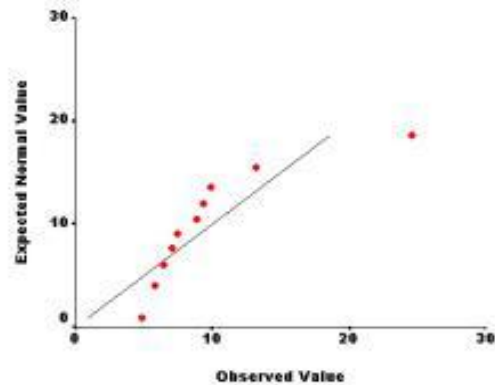
39

If the distribution is skewed to the right it is shown in the Q-Q plot as an arch starting below the $y=x$ line, crosses over and ends below the line again. The arch would be mirrored around the $y=x$ line if the population was skewed to the left. *Figure 4.2-2* shows examples with the same population sizes, but they are skewed to the right.

Figure 4.2-2



Sample size = 40



Sample size = 10

40

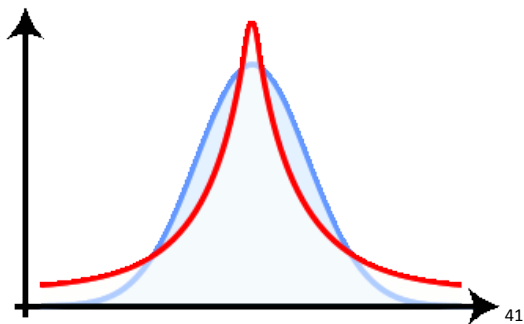
A distribution that is skewed to the right will have short left hand tail, and a long right hand tail. This means that in a right skewed distribution has its peak to the left of a normal distribution.

³⁹ Plots from <http://www.cms.murdoch.edu.au/areas/maths/statsnotes/quizzes/qg/qgnormalpuzzle.html>

⁴⁰ Plots from <http://www.cms.murdoch.edu.au/areas/maths/statsnotes/quizzes/qg/qglognormalpuzzle.html>

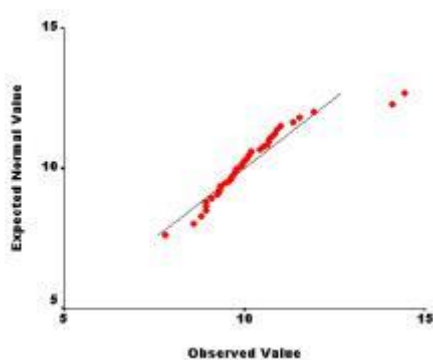
Another form of deviation from the normal distribution is a heavy tailed distribution or large kurtosis. This distribution is in contrast to a skewed distribution, symmetric but with longer and fatter tails than a normal distribution.

Graph 4.2-1

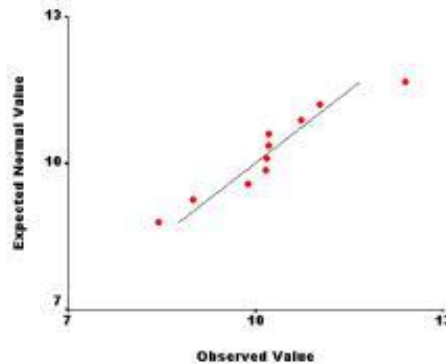


In the Q-Q plot below, this will show as a weakly s-bent shape. Again, two examples are provided from Murdoch University

Figure 4.2-3



Sample size = 40



Sample size = 10

42

⁴¹ From <http://www.quasol.net/portfoliooptimization.html>

⁴² Plots from <http://www.cms.murdoch.edu.au/areas/math/statsnotes/quizzes/qg/qgtpuzzle.html>

4.2.2 Autocorrelation

Autocorrelation in time series occurs if any disturbance term is related to the disturbance term of another observation.

Autocorrelation may occur for several reasons such as inertia or sluggishness of economic time series, mis-specification resulting from excluding important variables from the models or using incorrect functional form, the cobweb phenomenon, data massaging, and data transformation (Gujarati & Porter, 2009).

There are many tests for autocorrelation in the practical field of statistics. The reason being, so far no tests have been proven to be more statistically powerful than the others. We chose to use the Breusch-Godfrey (BG) test, also known as the Lagrange multiplier (LM) test, to check whether the data is affected by autocorrelation. The reason being that Breusch and Godfrey developed their test for autocorrelation so that it can test it in a more general sense than for example the Durbin Watson (DW) test, as it allows for non-stochastic regressors, like higher order Auto regression, AR(p), and moving averages, MA(q) (Gujarati & Porter, 2009).

4.2.2.1 The Breusch-Godfrey test

The hypotheses of the test are the following:

$$H_0: \text{covariance}(u_i, u_j) = 0 \quad i \neq j$$

$$H_A: \text{covariance}(u_i, u_j) \neq 0 \quad i \neq j$$

Hence, we are testing whether the error terms are uncorrelated for any given observed value of X in the dataset. The test statistics follows the Chi-squared distribution. If we obtain test-values that exceed the chosen critical-value, say the 95% level, we can reject the null hypothesis and assume autocorrelation in the residuals.

This model has, as mentioned above, several advantages over the DW test, but a drawback with the BG test is that the order of the autocorrelation function of the error term cannot be specified prior to

testing. Luckily we have methods to determine this, the Box-Jenkins method and selection criteria like Schwartz and Akaike are examples of this (Gujarati & Porter, 2009).

4.2.3 Multicollinearity

Multicollinearity relates to the issue of an existing relationship between some, or all, of the explanatory variables. Explanatory variables that are collinear are sometimes unavoidable. For example, it would be reasonable to assume that number of employees and amount of total assets will be collinear. Such correlations are a fact of life, and to regard a “natural law” as a problem is not at all constructive (Leamer, 1983). The real problem occurs in the presence of high or perfect multicollinearity. This will lead to large or infinite standard errors for the coefficients. Thus one cannot estimate coefficients with great accuracy.

Multicollinearity can occur for several reasons, such as the data collection method employed, constraints on the model or in the population being sampled, wrong model specification, or that the regressors` in the model share a common trend (Gujarati & Porter, 2009).

But multicollinearity does not violate any regression assumptions. The estimation will be BLUE, but it can be hard to get coefficients without large standard errors, hence if multicollinearity is present one can get a problem with statistically insignificant variables. This is the same problem that often occurs when having too few observations. So the answer for what to do with multicollinearity is the same as what to do with few observations; there is no correct statistical answer (Achen, 1982).

To detect if there are signs of multicollinearity in our data we will make a scatter plot to check for any signs of relations between the explanatory variables.

For a presentation review, Gujarati and Porter (2009)⁴³, for a more detailed presentation of Multicollinearity see Frisch (1934)⁴⁴.

⁴³ Gujarati, D. N. & Porter, D. C., 2009. *Basic Econometrics*. Fifth Edition red. New York: McGraw-Hill, pages 320 - 350

⁴⁴ Ragnar Frisch, *Statistical Confluence Analysis by Means of Complete Regression Systems*, Institute of Economics, Oslo University, publication no. 5, 1934.

4.2.4 Heteroscedasticity

Homoscedasticity applies to the assumption of constant variance in the disturbance term. If heteroscedasticity is present, the OLS method will be unbiased, but not provide the minimum variance or efficient estimators, hence, it will not be BLUE. In time series heteroscedasticity often takes form as autoregressive conditional heteroscedasticity, or ARCH. This form of volatility clustering is often present in financial time series, such as stock prices and inflation rates (Gujarati & Porter, 2009). Because of this, it is very important to check for ARCH effects in our models.

4.2.4.1 Engles Lagrange Multiplier test

We will use the Engles LM test, as provided by SAS Enterprise Guide 4.3, to check for the presence of ARCH effects. The LM test is performed by regressing the squared residuals of the original model on a constant term and q lagged values of squared residuals.

Equation 4.2-1

$$\hat{u}_t^2 = \alpha + \beta_1 u_{t-1}^2 + \beta_2 u_{t-2}^2 + \cdots + \beta_q u_{t-q}^2 + \varepsilon_t$$

The LM-test statistics are given by

Equation 4.2-2

$$(n - q)R^2 \sim \chi_{df=q}^2$$

The hypotheses of the LM test are:

H_0 : no coefficients are statistically different from zero

H_A : at least one of the coefficients is statistically different from zero

If the test statistics exceed the critical value of the chi-squared distribution, we can reject the null hypothesis of no autoregressive conditional heteroscedasticity and assume heteroscedasticity. We chose SAS to report p-values from lags 1 through 12 for the LM test, as we argue that this is sufficient to provide a good picture of if there are ARCH effects present.

4.2.5 Stationarity

Non-stationary and stochastic trend values are often observed in time series. Asset prices, such as stock prices and exchange rates, are often said to follow a random walk, i.e. they are non-stationary (Gujarati & Porter, 2009).

Regressing non-stationary variables might lead to the results being spurious. A Spurious regression refers to a regression that tends to accept a false relation, or reject a true relation, by flawed regression schemes (Chiarella & Gao, 2007). A way to avoid spurious regressions can be to take the first difference of a non-stationary time series in order to transform it into a stationary time series.

It is evidence showing that regressions based on differentiated variables may not be able to retain the long-term information contained in levels of variables and avoiding spurious regressions to reject a true relation, but in applied econometrics this is still an often used way of handling non-stationary time series.

Another problem with non-stationary time series is that one cannot generalize the results to other time periods. Since the main idea of our FI is to help policymakers through forecasts of GDP, generalization with regards to validity and causality is an absolute requirement.

Perron (1989)⁴⁵ advocates not to use tests like the Augmented Dickey-Fuller to test for unit-root as macroeconomic variables are not unit-root processes, but rather stationary series with a trend and structural breaks, where the financial crisis of 2008-2009 is an example of such a structural break.

We have used the autocorrelation function (ACF), and Partial autocorrelation function (PACF) correlograms to graphically check if our models were non-stationary. As we did not find any indications of non-stationarity in our models we have not performed any additional tests. As we have a realization of a stochastic process, we compute the Sample ACF and PACF (Gujarati & Porter, 2009). For simplicity we will simply address them as ACF and PACF in this thesis.

⁴⁵ Perron, Pierre, 1989. *Testing for a Unit Root in a Time Series with a changing mean*. Econometric Research Program Princeton University. Research Memorandum No. 347.

4.2.6 Ramsey's RESET test

RESET is short for Regression Specification Error Test. As the name implies Ramsey's RESET test is a general test of specification error. For description purposes you can find a simple walk through of the model presented in Gujarati and Porter (2009)⁴⁶.

The hypotheses of the test are as follows:

H_0 : R_{new}^2 is not statistically different from R_{old}^2

H_A : R_{new}^2 is statistically different from R_{old}^2

Ramsey's RESET test follow the F-distribution, and is used to check if the increase in R^2 is statistically significant. So if one finds the F-value to be significant at, say the 95% level, this means that the old model is mis-specified.

The main advantage of the RESET test is how easy it is to perform; it does not even require a specified alternative model. This takes us directly to its main disadvantage. The RESET test does not help you to specify a better model, it simply indicates that something is wrong (Gujarati & Porter, 2009).

4.3 Parameter Stability

When we use a regression model involving time series data we may observe a structural change. That is, there is a change in the relationship between the dependent and explanatory variables. This means that the parameters in the model would be statistically different if we estimated a model before and after the structural break. When this happens, an unrestricted model for the whole period would underperform both before and after the break. To check for structural breaks in our indicators we will use the Chow test.

4.3.1 The Chow test

The idea behind the Chow test is that the residual sum of squares (RSS) of the full period model, and the sum of the RSS of the two sub-period models, should not be statistically different if there is no structural break in the time series. In order to use the Chow test the following conditions must be fulfilled:

⁴⁶ Gujarati, D. N. & Porter, D. C., 2009. *Basic Econometrics*. Fifth Edition red. New York: McGraw-Hill, pages

- The time of the structural changes are known.
- The errors terms before and after the change are normally distributed and homoscedastic.
- The error terms in the two periods are independently distributed.

We have performed an F-test to see if the error variances of the two periods are the same. As we cannot observe the true variances, we can find the F-statistics follow this equation;

Equation 4.3-1

$$F = \frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(n - 2k)} \sim F_{[k, n-2k]}$$

where n the number of observations, k is the number of parameters in the model, and $(n - 2k)$ is the sum of degrees of freedom of the two sub-periods. If we obtain a significant F statistic we will reject the null hypothesis of parameter stability.

The problem with the Chow test when checking for a structural break is, that it can only check for known breaks. If the number of breaks is unknown or when they are is unknown, the Chow test is not applicable. The Cumulative Sum Control Chart (CUSUM) "test" is a sequential analysis that checks for structural breaks. This is a more complex and time consuming method.

To ensure that we can use the Chow test we have examined if the assumption of equal variance in the sub-periods are fulfilled. To do this we will perform an F-test, where the test-value calculated in the following way;

Equation 4.3-2

$$F = \frac{\hat{\sigma}_1/\sigma_1}{\hat{\sigma}_2/\sigma_2} \sim F_{(n_1 - k), (n_2 - k)}$$

where n_1 and n_2 are the number of observations in subsample 1 and 2, respectively, and k is the number of regressors in the model. This test value for F is equal to: $\frac{\hat{\sigma}_1}{\hat{\sigma}_2} \left(\frac{RSS_1}{RSS_2} \right)$ if the assumption; $\sigma_1 = \sigma_2$, is fulfilled. In practice this is what we used.

4.4 Causality test

Even though one can prove a relationship between two or more variables, this does not prove causality or the direction of influence between them (Gujarati & Porter, 2009). In time series regressions one can find a somewhat different situation because, as explained by Gary Koop:

...time does not run backward. That is, if event A happens before event B, then it is possible that A is causing B. However, it is not possible that B is causing A. In other words, events in the past can cause events to happen today. Future events cannot (Koop, 2000).

It should be noted that the question of causality is both deeply philosophical and controversial. Within this science, as with most others, there are extreme views where some say that “everything causes everything,” while others deny any existence of causality whatsoever (Gujarati & Porter, 2009). Yet others use other terms, like precedence used by for example Edward Leamer(1983), or predictive causality used by Francis Diebold. Diebold (2001) argues the following:

...the statement “ Y_i causes Y_j ” is just shorthand for the more precise, but long-winded, statement “ Y_i contains useful information for predicting Y_j , over and above the past histories of the other variables in the system.” To save space, we simply say that “ Y_i causes Y_j ” (Diebold, 2001).

4.4.1 Granger Causality test

The Granger causality tests are performed to explore whether a change in one variable causes a change in the other variable, vice versa, both directions, or if there are no causal relationship at all. It is assumed in the test that all relevant information for predicting the respective variables is contained in the time series data for these variables (Gujarati & Porter, 2009).

As an example we will use the regression:

Equation 4.4-1

$$GDP_t = \sum_{i=1}^n \beta_i FI_{t-i} + \sum_{j=1}^n \beta_j GDP_{t-j} + u_t$$

The hypotheses of the Granger causality test for whether FI Granger causes GDP are:

$$H_0: \beta_i = 0, i = 1, 2, 3, \dots, n.$$

$$H_A: \beta_1 \text{ or } \beta_2 \text{ or } \dots \text{ or } \beta_n \neq 0$$

This is how to find the tests F-value:

Equation 4.4-2

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n - k)}$$

The RSS_R is from Equation 4.4-1, only containing lagged values of the dependent variable, whereas the RSS_{UR} also contain FI-, and possibly lagged FI-variables. If one finds that the computed F value exceeds the critical F value at the chosen level of significance, 95% in our case, we reject the null hypothesis, meaning that at least one of the FI-variables belong in the model. In other words, that FI granger causes GDP. If we find that the computed F-value does not exceed the critical value we do not reject H_0 and assume that all of the FI-coefficients jointly are statistically not different from zero.

Using Equation 4.4-3 and Equation 4.4-4, presented below, we will explain the four different possible cases of causality:

Equation 4.4-3

$$GDP_t = \sum_{i=1}^n \beta_i FI_{t-i} + \sum_{j=1}^n \beta_j GDP_{t-j} + u_{1t}$$

Equation 4.4-4

$$FI_t = \sum_{i=1}^n \beta_i FI_{t-i} + \sum_{j=1}^n \beta_j GDP_{t-j} + u_{2t}$$

- *Unidirectional causality from FI to GDP* exists when we reject the null hypothesis of the Granger causality test for Equation 4.4-3, and in Equation 4.4-4 we do not reject the null hypothesis.
- *Unidirectional causality from GDP to FI* are indicated if one does not reject the null hypothesis in Equation 4.4-3, while in Equation 4.4-4 the null hypothesis is rejected.

- *Feedback, or bilateral causality*, is suggested if both regressions reject the null hypothesis.
- *Independence* is indicated when none of the sets of coefficients are found to be statistically different from zero in either of the regressions, hence we do not reject the null hypothesis.

Since, as mentioned above, the future cannot predict the past, we say that for the explanatory variable to granger cause the dependent variable, its changes should precede those of the dependent variable, and hence it must be leading. Therefore we say that the regressor granger causes the regressand if the inclusion of lagged regressors` significantly improves the prediction of the regressand in a model containing lagged values of the dependent variable as well as those of the explanatory variable (Gujarati & Porter, 2009).

There are some things that one needs to keep in mind when performing a Granger causality test;

- The variables are assumed to be stationary, if they are not, one have to transform them so that they are.
- To decide how many lagged terms to include in the causality test is not only a practical, but also important question. One can use for example the Schwartz Information Criteria.
- It is assumed no autocorrelation, if not the error terms need to be transformed.
- As our interest is testing the granger causality, the parameter estimates in the tests are not of interests, and is therefore not presented. Hence, we only present the results from the F-test.
- One can experience “spurious” causality. This can occur if our FI influence another variable, for example interest rates, and the interest rates again Granger cause the GDP. Hence, if one does not include the interest rates in the model and find that our FIs Granger causes GDP, this relationship may be spurious. Investigation of spurious causality is out of scope for this thesis.

(Gujarati & Porter, 2009).

As a conclusion of the introduction to Granger causality, keep in mind that what we are examining is whether or not one can detect the direction of causality when there is a temporal lead-lag relationship between two variables, in our example the FI and GDP. It is suggested that one may be

better able to predict the other variable than simply using its own past history if causality is established (Gujarati & Porter, 2009).

For a short presentation on Granger causality see Gujarati and Porter (2009)⁴⁷, for a more detailed presentation see Granger (1969)⁴⁸

4.5 Box-Jenkins modeling

The Box-Jenkins (B-J) model is also called ARIMA modeling. Formally written as ARIMA (p, d, q), where p represents the number of autoregressive orders to include, d the number of transformations needed, and q represents the number moving average parameters to include in the model. In short, the B-J methodology consists of five steps. First, check that the variance is stationary, and if not it needs to be transformed into stationarity. Second, identify a model, meaning that you chose tentative numbers of p, d, and q. Third, estimate the tentative model. Fourth, do a diagnostics check to see if the residuals are white noise. If they are, proceed to step five, namely forecasting, if not start over at step two.

In identifying the proper ARIMA model one use the ACF and PACF correlograms, as they indicate what should be done to get white noise residuals.

For a quick introduction see Gujarati and Porter (2009)⁴⁹, for a more thorough walk through we recommend Makridakis et al. (1998)⁵⁰.

4.6 Dynamic regression models

Sometimes the effect of a change in an explanatory variable does not show up in the forecast variable instantaneously, but is distributed across several time periods (Makridakis, et al., 1998). This means that changes in the explanatory variable today may impact the dependent variable both instantaneously, and several periods in the future. The main objective of dynamic regression

⁴⁷ Gujarati, D. N. & Porter, D. C., 2009. *Basic Econometrics*. Fifth Edition red. New York: McGraw-Hill, pages 652 – 658.

⁴⁸ C.W.J. Granger, 1969, *Investigating Causal Relations by Econometric Models and Cross-Spectral Methods*. *Econometrica* July 1969, pp. 424-438.

⁴⁹ Gujarati, D. N. & Porter, D. C., 2009. *Basic Econometrics*. Fifth Edition red. New York: McGraw-Hill, pages 777 – 784.

⁵⁰ Spyros Makridakis, Steven C. Wheelwright, Rob J. Hyndman, 1998. *Forecasting Methods and Applications*, Third edition, John Wiley & Sons, Hoboken, pp. 335-346

modeling is to identify the role of a leading indicator, the input series, in determining the variable of interest, the dependent variable. The intercept coefficient in such models has no real economic interpretation, therefore we will not further comment on this. For simplicity we will use one of our chosen models for our comparable leading indicators` as an example, but it can easily be generalized. This is the chosen model for OECDs CLI-index:

Equation 4.6-1

$$GDP_{t+1} = \alpha + \beta_1 CLI_t + \beta_2 CLI_{t-1} + \beta_3 GDP_{t-3} + \varepsilon_t$$

The values of β_1 and β_2 coefficients are called impulse response weights, and are a measure of how the future change in GDP responds to a change in the CLI-index today, β_1 , and one quarter back, β_2 . β_2 can also be interpreted as an intermediate effect of a change in CLI at time t. The lagged explanatory variables, here exemplified by CLI_{t-1} , are usually collinear, so caution is needed in attributing much meaning to each coefficient. If many lagged explanatory variables are included, the effect of it may become so small that it will be swamped by the random error term after some time (Makridakis, et al., 1998). For example: the lagged values may eventually turn negative even though a positive relationship is expected. This does not mean that there necessarily is a reverse in time, but the effect may have become so small that it is indistinguishable from zero and the coefficient came out negative by chance (Makridakis, et al., 1998).

In dynamic regressions both the input and output series should be appropriately transformed to take care of non-stationarity in variance and mean, and possibly also seasonally adjusted to make simpler models.

When selecting the model order, we have used a version of the “linear transfer function.” Below we will present a simplified explanation as given by Makridakis et al. (1998). For a more detailed description see Forecasting with dynamic regression models by Alan Pankratz (1991)⁵¹.

Step 1:

⁵¹ Alan Pankratz, 1991, *Forecasting with dynamic regression models*. John Wiley and Sons, New York.

The first step in identifying the appropriate dynamic regression model is to fit a multiple regression model of the form:

Equation 4.6-2

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \cdots + \beta_j X_{t-k} + N_t$$

Where k , which is the number of periods back in time, must be sufficiently large so that the model captures the longest time-lagged response that is likely to be important. N_t is an ARIMA process that needs its own selection process, described in section 4.5, and X_t is the leading indicator. Since the form of the noise is relatively unimportant at this stage, it is convenient to use a low-order proxy AR model for N_t . What we did in practice was to include all statistically significant lagged explanatory variables. We used the rule of thumb saying that the coefficients should be within the 95% statistical significance level.

Step 2:

If the error terms from the regression appear to be non-stationary and differencing seems appropriate, then differentiate the Y and X variables. Then re-estimate the model using a low-order autoregressive model for the errors, this time with differenced variables. We had transformed all of our variables before doing the regressions, hence we did not encounter the problem with non-stationary errors.

Step 3:

The next step is to identify an appropriate transfer function. That is, one must select the number of periods before X_t influence Y_t . The dead time is equal to the number of β -weights that are not significantly different from zero.

Step 4:

Here one calculates the errors from the regression model

Equation 4.6-3

$$N_t = Y_t - (\alpha + \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_j X_{t-k})$$

and identify the appropriate Autoregressive Moving Average (ARMA) model, described in *section 4.5*, for the error series. We found that for our FIs, the low order AR model was sufficient, but for some of our comparable indicators we had to go through more of an ARMA process. Here we had a more practical approach where we did different regressions and used the traditional model selection criteria SBC to select the most appropriate models. We ended up using different low order AR models also for the comparable indicators. We used the rule of thumb saying that the coefficients should be within the 95% statistical significance level. We had to make some exceptions, but they are described more thoroughly in later sections.

Step 5:

Re-estimate the entire model using the new ARMA model for the errors and the transfer model for X.

Step 6:

Check that the fitted model is adequate by analyzing the residuals to see if they are significantly different from a white noise series. Here the usual diagnostic tests can be used. Note that a wrong dynamic regression model may include significant autocorrelations in the residual series. If the residuals show any problems, it is good practice to reconsider the appropriateness of the transfer function as well as the error model.

We found different models for each of our comparable leading indicators, where some of them included the 2nd and 4th lagged values of GDP. The interpretation of these coefficients is not as straight forward as it may seem, because of long term effects in dynamic models. Therefore one should be careful when interpreting these. We will explain how these effects impact in *section 4.6.1.2*. It should be noted that in our model the 2nd lagged value is the latest available reported number at time t. For our FIs, we only used an AR process with the 2nd lagged dependent variable included.

4.6.1 Intermediate and long term effects in dynamic models

In dynamic regression models the coefficients is not always straight forward to interpret. The effect of the variables can be divided into three categories:

- Short-run effect
- Intermediate effect
- Long-run effect

4.6.1.1 Distributed lag

In a distributed lag model:

Equation 4.6-4

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + \beta_5 X_{t-5} + \beta_6 X_{t-6} + u_t$$

Here the short-run effect on the dependent variable is the β_0 coefficient, the first impact of X on Y.

The intermediate effect and the long-run effect of this model are similar. A one unit change in X will have a β_0 effect on Y. But the next period, the same unit change will also have a β_1 on Y and so on. An intermediate effect of this model could be after 3 time periods. The intermediate effect of one unit change in X will then be $\sum_{i=0}^3 \beta_i$. The long run effect would be exactly the same, but it will be the sum of all the coefficients in the model, here until β_6 .

Including this many lags, or even more in a model is inconvenient, and it will also reduce the degrees of freedom. A better solution might be to include a lag of the dependent model, thus making it an autoregressive distributed lag model.

4.6.1.2 Short- and long-run effect in dynamic models

When adding a lag of the dependent variable in the regression, the intermediate and long-run effects become somewhat different. The lag contains the previous impact or impacts of a change in the explanatory variable. All of our FI based indicators are modeled the same way, so let us use this to illustrate:

Equation 4.6-5

$$GDP_{t+1} = \alpha + \beta_0 FI_t + \beta_1 GDP_{t-1} + u_t$$

Though short-run impact usually is defined as the immediate impact of a regressor on the regressand, we argue that in our case the short run impact is represented by β_0 in *Equation 4.6-5*, as this is the first impact our explanatory variable has on the regressand. As we try to forecast GDP it makes sense to address this as the short term. Two quarters from now, the GDP affected by the present FI will be an explanatory variable. E.g. the intermediate effect of a change in the FI will be introduced. Special in our case is that the first lagged GDP is two periods behind the dependent GDP. That is, after one period, we will not get an intermediate effect. When calculating the intermediate effect, we must take this into account that. So if we want to see the intermediate effect of a change in FI e.g. 20 periods into the future, the h in *Equation 4.6-6* must be 19.

Equation 4.6-6

$$\beta_0 \sum_{i=0}^h \beta_1^i$$

The long-run response of a percentage change in our FI takes it one step further. Again $i = 0$ is at time $t+2$.

Equation 4.6-7

$$\beta_0 \sum_{i=0}^{\infty} \beta_1^i = \frac{\beta_0}{1 - \beta_1}$$

4.7 Model specification

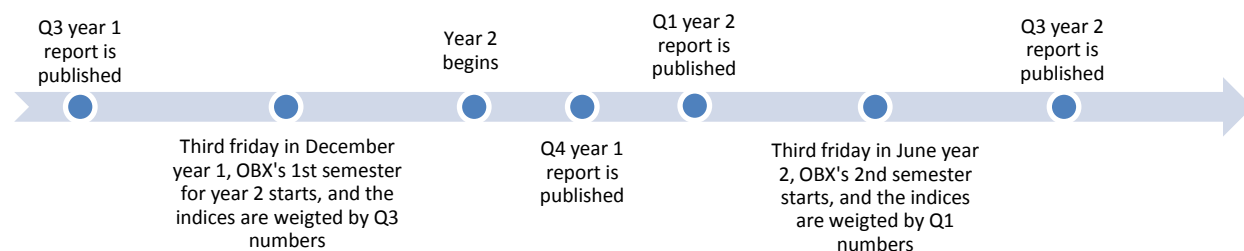
To specify our model we have followed the specification criteria of Hendry and Richard (1983). They say that an empirical model should satisfy the following:

1. *Be data admissible – Saying that the model must be logically possible*

To specify our leading indicator, there are a number of data availability questions that have to logically add up. First, in building our fundamental indices; our FI, based on the OBX, includes and

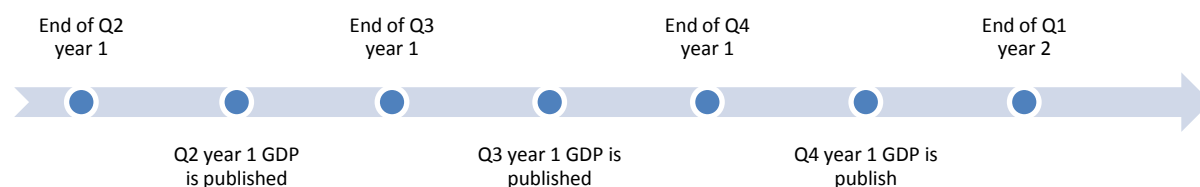
excludes companies, and reweigh the index semi-annually, the third Friday in June and December. This means that the most recent fundamental numbers would be Q1 and Q3 numbers accordingly for an index based on quarterly reported numbers. For the period when we only have Q4 numbers we use the latest reported numbers. E.g. for the reweighing in December 1999 we use Q4 numbers from 1998, and for the reweighing in June 2000 we use 1999 Q4 numbers.

Figure 4.7-1



As with company data the quarterly GDP is published a while after the period ends. Norwegian GDP is published through the Norwegian quarterly national account, which is published about 50 days after the end of the quarter. When creating our model, the last available GDP is at any time, t , the $t-1$ numbers.

Figure 4.7-2



Several papers on leading indicators that use lagged GDP to forecast GDP uses numbers from time t to forecast GDP at time $t+1$. If they are not assuming to be in time $t+\sim 50$ days when forecasting time $t+1$ this will be a violation on the criteria that the data must be admissible.

2. The model must make economic sense

The intuition behind our model is fairly straight forward; GDP is a measure of economic activity and so is a stock index. We believe that a stock index weighted on current fundamental data, i.e. current state, and not an average of possible future states, will better than a MCAP weighted index describe the present, and one quarter forward state of the gross domestic product. Using market prices, we obtain much of the expectations in the market, and can use frequently updated data.

3. Be weakly exogenous regressors.

Regressors are said to be weakly exogenous if they are uncorrelated with the error term. Strictly exogenous regressors are in addition uncorrelated with both past and future values of the error term.

4. Exhibit parameter constancy

The estimated parameters should not have large variations when changing which period that is used to estimate the model. We will use the Chow test to check for a significant structural break in the period.

5. The residuals from the model must be white noise

This is a classical OLS assumption and is covered in *section 4.1*

6. Other models should not be an improvement the chosen model

This is a question of model specification or mis-specification. Finding the best model for the indicators proposed in this paper, and other leading indicators will be subject to definitions and preferences. Which selection criterion is to be used? Akaike, Schwartz, and so on. These are criteria for the best model and will be discussed later in this section.

(Hendry & Richard, 1983)

4.8 Model misspecification

When specifying the model there are several errors that can be committed that will make the model falsely specified. Some of them typically are:

- Errors of measurement
- Underfitting
- Overfitting
- Wrong functional form
- Incorrect specification of the stochastic error term
- Assumption that the error term is normally distributed

When we have in mind a correct or “true” model that for some reason has been modeled incorrectly, it is called a *model specification error*. The four first bullet points are often associated with such errors. When forecasting, or explaining GDP, there is no known correct model, thus we often talk about model mis-specification errors, which are associated with the latter two bullet points. In any regression all these errors can have a severe impact on the results.

4.9 Errors of measurement

Both GDP and our FIs are subject to possible estimation errors, and errors in the underlying data. This may have an impact on the validity of our leading indicators. For more details on these errors we refer to the section dedicated to GDP and construction of the FI.

4.9.1 Dependent variable

As described in the section about Norwegian GDP, there might be errors in their estimations. The impact of errors in the dependent variable is described using the example of Gujarati and Porter (2009):

Equation 4.9-1

$$Y_i^* = \alpha + \beta X_i + u_i$$

Y_i^* = True Norwegian GDP

X_i = Explanatory variable

u_i = Error term

Since Y_i^* is not directly observable, we use the estimate \hat{Y}_i which is the GDP published in the quarterly national accounts.

Equation 4.9-2

$$\hat{Y}_i = Y_i^* + \varepsilon_i$$

So we do not estimate Y_i^* but \hat{Y}_i and therefore we are estimating the equation:

Equation 4.9-3

$$\hat{Y}_i = \alpha + \beta X_i + u_i + \varepsilon_i$$

Assuming $E(u_i) = E(\varepsilon_i) = 0$, $cov(u_i, \varepsilon_i) = 0$, $cov(X_i, u_i) = 0$, and $cov(X_i, \varepsilon_i) = 0$, which are classical assumptions for error terms, the β estimate and its variance will be unbiased. But as the error term now is $u_i + \varepsilon_i$, the variance of the parameters will be larger than in the case of no measurement error.

4.9.2 Explanatory variables

Now, let us assume that the dependent variable, here GDP, is correctly estimated, but the explanatory value has measurement errors. To show how this will affect the leading indicator, we will again use an example given by Gujarati and Porter (2009):

Equation 4.9-4

$$Y_i = \alpha + \beta X_i^* + u_i$$

Y_i = Norwegian GDP

X_i^* = True explanatory variable

u_i = Error term

Since X_i^* contains underlying estimates we can assume that

Equation 4.9-5

$$\hat{X}_i = X_i^* + w_i$$

This means that we are actually estimating the equation:

Equation 4.9-6

$$Y_i = \alpha + \beta(\hat{X}_i - w_i) + u_i$$

$$Y_i = \alpha + \beta\hat{X}_i + (u_i - \beta w_i)$$

$$Y_i = \alpha + \beta\hat{X}_i + z_i$$

Assuming $E(z_i) = 0$

$$\text{cov}(z_i, X_i) = E[z_i - E(z_i)][X_i - E(X_i)]$$

$$= E(u_i - \beta w_i)(w_i)$$

$$= E(-\beta w_i^2)$$

$$= -\beta\sigma_w^2$$

This shows that the explanatory variable and the error term are correlated. This will make the estimates biased, and it can be shown that they are also inconsistent. Measurement errors in explanatory variables may be a serious problem. As the fundamental data are estimated for only a few companies, never more than one company per semester, and usually only for smaller companies, the error in the weights will not contribute to a significant error in the Fundamental indices. As we believe that the measurement error in the FI will be very small, we will assume it to be equal to zero.

4.9.3 Underfitting and overfitting

In the following sections we will present the consequences of model specification errors. We will mainly use the examples given by Gujarati et.al. (2009).

4.9.3.1 Underfitting

The true model is defined as

Equation 4.9-7

$$Y_t = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

The model is wrongly specified as

Equation 4.9-8

$$Y_t = \alpha_1 + \alpha_2 X_{2i} + v_i$$

A number of different problems may occur when this happens:

1. If $\text{Corr}(X_3, X_2) = 0$ is α_2 would be unbiased, but the constant term α_1 will be biased.
2. If $(X_3, X_2) \neq 0$, the coefficients α_1 and α_2 would become correlated, and because of that both biased and inconsistent.
3. The disturbance variance σ^2 is incorrectly estimated
4. The conventionally measured variance of $\hat{\alpha}_2 (= \frac{\sigma^2}{\sum x_{2i}^2})$ is a biased estimator of the variance of the true variable.

Point number 3 and 4 has further consequences. The confidence intervals and then of course the hypothesis testing will be misleading, and forecasts confidence intervals, what we call prediction intervals, will also be unreliable.

4.9.3.2 Overfitting

Let's assume that the true model is

Equation 4.9-9

$$Y_t = \beta_1 + \beta_2 X_{2i} + u_i$$

The model is wrongly specified as

Equation 4.9-10

$$Y_t = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + v_i$$

I.e. the X_3 variable does not belong in the model.

The estimated coefficients α_1 , α_2 , and α_3 are all blue, but inefficient. The variances of these estimated parameters are larger than of the true model.

4.10 Model selection and comparison:

Choosing the best model is, as mentioned earlier, a matter of preferences. For in-sample forecasting, there is a number of selection criteria that can be used, where the most basic is R^2 . The R^2 is defined as:

Equation 4.10-1

$$R^2 = 1 - \frac{RSS}{TSS}$$

By definition R^2 is between 0 and 1, and the closer the R^2 is to 1, the better the fit of the model. A flaw of this measure of fit is that it cannot fall as variables are added to the model. By adding more variables, the R^2 may increase, but so may variance of the forecast error. To account for this flaw multiple information criteria have been developed to penalize the use of extra variables. Two of the most known are the Akaike, and the Schwartz information criterion.

The Akaike Information Criterion (AIC) imposes a penalty for adding extra regressors, and is calculated by

Equation 4.10-2

$$AIC = e^{2k/n} \frac{RSS}{n} \leftrightarrow \ln AIC = \frac{2k}{n} + \ln \frac{RSS}{n}$$

Where k is the number of parameters, and n is the number of observations. An important advantage of the AIC over R^2 , in addition to imposing penalties for extra regressors, is that it is useful for both nested and non-nested models.

The Schwartz Information Criterion (SIC or SBC) is in same spirit as the AIC, but it imposes harsher penalties for including extra variables, as we can see from its definition:

Equation 4.10-3

$$SIC = n^{k/n} \frac{RSS}{n} \leftrightarrow \ln SIC = \frac{k}{n} \ln n + \ln \frac{RSS}{n}$$

As with the AIC, this selection criterion can be used to compare models in and out of sample.

We have used these criteria to help specify our models, but when we have compared the performance of the models, both in and out of sample, we have compared the root mean squared errors (RMSE) in sample and the root mean squared forecasting errors (RMSFE) out of sample.

Equation 4.10-4

$$RMS(F)E = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i^2}{n}}$$

This measures the size of the error in the predictions and forecasts. We believe that the most important criteria when comparing the leading indicators is how close the predictions are to the true values, and that is why we rank the models according to RMS(F)E.

4.11 Constructing the fundamental indices

When Research affiliates created their fundamental index they find the 1000 largest companies of each metric they use to weigh their indices, and not simply use Russell 1000 or S&P 1000. If they chose to use the companies in these indices they would end up with using companies with high capitalization. 1000 companies are out of scope for this thesis. This is why we use the OBX which consists of the 25 most liquid stocks, and not necessarily the largest companies. Though many of the companies in the OBX are among the largest on the Oslo Stock Exchange, we argue that the companies in OBX are also highly liquid because their current importance for the country's economy. The number of companies and how they are selected is why we chose OBX as the index we want to reweigh. The fundamental metrics we have used are as follows:

- Earnings before interests, taxes, depreciation and amortization (EBITDA)
- Sales/Revenues
- Total Assets
- Number of Employees
- Dividends

We have chosen these fundamental data as our basis of our indices because they are well known as size and financial condition measures also outside the financial world, but as Dr. Denis Chaves from Research Affiliates wrote to us: *it is not that important which fundamental data that is used, as long as it breaks the relationship between size and market capitalization.*

4.11.1 Special notes

4.11.1.1 Composite index

In the composite index we will, as Research Affiliates do in their composite index, not use number of employees. This is because a number of companies on the OBX are holding companies or similar and the number of employees in the holding company does not represent the real size of the company.

4.11.1.2 EBITDA

EBITDA is not a relevant measure for financial companies, therefore we have used pretax profit for these companies when constructing the index that are weighted by EBITDA. We will from this point on only refer to this as EBITDA. When EBITDA is negative, we have used the average weights of the Sales index and the Total Assets index. If we would use the negative EBITDA when weighing this index, we would get short positions. This would have a positive impact on the index if the share price drops, and would be relevant if the index is used as an exchange traded fund (ETF), but not in our case where we seek to forecast the growth of the overall economic state of a country.

4.11.1.3 Dividend

The treatment of the dividend index is somewhat special. Because not paying dividends is not necessarily a sign of weakness, but can simply be for example tax-reasons, we have treated non-dividend paying companies in the same way as Research Affiliates. That is, for non-dividend paying companies their weight is the average weight of the other indices (excluding employees). We recognize that we may get average of averages here because some non-dividend paying companies may also have a negative EBITDA. As with the other indices, we use total numbers (gross dividends) and not dividend per share.

4.11.1.4 Capping rules and index updating

Our indices follow the OBX capping rules. The largest company's weight can maximum be 30 % of the index, and the second largest can be maximum 15 %. We have not followed the capping rules saying that the aggregate weight of non-EEA companies can maximum be 10 %, as this rule was implemented in 2012. When a company's weight is capped, the remaining weight is distributed to the other companies according to their weights.

As the OBX we update our indices semi-annually, and implement the changes the third Friday of June and December. In contrast to the OBX, our indices' weights are kept constant over the period. The sample period is from 1996 (when the OBX started to adjust for dividends) until 2011. A longer period would be preferable, but the OBX would be inconsistent and the data availability more limited.

4.11.1.5 Seasonality

We assume that stock indices on average become seasonally "adjusted," therefore we chose to use seasonally adjusted GDP.

4.12 Data

4.12.1 Data gathering

We have used Datastream as our main source for both company data and stock prices. We used adjusted close prices, and last reported fundamental data, and therefore the fundamental data had to be leaded once, so that the numbers from Datastream match the reporting period.

Some companies, like Nycomed Amersham where Datastream do not provide data and annual reports are no longer available, are excluded from the indices. In the case of Nycomed Amersham, the company merged, demerged, was bought, and so on, many times. There are no available annual reports this far back, as it is not required by law. From 2001 and backwards, Datastream only provides the fourth quarter numbers. So in this period we were forced to use this number, we argue that this will not have a large impact on our indices as this will be the same for all companies. As explained, the OBX includes and excludes companies with every reweighing. This means that the index consists of different companies each semester. Oslo Stock Exchange provided us with the list of which companies that was in the index when, and the historical OBX values.

4.12.2 Data transformation

The GDP series, the fundamental indices, and most of the comparable indices are transformed into returns. This made all the time series stationary, and the models easy to interpret. The time series that is not transformed into percentage changes is the OECD CLI, the PMI, and the interest rate spread.

When presented graphically, we converted the returns into accumulated growth with $12/31/1995 = 1$.

4.12.3 Flaws in the data

There are some flaws in our fundamental data, and this will affect our regressions. We explained shortly how in *section 4.9.2*. A flaw in our data occurs when we have missing fundamental data. We have in some instances made a simple interpolation to estimate the missing information. This will possibly give small errors in the weights, which in turn will affect our FI. For flaws in GDP we refer to *section 2.2.4*.

5 Comparable Leading indicators

Financial variables may be leading indicators for GDP because they are; priced based on expectations of the future, influence the economy with a lag, or if they are published more frequently and earlier than GDP. These are the criteria we used when selecting which indicators to compare to. We found house prices, stock prices, credit growth, growth in money supply, and the short term vs. long term interest rate spread to be possible single indicators for change in GDP (Gerdrup, et al., 2006).

5.1 Single indicators

5.1.1 OBX index

As we mentioned earlier, we will use the OBX index as a representative for MCAP indices. For more information about the OBX index, see *section 2.4*, and as for why this could be a good leading indicator see *section 2.3*

5.1.2 Interest rate spread

Several empirical papers indicate that an inverted interest rate yield curve might be able to work as a leading indicator for future recessions, see for example Stock and Watson (2001)⁵². Inverted yield curves result from negative expectations of market participants about the future economic growth potential. If short term interest rates are expected to increase because of higher inflation, as opposed to future economic growth, the interest rate spread will be weakened as a leading indicator for change in the real economy (Gerdrup, et al., 2006). Like Gerdrup et al. (2006) we used the spread between the three months Norwegian Interbank Offering Rate (NIBOR) and the five year Norwegian Treasury bond for predicting change in the GDP. The spread between the three month NIBOR and five year Treasury bond was chosen because in efficient financial markets long term interest rates will reflect the market participants' expectations of future short term interest rates. These expectations are influenced by expectations of future economic growth and inflation.

Long term interest rates can also be influenced by risk-premiums. Owning long term bonds introduces the risk that the real return will be lower than expected if for instance inflation turn out to be higher

⁵² James H. Stock, Mark W. Watson, 2001, *Forecasting Output and Inflation: The Role of Asset Prices*. National Bureau of Economic Research, Working Paper 8110.

ex post than ex ante. This means that long term interest rates may incline when investors are more uncertain about future development in for example the inflation. This may also weaken the correlation between the interest rate spread and future production (Gerdrup, et al., 2006).

5.1.3 Money supply

As a measure of money supply we chose to use M1 which is the money-holding sectors, public and financial enterprises other than banks and state lending institutions, stock of Norwegian notes and coins, plus the sector's deposits in transaction accounts in the Norwegian National Bank (Norges Bank) and commercial and savings banks. Bank deposits in transaction accounts include deposits that are readily converted to notes and coins or payments can be made directly without incurring any costs other than normal transaction and arrangement fees. M1 is described as the narrow money supply term (Norges Bank, 2011).

The change in money supply is registered every month, hence, it can provide information about real economic development earlier and more frequently than the national account. Consequently, changes in money supply may be a good leading indicator for the GDP. However, whether the money supply provides any additional information to what is given by changes in the credit growth is uncertain.

The quarterly changes in money supply are used to predict the quarterly change in GDP in this thesis.

5.1.4 Credit indicator

Credit indicator K3, from here on only referred to as K3, is an important indicator for the total gross debt of local government, non-financial corporations and households to all domestic and foreign creditors. It is an approximation, and not an exact number of how large the total debt is, as the foreign debt also depends on change in the exchange rate. K3 is published about one month after the K2 which only contains debt from domestic creditors. As the OBX contains international companies with foreign debt, cash, and so on, we assume K3 to have a better foundation to compare with our FI. This may not be correct as the goal of the credit indicator is to predict GDP, but we argue that for the Norwegian economy overall, foreign debt should be included as we are an open economy (Statistics Norway, 2012).

If corporations increase their credit this will be registered in the credit statistics, and it may provide indications about the development in the corporations' real investments before impacts are observable in the national accounts. It is also reason to believe that the corporations to some degree are granted credit before larger real investments are transacted, hence, they may contain leading information about the development in the real economy (Gerdrup, et al., 2006).

If the growth in the corporations' real credit is motivated by other factors than real investments, like purchasing domestic goods, it weakens K3s predictive power to GDP.

5.1.5 House Price Index

The demand for houses depends, amongst other things, on the households expectations on the future development in the economy. Normally it takes time to increase the total supply of houses through building new ones; therefore increased demand will immediately lead to increased house prices, hence it might be used as a leading indicator for GDP. It can also amplify the real economic development through other canals, like increased access to credit (Gerdrup, et al., 2006).

5.2 Comparable indices

5.2.1 Norwegian Purchasing Manager Index

Norwegian Purchasing Manager Index (PMI) measures change in the activity in the Norwegian industry based on monthly surveys from a fixed sample of about 300 purchasing managers in the manufacturing industry. The PMI is a combination of five different sub-indices with direct relations to purchasing. Its goal is to give a quick conjuncture indication for the Norwegian economy, and therefore seems suitable to indicate economic prosperity. Fokus bank and the Norwegian association for purchase and logistics (Norsk forbund for Innkjøp og Logistikk, NIMA) created the index in 2004. The idea of the index is equivalent to the conjuncture indicator in the US; The Industrial Purchasing Manager Index.⁵³

⁵³ Fokus Bank, <http://www.fokus.no/nb-no/Bedrift/Store-bedrifter/Markets/Pages/Norsk-PMI.aspx>.

5.2.2 OECD Composite Leading Indicators

For a description of the OECD CLI, see OECD Composite Leading Indicators under the section Comparable leading indicators. Here we will explain shortly why we expect that the OECD CLI will be able to predict one quarter change in GDP.

The OECD CLI should be a top performer as it is constructed from six different component series, picked especially for their similar, preferably preceding, fluctuations of the Norwegian business cycle, represented by GDP. OECD CLI and PMI are included in order to check if our FIs, which are single indicators, can compete with more complex models that include multiple indicators. One possible weakness is that we use it outside of its original context of indicating changes in business cycles in GDP, and use it to forecast next quarters change.

6 Results

6.1 In sample predictions

6.1.1 Our fundamental indices

Out of the competing regressions, the best results have been achieved for each of our fundamental indices using the Linear Transfer Function identification method described in *section 4.6*. The best way for our fundamental indices to predict the Norwegian GDP was the same for all of the different sub-indices and also the composite index. To illustrate the chosen regression we present the Dividend-model:

Equation 6.1-1

$$GDP_t = \alpha + \beta_1 DIV_{t-1} + \beta_2 GDP_{t-2} + \varepsilon_t$$

6.1.1.1 Autocorrelation

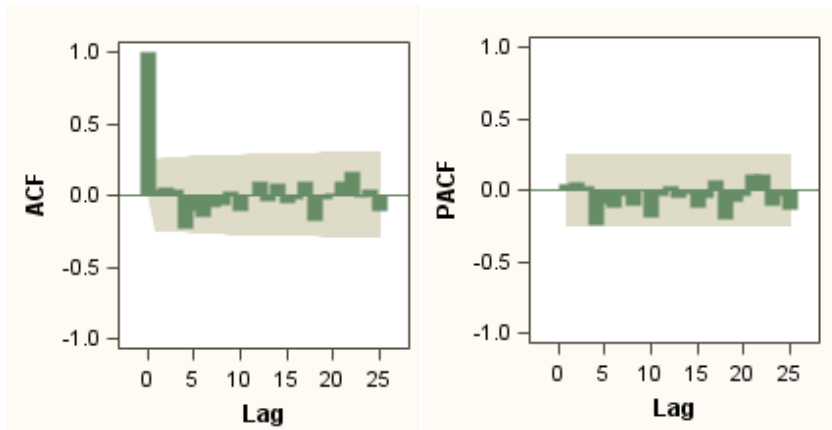
To check for autocorrelation we used the Breuch-Godfrey (BG) test, and none of our FIs showed any sign of autocorrelation. In the table below we have presented the BG test results of the Dividend FI based indicator. The remaining results can be found in the Appendix A.

Table 6.1-1

Godfrey's Serial Correlation Test		
Alternative	LM	Pr > LM
AR(1)	0.0127	0.9102
AR(2)	1.1032	0.5760
AR(3)	1.1252	0.7710
AR(4)	3.7350	0.4431

We also did a visual check for autocorrelation through the ACF and PACF. Here we saw some borderlines, but none of them were significant, and there were no significant signs of seasonality. We have added the ACF and PACF correlograms from Sales (Revenue) as an example in the text, the other FIs exhibited a similar pattern.

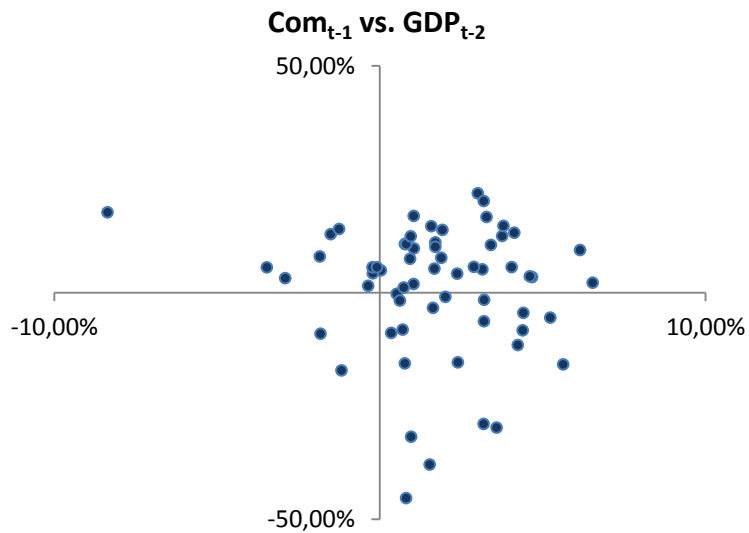
Figure 6.1-1



6.1.1.2 Multicollinearity

When testing for multicollinearity we created scatter plots of the regressors. For all our FI indicators we found that the scatter plots resemble white noise. The scatter plot in *Figure 6.1-2* shows our Composite FI-models growth on the Y-axis and the lagged GDP growth on the X-axis.

Figure 6.1-2



The graph might suggest negative correlation, which is confirmed by the numerical results of -0.095. The FI based indicator that exhibits the most multicollinearity is the total assets weighted index, with a correlation between the regressors of -0.181. Due to the low impact we do not expect multicollinearity to be a problem in our regressions.

6.1.1.3 Normality

The normality of the residuals was tested using the Jarque–Bera test, and shows no sign of incorrectly distributed residuals. The FI based indicator with the most significant Jarque-Bera was the number of employees based FI indicator, which had a p-value of 0.22, but none of the other FI indicators had a lower p-value than 0.5. The Q-Q plots indicate that some of the FI based indicators suffers from large tails, but this is not clear, and we will assume that for all of the FI based indicators the residuals are normally distributed. The Q-Q plots and the Jarque-Bera results can be found in Appendix B.

6.1.1.4 Heteroscedasticity (ARCH)

To check for ARCH disturbances we used the Lagrange Multiplier. We did not experience any significant values, hence there does not seem to be any volatility clustering in the residuals from our FI-models. An example from our total asset model is seen below, and yet again our other FIs produced similar results.

Table 6.1-2

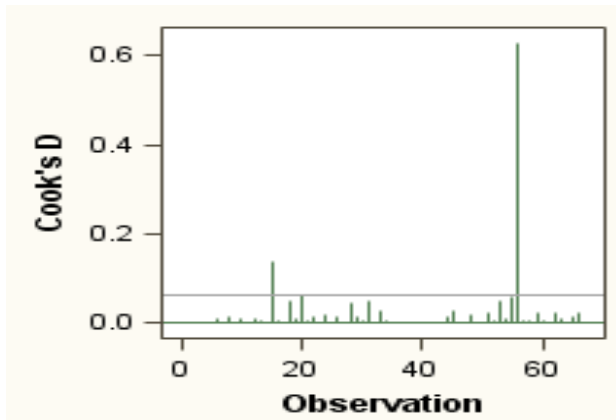
Tests for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	0.0867	0.7685	0.0690	0.7929
2	0.4084	0.8153	0.2897	0.8652
3	1.9434	0.5842	1.8615	0.6016
4	1.9568	0.7437	1.9493	0.7451
5	2.1302	0.8308	2.4449	0.7848
6	2.1825	0.9022	2.5238	0.8658
7	2.6334	0.9167	2.7440	0.9076
8	2.8133	0.9455	2.8909	0.9410
9	4.3951	0.8835	4.0889	0.9055
10	4.5128	0.9213	4.0904	0.9432
11	4.5721	0.9501	4.1028	0.9668
12	4.5957	0.9701	4.3350	0.9766

6.1.1.5 Outliers

When looking at the Cook's D, there are one or two outliers with significant influence on the coefficients. One of those, and the largest by far, seem to stem from the period of the financial crises in 2008/2009. Due to the small number we assume the overall impact to be negligible, hence chose

not to correct for this. But, if necessary, we could have included a dummy variable to account for this , we describe these results in later sections. We have included an example from our Dividend model here, as for the rest of our FIs they exhibit a similar pattern.

Figure 6.1-3



6.1.1.6 Model specification

As we have stated, we use Ramsey's RESET test to check if the models are correctly specified. The indicators based on total assets and gross dividend returns significant results on a 90 % level in the 2nd and 3rd power respectively. None of the other FI based indicators returns significant RESET results. Through this thesis we operate with a 95 % significance level and we will conclude that our FI based indicators are correctly specified. Below is the Ramsey's RESET test result of the total assets model. The rest can be found in Appendix C.

Table 6.1-3

Ramsey's RESET Test		
Power	RESET	Pr > F
2	3.3117	0.0740
3	2.2734	0.1122
4	1.5768	0.2052

6.1.1.7 Coefficients

Our results showed positive coefficients, as expected, meaning that an increase in one of our FIs will lead to an increase in the GDP. To illustrate the effect we will give a simplified example from the Dividend-model here where the $\beta_1 = 0.1036$. This means that we expect the short-run effect on GDP

of a 1 percentage point increase in DIV_{t-1} is an increase of 0.1036 percentage points in GDP. The lagged GDP coefficient can be used to measure how a change in the FI will affect the GDP in the long run. The long-run effect on GDP from a percentage point change in the dividends based FI will be.

Equation 6.1-2

$$0.1036 \sum_{i=0}^{\infty} 0.2748^i = \frac{0.1036}{1 - 0.2748} = 0.1429$$

Table 6.1-4

Model	R ²	RMSE	Intercept	FI t-1	GDP t-2	Long-run effect
Dividends	0.4007	0.0196	0.0106***	0.1036***	0.2748***	0.1429
EBITDA	0.3999	0.0197	0.0107***	0.1073***	0.2686***	0.1467
Composite	0.3816	0.0200	0.0107***	0.1020***	0.2785***	0.1414
Sales/Revenue	0.3306	0.0208	0.009***	0.0966***	0.2941***	0.1368
Total Assets	0.311	0.0211	0.0082**	0.0911***	0.3172***	0.1334
Employees	0.2848	0.0215	0.0096***	0.0764***	0.2966***	0.1086

*, **, *** represents the level of significance, 90, 95, and 99% respectively

6.1.2 Model comparison

In Table 6.1-4 above the different models are ranked according to the best fit in our in-sample predictions according to RMSE. The Dividend model is our best one, but only marginally better than the EBITDA-model. It explains 40.07% of the variation, and has a RMSE of only 0.0196

6.1.3 Comparable leading indicators

We used the same Linear Transfer Function method as we did for our FIs to find the best model for each of our comparable leading indicators. This way we test our leading indicators against their best peers. In Table 6.1-5 below you find the best model for each of the different comparable leading indicators.

Table 6.1-5

Leading Indicator	Preferred model
OBX	$GDP_{t+1} = \alpha + \beta_1 OBX_t + \beta_2 GDP_{t-1} + \varepsilon_t$
CLI	$GDP_{t+1} = \alpha + \beta_1 CLI_t + \beta_2 CLI_{t-1} + \beta_3 GDP_{t-3} + \varepsilon_t$
M1	$GDP_{t+1} = \alpha + \beta_1 M1_t + \beta_2 M1_{t-1} + \beta_3 GDP_{t-1} + \beta_4 GDP_{t-3} + \varepsilon_t$
HPI	$GDP_{t+1} = \alpha + \beta_1 HPI_t + \beta_2 GDP_{t-1} + \beta_3 GDP_{t-3} + \varepsilon_t$
Spread	$GDP_{t+1} = \alpha + \beta_1 SPR_{t-1} + \beta_2 SPR_{t-2} + \beta_3 SPR_{t-3} + \beta_4 GDP_{t-1} + \beta_5 GDP_{t-3} + \varepsilon_t$
K3	$GDP_{t+1} = \alpha + \beta_1 K3_t + \beta_2 K3_{t-1} + \beta_3 GDP_{t-1} + \beta_4 GDP_{t-3} + \varepsilon_t$
PMI	$GDP_{t+1} = \alpha + \beta_1 PMI_{t-1} + \beta_2 GDP_{t-3} + \varepsilon_t$

Table 6.1-6

	R ²	RMSE	Intercept	B ₁	B ₂	B ₃	B ₄	B ₅	Long-run effect
CLI	0.462	0.019	-0.442 ^{**}	0.029 ^{***}	-0.024 ^{***}	-0.195 ^{***}			0.004
PMI	0.275	0.024	-0.064 [*]	0.002 ^{**}	-0.421 ^{**}				0.001
OBX	0.354	0.020	0.008 ^{**}	0.103 ^{***}	0.285 ^{***}				0.143
M1	0.294	0.022	0.008 [*]	0.214 ^{***}	0.230 ^{***}	0.247 ^{**}	-0.345 ^{***}		0.405
HPI	0.200	0.023	0.011 ^{**}	0.234 ^{**}	0.276 ^{**}	-0.288 ^{**}			0.231
Spread	0.251	0.023	0.018 ^{***}	-0.013 ^{***}	0.020 ^{***}	-0.011 ^{**}	0.277 ^{**}	-0.370 ^{***}	-0.004
K3	0.132	0.024	0.014 [*]	-0.042	0.155	0.274 ^{***}	-0.314 ^{***}		0.109

^{*}, ^{**}, ^{***} represents the level of significance, 90, 95, and 99% respectively

The long-term effects of the different indicators are calculated the same way as with the FI based indicators, though some are more complex. We will illustrate with the Spread model as this is the most complex. The long term effect in a change in the Spread is:

Equation 6.1-3

$$= \frac{(Spread_{t-1} + Spread_{t-2} + Spread_{t-3})}{(1 - GDP_{t-2} - GDP_{t-4})}$$

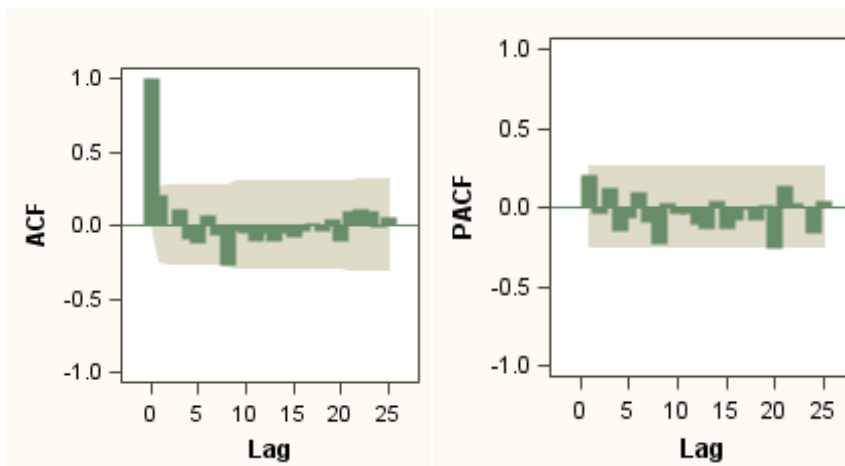
6.1.3.1 Autocorrelation

The BG test indicates autocorrelation in the K3 model in the fourth lag, see *Table 6.1-7*, one can also see that there are some borderline spikes in the ACF and PACF correlogram, see *Figure 6.1-4*, but none are significant. The test showed no signs of autocorrelation in the rest of the models. BG results for all of the comparable indicators can be found in Appendix A.

Table 6.1-7

Godfrey's Serial Correlation Test		
Alternative	LM	Pr > LM
AR(1)	2.6216	0.1054
AR(2)	2.6409	0.2670
AR(3)	3.8825	0.2744
AR(4)	9.0102	0.0608

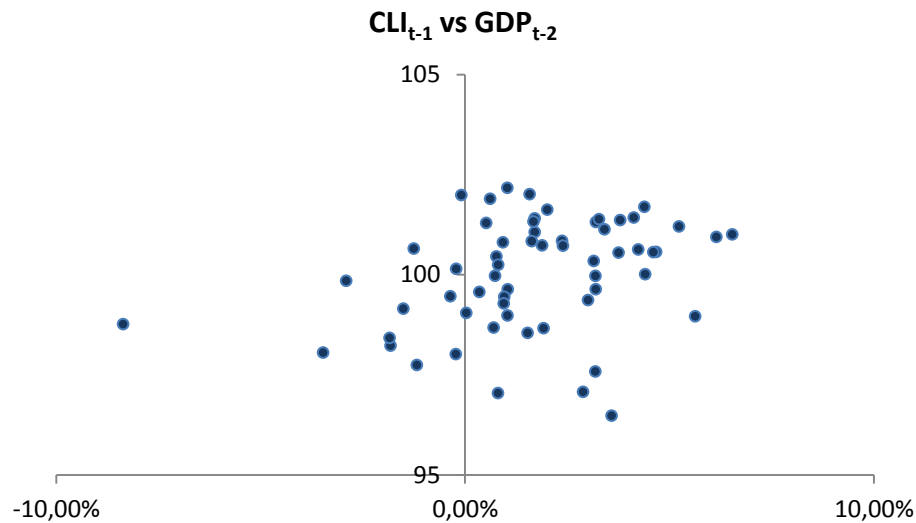
Figure 6.1-4



6.1.3.2 Multicollinearity

Among the comparable indicators we will focus on the CLI model. The multicollinearity in this indicator is a lot higher than what we found in our FIs. Between the regressors shown in the scatter plot the correlation is 0.310. Between CLI_{t-2} and GDP_{t-2} the correlation is 0.393.

Figure 6.1-5



There are some indications of correlation between the regressors in the models that can lead to higher standard deviations for our coefficients. The correlation between lags of GDP and other regressors is relatively low, but between non-GDP regressors and their lags the correlation is in most instances higher. The correlation between CLI_{t-1} and CLI_{t-2} is at 0.918, this is very high, but the coefficients in this indicator do not suffer from high standard deviations, so we have not adjusted the model for this.

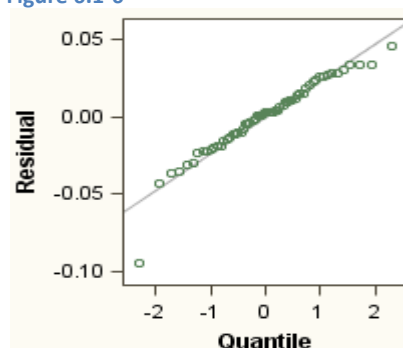
6.1.3.3 Normality

Most of the comparable indicators have normally distributed residuals, but the K3 model is skewed to the right. The BG test is very significant and the Q-Q indicates skewness:

Table 6.1-8

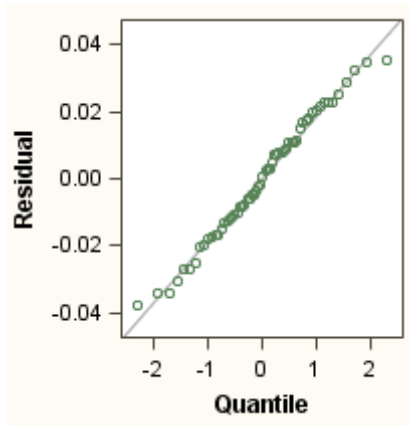
Jarque-Bera			
Statistic	Value	Prob	Label
Normal Test	28.1234	<.0001	Pr > ChiSq

Figure 6.1-6



The rest of the indicators return non-significant Jarque-Bera results, but some Q-Q plots indicate some skewness or kurtosis, like the CLI:

Figure 6.1-7



All of the Q-Q plots with Jarque-Bera results can be found in Appendix B.

6.1.3.4 Heteroscedasticity (ARCH)

The only comparable indicator that exhibits ARCH effects is the CLI model. It is significant at a 90 % level in the fourth order.

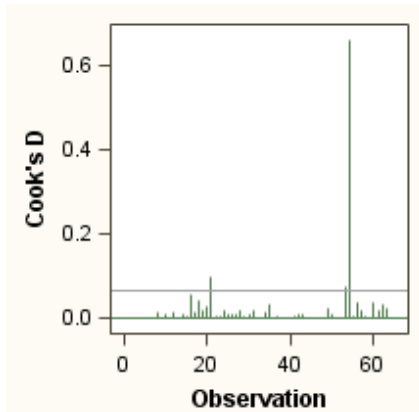
Table 6.1-9

Tests for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	0.0249	0.8746	0.0000	0.9988
2	0.7073	0.7021	0.0722	0.9646
3	15.8063	0.0012	5.5796	0.1340
4	26.4204	<.0001	8.3071	0.0810
5	26.4377	<.0001	8.5616	0.1279
6	29.9747	<.0001	8.6563	0.1938
7	41.9405	<.0001	9.7228	0.2048
8	42.0275	<.0001	9.7307	0.2844
9	42.6375	<.0001	9.9937	0.3510
10	49.9487	<.0001	10.0936	0.4323
11	51.1424	<.0001	11.6975	0.3868
12	51.1629	<.0001	12.2267	0.4276

6.1.3.5 Outliers

With respect to the Cook's D, we see a trend in all of our analyses, namely that there is a highly influential residual outlier around the time of the financial crises. For the OBX-model, we also see a few others around the same time. In the rest of the models we also see a significant, but much smaller spike in the beginning of our dataset. As we did not correct for this in our FI models we chose not to do so here either by the same reasoning. Below you can see an example from the HPI-model.

Figure 6.1-8



6.1.3.6 Model specification

Two of the comparable indices exhibit significant Ramsey's RESET test results. The OECD CLI based indicator has a significant RESET result in the fourth power. This result is presented in the table below, the rest can be found in Appendix C.

Table 6.1-10

Ramsey's RESET Test		
Power	RESET	Pr > F
2	1.0658	0.3064
3	0.9909	0.3779
4	4.3924	0.0078

We ran many different models to address this problem, but the model presented was the best specified, and performed furthermore best according to Schwartz information criterion.

6.1.3.7 Coefficients

The Spread model indicates that changes in GDP are negatively correlated with the interest rate spread from two and four quarters back in time. Therefore, a decrease in the spread between the three month NIBOR and the 5 year Norwegian treasury bonds today indicates that in the short-run the GDP will decrease, but this will shift twice in the intermediate effect. In the long run, we expect that an increase in the spread will decrease the GDP. The positive coefficient surprised us as we expected that an increase in the spread between long-term and short-term interest rates is a sign of increased uncertainty in the market, and therefore should indicate a decline in the GDP. The trouble the spread model experience may be due to changes in inflation during the period. This indicator is also the only with a negative long-run relationship with the percentage point change in GDP. When performing a simple correlation analysis Gerdrup et al. (2006) found a positive relation between all lags of their spread and GDP data (up to eight lags). They conclude that that an interest rate spread model could be a potential leading indicator of Norwegian mainland GDP. According to our results, although we are using total GDP, this is not correct despite the correlation proved in Gerdrup et al. (2006)

The Money supply model was pretty much as expected, as both of the coefficients are positive. This means that a one percentage point increase in the money supply indicates an increase in the GDP in the next quarter. Gerdrup et al. (2006) found a positive relation between M1 and mainland GDP, which is in line with our results, back until 7 lags of M1.

The Credit indicator model did not perform well and we got no significant values in our in sample predictions except for the lagged GDP values. This might indicate that a unit-increase in K3 can in the long-run have a positive impact on GDP, even though the impact is insignificant in the short-run. The K3 coefficients were also inconsistent with a negative one-lagged value and a positive second-lag, but then again, neither of them was statistically significant. We expected a positive correlation between credit growth and the change in GDP as an increase in credit should mean increased money accessible to the public, hence increased spending and increased GDP. There are several possible reasons for this; one might be that we chose a wrong credit indicator as the K3 numbers are released one month later than other credit indicators and therefore might actually be a lagging, and not leading, indicator.

When performing their correlation analysis, Gerdrup et al. (2006) divide K3 into mainland Norway, and mainland companies, and found a negative, and positive, relation respectively. K3, for mainland companies, change to a negative correlation further back in time. This may be an explanation as to why we got non-significant coefficients for the explanatory variables in our K3 model, as this can be a sign of different influences as to change in credit depending on whether it is for households or companies and our K3 number combines them.

The House Price Index (HPI) model indicated a positive correlation between the HPI and GDP. So if there is an increase in the HPI one can expect to see an increase in the GDP as well. Just as we expected; an increase in housing prices might, as we discussed in section 5.1.5, give the households access to more, or cheaper, credit and so on. This is in line with Gerdrup et al. (2006) who found a positive correlation with a one period lagged real house price growth to GDP.

The OBX-model also indicates a positive relationship between the OBX-index and GDP with a positive and significant coefficient for one-lagged OBX changes. It seems like it takes one quarter for the OBX to influence GDP, which was as we expected. This lends support to the idea that stock markets are driven by the market participants' expectations and can change more rapidly on news etc. than GDP does. This, again, is in line with Gerdrup et al. (2006), and even though they look at real stock returns, adjusted by CPI, they find a positive correlation between GDP and stock returns.

The CLI-model also surprised us as the one and two lagged coefficients are positive and negative, respectively, yet highly statistically significant. This could indicate that it is reverting over time.

The PMI-model was somewhat disappointing. This is an index created by five sub-indices and its purpose is to give a quick and readily accessible conjuncture indication on the Norwegian economy. Still, we only got one significant variable, and that was with 2 lags.

6.1.3.8 Comparison of the models

In

Table 6.1-6 we see the models and their fit for predicting GDP. Here the CLI-model is the best by far, also better than all of our FIs. This was as expected as the CLI index, as mentioned earlier, contains six

different leading indicators with a proven long-term relationship with the reference value, GDP. The PMI-index was disappointing as it is specially made for Norway. It has the second lowest RMSE, and a relatively low R^2 . We had expected that the purchasing managers' expectations for the future should be better able to predict economic development, but in its defense it has fewer observations than the rest as it starts in 2004. The OBX-index was the second best comparable leading indicator with respect to both RMSE and R^2 . The K3-model was the worst performer with respect to both RMSE and R^2 , this was very disappointing as we expected that a credit indicator would be able to explain change in GDP well. The HPI-model was also somewhat disappointing; this was the second worst performer, it only outperformed the K3-model, and it could only explain 20% of the variation. The M1 model had a decent R^2 of 29.36%, despite this it outperformed only one out of six of our FIs, and its RMSE was outperformed by all of our FIs respective RMSEs. The most disappointing model was the one for the Spread measure. Interest rates, and especially interest rate spreads based on commercial paper rate and treasury, are described in many papers, see for example Stock and Watson (1989)⁵⁴, as a good predictor of economic activity. Regardless, in our tests it is one of the worst performers. This might be due to model construction, as this is the biggest model containing five explanatory variables, or it could be due to inflation. We found some evidence of model construction error, as it was significant at the 90% level in Ramsey's RESET test. It is the third worst performer in regards to both RMSE and R^2 . But as described in *section 2.3*, the Norwegian economy is quite different than for example the American which is where most of the research are focused, so this may also be an explanation. However, this is also consistent with Bernanke (1990)⁵⁵ who found weakened predictive power for this type of spread measure in the American economy.

6.2 Comparison of our models to a Naïve model

We report the RMSE of each model relative to the Naïve benchmark (RRMSE). A number below one indicates that the model's forecasts are more accurate than naïve model.

⁵⁴ James H. Stock, Mark W. Watson, 1989, *New Indexes of Coincident and Leading Economic Indicators*. NBER Macroeconomics Annual 1989, Volume 4, pp. 351 - 409

⁵⁵ Ben Bernanke, 1990, *On the Predictive Power of Interest Rates and Interest Rate Spreads*. NBER Working Paper No. 3486, Issued in October 1990.

Table 6.2-1

Model	RRMSE	RMSE
CLI	0.7725	0.0190
Dividends	0.8011	0.0197
EBITDA	0.8019	0.0197
Composite	0.8137	0.0200
OBX	0.8320	0.0204
Sales/Revenue	0.8467	0.0208
Total Assets	0.8589	0.0211
Employees	0.8753	0.0215
M1	0.8936	0.0219
Spread	0.9287	0.0228
HPI	0.9421	0.0231
PMI	0.9780	0.0240
K3	0.9906	0.0243
<i>Naïve</i>	<i>1</i>	<i>0.0245</i>

All of the models besides the naïve itself have a number below one, hence, they outperform the naïve model. This was not very surprising to us, as all of the different models include the same explanatory variable as the naïve model, and more. Thus these results might not be that interesting, but when comparing between the different models we find some very interesting results. We find that the CLI-model has the lowest RRMSE with its 0.7725, meaning that this is the best predictor for quarterly changes in the Norwegian GDP, despite that its originally purpose is to predict turning points in the economy. Another peculiar detail is that the OBX-model seems to be better than half of our FIs, Total Assets, Sales (Revenue), and Employees, while the other three, namely our Dividends, EBITDA, and Composite models outperforms it. This is somewhat in line with our hypothesis that a fundamentally weighted index can outperform a market capitalization weighted index as a leading indicator for GDP.

6.3 Out of Sample

When testing the models out of sample we estimated the models by using observations from the previous 10 years. In other words, all of our predictions, from Q2 2007 until Q1 2012 are based on a model estimated by the previous 40 observations. We chose this length of our “rolling” model based on OECD (2012)⁵⁶. Based on the GDP from 1960 to 2011, they found that an average Norwegian

⁵⁶ OECD, March 2012, *OECD Composite Leading Indicators Country reviews*. Pp 11-13.

business cycle has a duration of 60 months. Our model will therefore be based on data from 10 years back, which on average is two business cycles. We chose this “rolling” estimation rather than adding new observations to the existing model, because adding observations will result in less weight given to the most recent observations. This “problem” will not occur with a rolling model, but on the other hand, it can be argued that less than 40 degrees of freedom is on the short side of what is reasonable when making an efficient model. This is also in accordance with Gerdrup et al. (2006).

We used the models found to work best in sample when estimating the out of sample predictions, both for our own FIs and the comparable leading indicators.

Table 6.3-1

	Regression	RMSFE
EBITDA	$GDP_{t+1} = \alpha + \beta_1 EBI_t + \beta_2 GDP_{t-1} + \varepsilon_t$	0.0232
Composite	$GDP_{t+1} = \alpha + \beta_1 COM_t + \beta_2 GDP_{t-1} + \varepsilon_t$	0.0235
Dividends	$GDP_{t+1} = \alpha + \beta_1 DIV_t + \beta_2 GDP_{t-1} + \varepsilon_t$	0.0236
Sales\Revenue	$GDP_{t+1} = \alpha + \beta_1 SAL_t + \beta_2 GDP_{t-1} + \varepsilon_t$	0.0252
Total Assets	$GDP_{t+1} = \alpha + \beta_1 TA_t + \beta_2 GDP_{t-1} + \varepsilon_t$	0.0258
Employees	$GDP_{t+1} = \alpha + \beta_1 EMP_t + \beta_2 GDP_{t-1} + \varepsilon_t$	0.0266

6.3.1 Our fundamental indices

As with our in-sample results the Composite, Dividends, and EBITDA based leading indicators performed the best among our models. In contrast to the in-sample results, the EBITDA model has the lowest RMSFE.

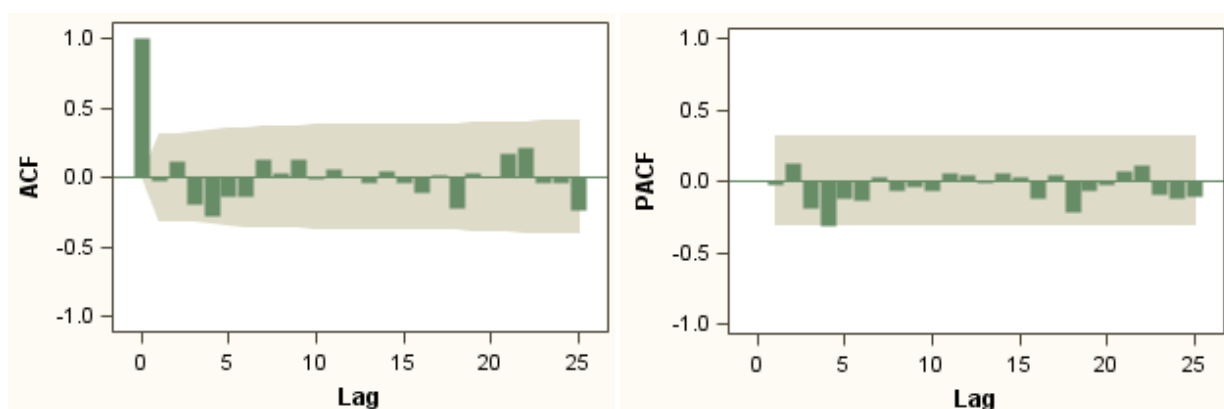
6.3.1.1 Autocorrelation

The BG test results show some evidence of autocorrelation in the dividend model when the estimation include the Q1 2009, which is the heavy outlier that we have discussed earlier. It is only significant at a 90 % level, but before this quarter, there is no indication for autocorrelation. *Table 6.3-2* shows the BG test results for the Dividend model when it is estimated just before the financial crisis, and when the estimation includes it.

Table 6.3-2

Before the crisis			Including the crisis		
Godfrey's Serial Correlation Test			Godfrey's Serial Correlation Test		
Alternative	LM	Pr > LM	Alternative	LM	Pr > LM
AR(1)	0.1391	0.7092	AR(1)	0.0111	0.9163
AR(2)	1.7739	0.4119	AR(2)	2.7870	0.2482
AR(3)	1.7861	0.6180	AR(3)	7.0386	0.0707
AR(4)	3.0677	0.5466	AR(4)	8.7114	0.0687

Figure 6.3-1



These are the ACF and PACF of the Composite model estimated at Q2 2010. As we see, there is one borderline in the PACF correlogram, but there are no signs of seasonality, and in other estimations they are well inside the borders. As a pattern across the leading indicators, it seems like the 4th/5th lag of the ACF and PACF becomes closer to significant when approaching the end of the series.

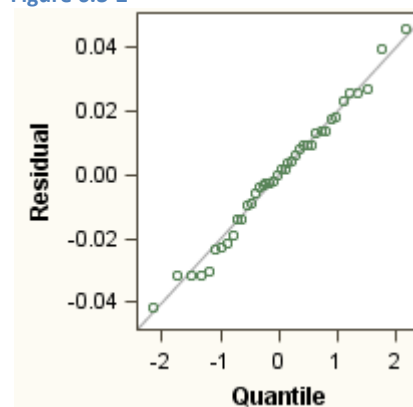
6.3.1.2 Normality

Jarque-Bera is insignificant and stable through the period. Further, the Q-Q plots do not exhibit signs of skewness or large kurtosis. Table 6.3-3 and *Figure 6.3-2* shows the Jarque-Bera and Q-Q results for dividend FI based indicator while the estimation includes the financial crisis.

Table 6.3-3

Jarque-Bera			
Statistic	Value	Prob	Label
Normal Test	0.2651	0.8759	Pr > ChiSq

Figure 6.3-2



6.3.1.3 Heteroscedasticity (ARCH)

The LM test results for ARCH effects in the FI based indicators are insignificant and stable over the period. We have as an example included the result for the composite FI based indicator:

Table 6.3-4

Tests for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	0.5942	0.4408	0.3803	0.5374
2	0.6855	0.7098	0.4388	0.8030
3	4.3330	0.2277	3.7266	0.2925
4	4.4587	0.3475	3.7357	0.4430
5	5.4480	0.3637	4.9113	0.4268
6	5.8085	0.4450	5.7912	0.4470
7	5.8430	0.5582	5.8421	0.5583
8	5.9833	0.6491	6.0800	0.6383
9	6.7226	0.6660	6.2145	0.7183
10	11.5248	0.3181	12.1689	0.2739
11	12.5237	0.3256	13.2165	0.2794
12	12.8122	0.3828	13.8868	0.3080

6.3.1.4 Ramsey RESET

When the estimation of the models include the financial crisis, or more precise, the drop in GDP during Q1 2009, all of our leading indicators get significant Ramsey RESET results indicating model mis-specification. The three indices with the least significant RESET test results are also those with the lowest RMSFE. However, even they are significant at a 95 % level, indicating a mis-specification in the

model. Since the sign of mis-specification only occurs when this large drop in GDP is included, we suspect that this will be an outlier that can have a large impact on our indices.

Table 6.3-5

Composite FI indicators RESET
results just before the crisis

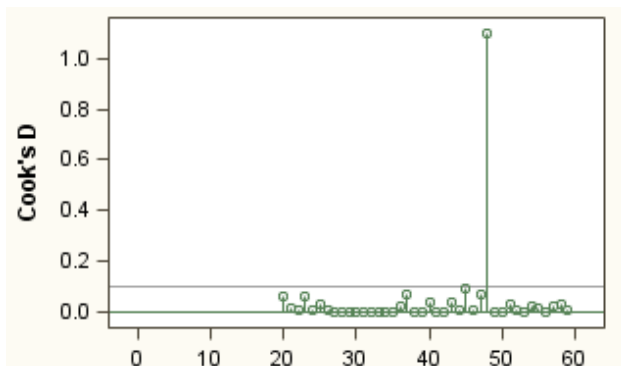
Ramsey's RESET Test		
Power	RESET	Pr > F
2	0.1872	0.6678
3	0.2068	0.8142
4	1.2612	0.3032

Composite FI indicators RESET
results including Q1 2009

Ramsey's RESET Test		
Power	RESET	Pr > F
2	4.4448	0.0420
3	2.8314	0.0725
4	1.8338	0.1596

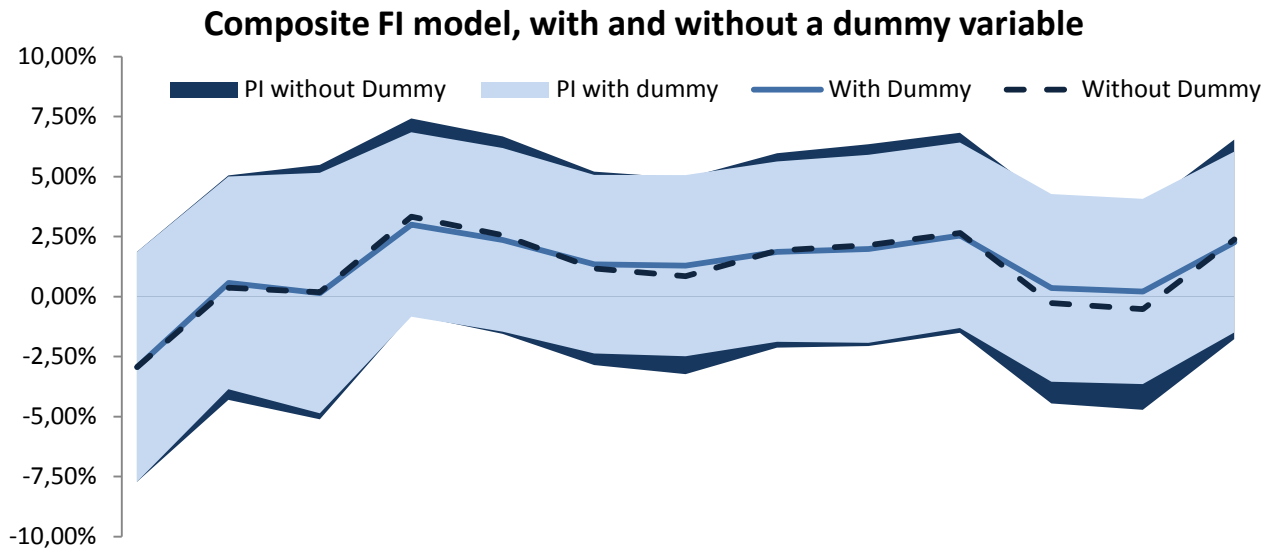
A look at the Cook's D graph for the Composite model estimating with observations from Q3 2011 back confirms our suspicion.

Figure 6.3-3



To further check for this we re-ran the forecasts, now including a dummy variable with the value of 1 at the Q1 2009 GDP growth, else zero. Now the Ramsey RESET test is again insignificant, indicating a correctly specified model. Though the difference in Ramsey's test is large, the difference in predicted values seems to be small. The RMSFE for the original model is as shown in *Table 6.3-1* is 0.0235 and with the dummy variable, the RMSFE is 0.0231. There is a larger difference in the prediction intervals (PI). The original model has on average 0.6 percentage point's wider PI than the model including the dummy variable. *Graph 6.3-1* shows the period from Q1 2009 to Q4 2011. As we see the estimates are pretty close to each other, but the prediction interval bands on 95 % confidence are narrower when the dummy variable is included.

Graph 6.3-1



6.3.1.5 Coefficients

The coefficients for all the FI based indicators behave more or less in the same manner. However, there is a clear pattern, all the β_1 estimates in the models doubles from before the finance crisis, Q2 2008, until Q1 2009. This is presented in Table 6.3-6 below.

Table 6.3-6

	Composite	Dividends	EBITDA	Employees	Sales	Total Assets
Estimation period Q2 2008	0.064	0.070	0.069	0.046	0.056	0.053
Estimation period Q1 2009	0.125	0.123	0.129	0.120	0.125	0.122
Estimate increase	195.8%	176.5%	188.1%	262.3%	220.9%	228.7%

6.3.2 Comparable Indices

Among the comparable indices the ranking of the leading indicators in terms of Root Mean Squared Forecasting Errors (RMSFE), the results are more or less the same. The only difference is that the PMI and K3 indicators have changed place.

Table 6.3-7

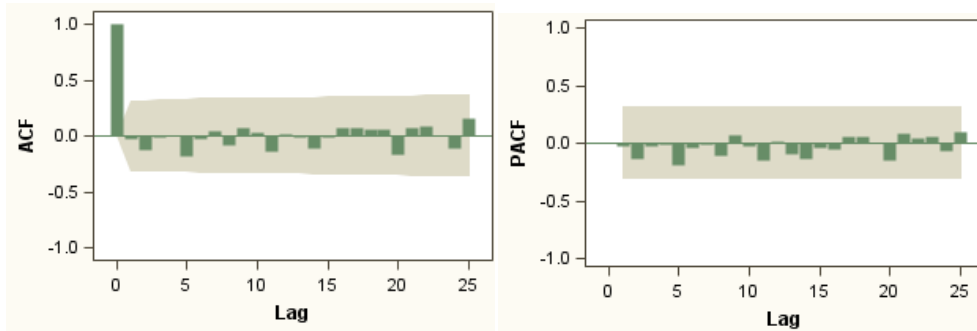
	Regression	RMSFE
CLI	$GDP_{t+1} = \alpha + \beta_1 CLI_t + \beta_2 CLI_{t-1} + \beta_3 GDP_{t-3} + \varepsilon_t$	0.0207
OBX	$GDP_{t+1} = \alpha + \beta_1 OBX_t + \beta_2 GDP_{t-1} + \varepsilon_t$	0.0247
M1	$GDP_{t+1} = \alpha + \beta_1 M1_t + \beta_2 M1_{t-1} + \beta_3 GDP_{t-1} + \beta_4 GDP_{t-3} + \varepsilon_t$	0.0290
Spread	$GDP_{t+1} = \alpha + \beta_1 SPR_{t-1} + \beta_2 SPR_{t-2} + \beta_3 SPR_{t-3} + \beta_4 GDP_{t-1} + \beta_5 GDP_{t-3} + \varepsilon_t$	0.0298
HPI	$GDP_{t+1} = \alpha + \beta_1 HPI_t + \beta_2 GDP_{t-1} + \beta_3 GDP_{t-3} + \varepsilon_t$	0.0299
Naïve	$GDP_{t+1} = \alpha + \beta_1 GDP_{t-1} + \varepsilon_t$	0.0325
K3	$GDP_{t+1} = \alpha + \beta_1 K3_t + \beta_2 K3_{t-1} + \beta_3 GDP_{t-1} + \beta_4 GDP_{t-3} + \varepsilon_t$	0.0333
PMI	$GDP_{t+1} = \alpha + \beta_1 PMI_{t-1} + \beta_2 GDP_{t-3} + \varepsilon_t$	0.0354

By RMSFE alone the best leading indicator is by far the CLI.

6.3.2.1 Autocorrelation

For all the comparable indicators, the BG is insignificant and stable over the whole period. In other words, we fail to reject the null-hypothesis of no autocorrelation. Looking at the ACF and PACF correlograms for the M1 based indicator when the financial crisis is included in the estimation, we see that all bars are well within the significance level.

Figure 6.3-4



6.3.2.2 Normality

The K3, M1 and PMI based indicators all have significant BG test results when the start of the financial crisis is included in the estimation. In the Q-Q plot we can for these series see signs of both skewness,

as in *Figure 6.3-5* from the K3 based indicator, and large tails, as in *Figure 6.3-6* from the PMI indicator.

Figure 6.3-5

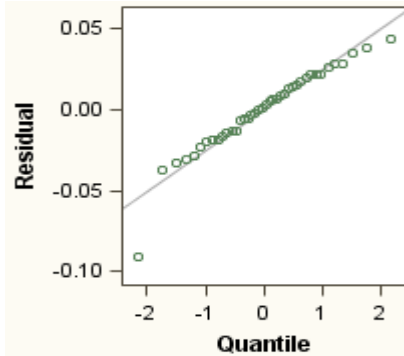
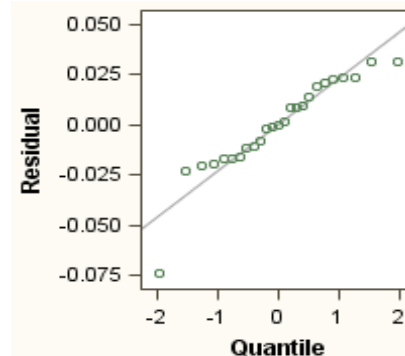


Figure 6.3-6



6.3.2.3 Heteroscedasticity (ARCH)

As with the FI based indicators the test results for the comparable are stable and insignificant through the period. The only exception is the M1 model just after the financial crisis start. Looking at the LM results in *Table 6.3-8* the significance is only at a 90 % level, and for only one lag.

Table 6.3-8

Tests for ARCH Disturbances Based on OLS Residuals				
Order	Q	Pr > Q	LM	Pr > LM
1	3.7318	0.0534	2.9588	0.0854
2	3.9451	0.1391	4.3607	0.1130
3	3.9597	0.2659	4.4945	0.2128
4	3.9705	0.4100	4.5493	0.3367
5	4.1249	0.5316	5.8390	0.3222
6	4.3305	0.6321	7.2136	0.3015
7	4.5410	0.7158	7.2145	0.4069
8	5.3083	0.7242	8.9274	0.3485
9	5.6131	0.7779	9.3422	0.4063
10	6.0629	0.8100	12.7845	0.2360
11	6.2552	0.8558	13.6641	0.2521
12	6.5759	0.8843	13.7051	0.3199

6.3.2.4 Ramsey RESET

Interestingly, one of the comparable indicators, K3, is the only indicator without significant results in the Ramsey RESET test. As we will see, this is also the indicator with the broadest prediction interval, indicating higher standard deviation for the model. Among the rest of the comparable indicators, we have the same problem as with our indicators; that the RESET test becomes significant when the estimating include the steep fall in GDP early 2009. Again these results become insignificant when adding a dummy to correct for the outlier. A more interesting note here is that the best indicator among all the models, the CLI based, is the only indicator with significant values (95 % level) in the RESET test before the financial crisis. It seems like the CLI are a bit unstable.

6.3.2.5 The Coefficients

For all the models, the coefficients behave more or less the same as with the in-sample estimates. One difference is with the third lagged GDP coefficient in the CLI model. The out of sample regression returns an insignificant coefficient. When this variable was removed from the in-sample regression we obtained significant Ramsey RESET results, indicating mis-specification. This is why it is included in our out of sample analysis. This could be the reason for the significant RESET results through much of the forecasting period in the CLI model, but we doubt it as the results after the finance crisis becomes insignificant once the outlier is corrected for.

6.3.2.6 Comparison of the models

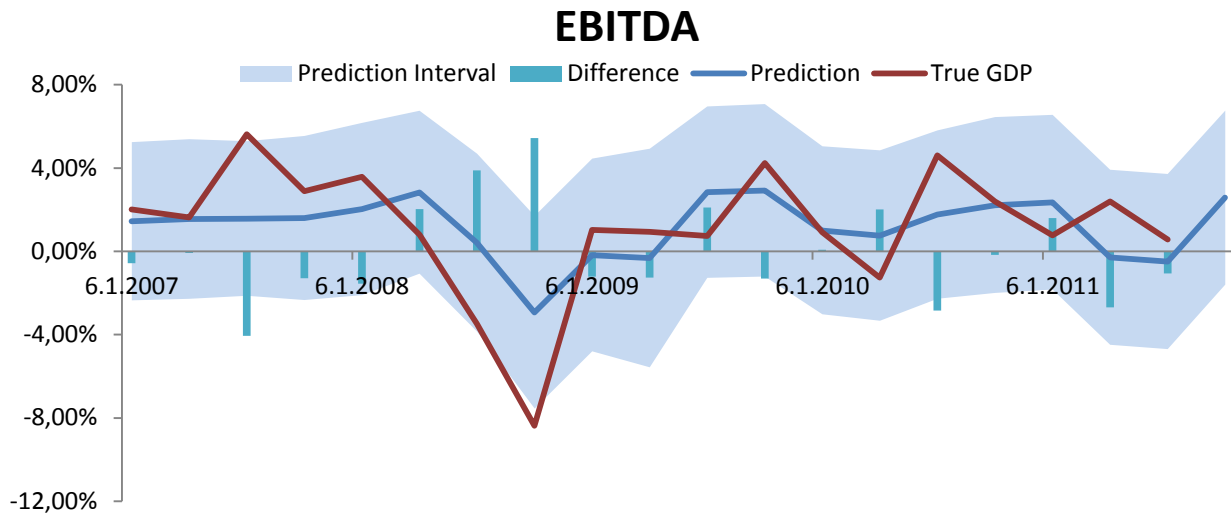
As mentioned earlier, the CLI is by far the best measure in terms of RMSFE, as with the RMSE in-sample. When comparing the size of the prediction intervals, only the 4th best indicator out of sample, the Dividends model, has a narrower average prediction interval, by 0.07 percentage points. Our best model, the EBITDA indicator has a 0.04 percentage points average wider prediction interval.

Table 6.3-9

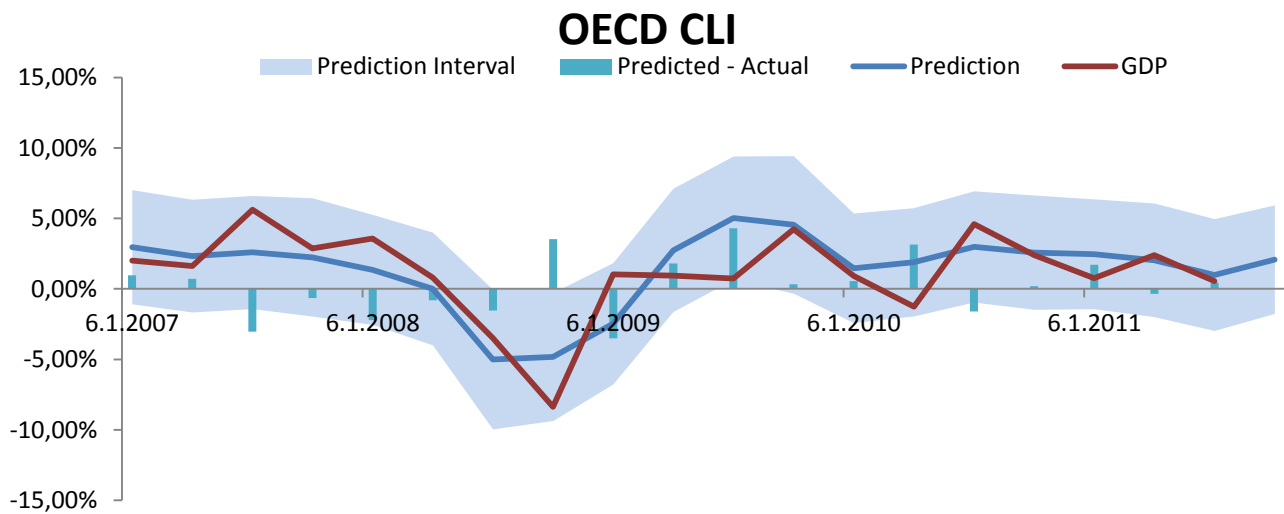
	Average PI	RMSFE ^{Ranking}
Dividends	8,24 %	0,0236 ⁴
CLI	8,31 %	0,0207 ¹
EBITDA	8,35 %	0,0232 ²
Composite	8,41 %	0,0235 ³
Sales/Revenue	8,70 %	0,0252 ⁶
OBX	8,77 %	0,0247 ⁵
Employees	8,81 %	0,0266 ⁸
Total Assets	8,84 %	0,0258 ⁷
PMI	9,24 %	0,0354 ¹⁴
Spread	9,38 %	0,0298 ¹⁰
M1	9,62 %	0,0290 ⁹
HPI	9,90 %	0,0299 ¹¹
Naïve	10,20 %	0,0325 ¹²
K3	10,58 %	0,0333 ¹³

As we see from *Table 6.3-9* there are some differences in the RMSFE ranking and the ranking of the prediction interval. When ranking the indicators by prediction interval, the best indicator, even better than the CLI, is our Dividend model. An interesting note is that among the FI based indicators, when ranking on narrowest to widest prediction interval we get almost the same order as when ranked on correlation between the regressors. The exception is that the dividends and EBITDA indicators switch places. So the multicollinearity in the different models might explain some of the differences in the prediction intervals. Then again, the CLI indicator is the one with the most correlation between the regressors, but still has a relatively narrow prediction interval.

Graph 6.3-2



Graph 6.3-3



As we see, the actual GDP values fall inside the CLI prediction interval for the whole period. As for our EBITDA indicator it falls outside the interval at two occasions. Although the CLI indicator performs better than our best, there are still reasons why our indices have an advantage over the OECD indicator.

1. It is more transparent and easy to understand
2. It does not rely on a general judgment, like the CLI
3. It can be updated more frequently

In fact, the FI leading indicators can be updated daily. How well this daily update fits as a leading indicator is out of scope for this thesis.

6.4 Comparison of our models to a naïve model

As for our in-sample-estimations we will here present a comparison of the Relative Root Mean Squared Forecast Error (RRMSFE) to that of a naïve model. The interpretation is the same.

Table 6.4-1

Model	RRMSFE	RMSFE
CLI	0.6369	0.0207
EBITDA	0.7138	0.0232
Composite	0.7220	0.0235
Dividends	0.7249	0.0236
OBX	0.7600	0.0247
Sales/Revenue	0.7745	0.0252
Total Assets	0.7945	0.0258
Employees	0.8177	0.0266
M1	0.8923	0.0290
Spread	0.9181	0.0298
HPI	0.9200	0.0299
<i>Naïve</i>	<i>1</i>	<i>0.0325</i>
K3	1.0246	0.0333
PMI	1.0886	0.0354

Most of the models presented in this thesis perform better than the Naïve model, but here we have two exceptions, namely the PMI and K3 Models. This is seen as they have a RRMSFE value above one. This means that these variables adds noise, and gives a higher variance in the predictions. This is surprising to us as we expected at least the PMI to perform better. This measure is a very important indicator in the US, for example, where they found the inspiration for the Norwegian PMI index. The Credit growth model was a bit more as expected, as this measure is seldom used in the previous research we have seen, but even though, we expected some form of contribution from it.

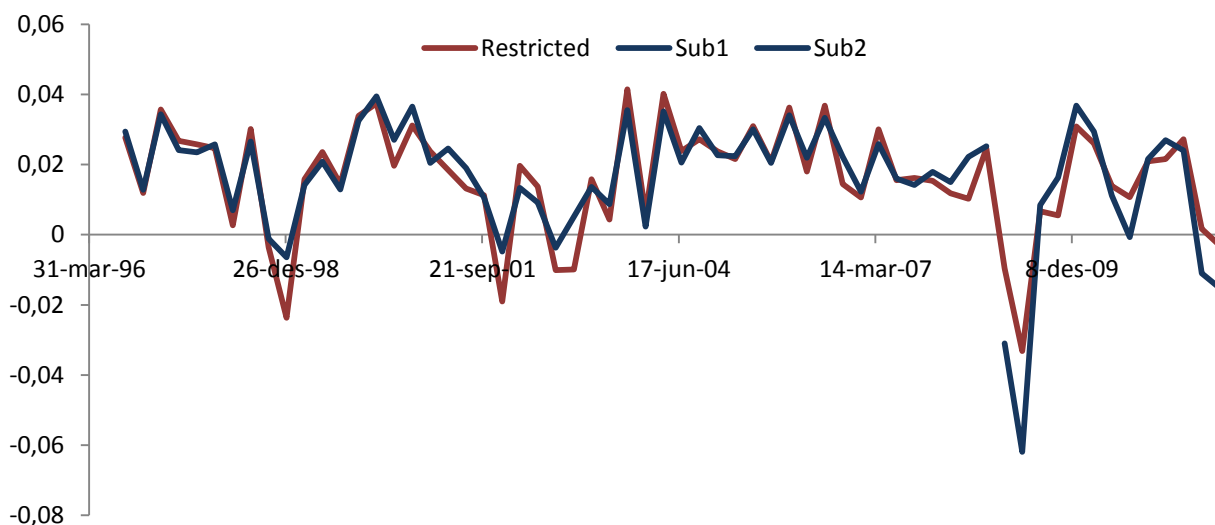
Besides this, we see the same trend as in the in-sample predictions; the CLI-model has the lowest RRMSFE, and the OBX are placed in the middle of our six FIs with Total Assets, Sales (Revenue), and Employees performing worse than the OBX-model with respect to the RRMSFE, while it is outperformed by the other three; our Dividends, EBITDA, and Composite models.

6.5 Structural breaks

We believe that the largest break in our data would be around the start of the financial crisis in 2008, so this is the most probable time to find a structural break. We will therefore start by testing for structural breaks from the second quarter of 2008 until the second quarter of 2009.

If we do a quick visual examination there is of course some difference between the restricted and unrestricted models. The unrestricted model is in *Graph 6.5-1* denoted Sub1 and Sub2.

Graph 6.5-1



This graph shows our Composite model, both as the full restricted model, and divided into two sub-periods. In the first period we see that the negative “peaks” is deeper in the restricted model than in the first sub-period, but this change in the second sub-period. This makes sense as the second sub-period contains the large negative outlier caused by the financial crises.

Looking at the Q-Q pots, the residuals seem a bit of, but no worse than that we will accept them. In the second sub-periods there are few observations, and one cannot expect to find clear evidence of normal distribution.

Figure 6.5-1

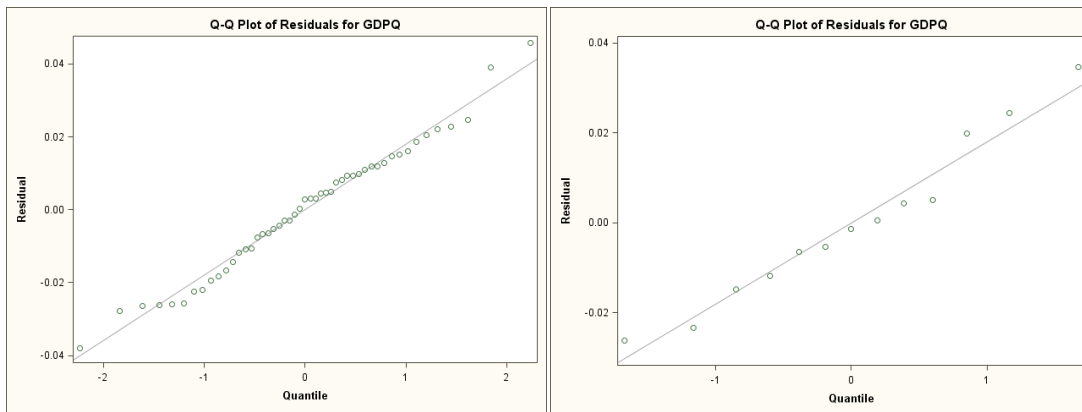


Figure 6.5-1 shows the Q-Q plot for the EBITDA indicator from the period Q1 1996 to Q3 2008 and from Q4 2008 to Q4 2011.

When applying the F-test for structural breaks in the variance, as described under Chow test in *section 4.3.1*, we found no evidence of structural breaks in the variance. Therefore we assume that the Chow test is sufficient when testing for structural breaks in our data.

As seen in *Table 6.5-1* below, the Chow test finds structural breaks in our leading indicators with up to 95 % confidence. We believe that the structural change that the Chow test speaks in favor of is due to the outlier at the structural break. Because of the few observations after the structural break, the outlier will have a significant impact on the second sub-period model.

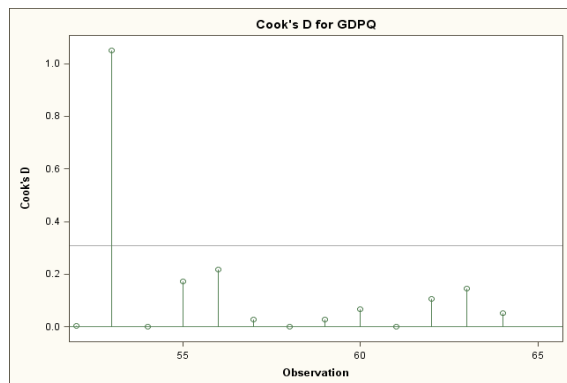
Table 6.5-1

Critical limit	0.1	0.05	0.01
Composite	Yes	Yes	No
Dividends	Yes	Yes	No
EBITDA	Yes	Yes	No
Employees	Yes	Yes	No
Sales/Revenue	Yes	Yes	No
Total Assets	Yes	No	No
CLI	No	No	No
HPI	Yes	No	No
K3	Yes	Yes	Yes
M1	Yes	Yes	No
OBX	Yes	Yes	No
PMI	No	No	No
Spread	Yes	Yes	No
Naïve	Yes	Yes	No

Yes = the test is significant at the level in the heading, No = we fail to reject the null hypothesis of no structural breaks.

When testing for the structural breaks, we tried to include the financial crisis in the first sub-period. The results then showed much less evidence of a structural break, supporting our statement that the outlier may have a severe impact on the second sub-period model due to few observations. The Cook's D plot for the second period, represented by the Composite model, further supports this statement, see Figure 6.5-2. Including a dummy variable with value one at the biggest drop in GDP and the index, Q1 2009, shows that the evidence of a structural break disappears. The outlier seen in Figure 6.5-2 is from Q4 2008. We found a structural break in our data in the quarter where the financial crisis had its largest impact on the financial markets.

Figure 6.5-2



According to the Chow test neither the CLI nor PMI indicator exhibit a structural break due to the financial crisis. We believe that this is because neither of them depends as heavily on the financial markets as the others. We are then a bit surprised that the HPI indicator shows some evidence of a structural break, but this is only at the 10 % level.

We recognize that structural breaks are a threat to our models, but believe that given more observations after the crisis, the one outlier will not impact the rest of the model as severely as it does now, and the structural change will diminish.

6.6 Granger causality tests

6.6.1 Fundamental Indices

Table 6.6-1

<i>FI on GDP</i>	Composite	EBITDA	Dividends	Employees	Sales	Total Assets
Restricted	0.0361	0.0361	0.0361	0.0361	0.0361	0.0361
Unrestricted	0.0235	0.0228	0.0227	0.0272	0.0255	0.0262
DF	59	59	59	59	59	59
F – value	31.5565	34.3145	34.4373	19.3025	24.6516	22.2824

<i>GDP on FI</i>	Composite	EBITDA	Dividends	Employees	Sales	Total Assets
Restricted	1.1887	1.1226	1.2172	1.5515	1.1557	1.2095
Unrestricted	1.1812	1.1187	1.2100	1.5355	1.1454	1.1815
DF	59	59	59	59	59	59
F – value	0.3764	0.2028	0.3482	0.6122	0.5287	1.3956

The Critical value found in the F-distribution is 4.004 for all of the tests here.

For our fundamentally weighted indices we got the results we expected. This means that we see indications of unidirectional causality from our FIs to GDP, while the GDP does not Granger cause our FIs.

As mentioned in section 2.3, this causality may be caused by the shared correlation to oil price fluctuations, resulting in possible spurious causality. The same applies to the OBX results below. Further investigation of this is out of scope for this thesis.

6.6.2 Comparable leading Indices

Table 6.6-2

<i>CI on GDP</i>	Spread	PMI	HPI	M1	K3	CLI	OBX
Restricted	0.0304	0.0178	0.0299	0.0264	0.0325	0.0201	0.0246
Unrestricted	0.0328	0.0202	0.0328	0.0328	0.0328	0.0355	0.0361
DF	54	27	56	55	55	56	59
F-test value	1.4310	3.5356	5.4104	6.6525	0.2864	21.4954	27.6553
Critical value	2.7758	4.2100	4.0130	3.1650	3.1650	3.1619	4.0040

<i>GDP on CI</i>	Spread	PMI	HPI	M1	K3	CLI	OBX
Restricted	23.9931	659.5622	0.0524	0.0894	0.0125	4.5471	1.0914
Unrestricted	24.5173	675.2758	0.0532	0.0965	0.0164	5.0851	1.0948
DF	54	25	56	55	55	56	59
F-test value	0.3933	0.5956	0.8316	2.1929	8.7325	3.3125	0.1810
Critical value	2.7758	4.2417	4.0130	3.1650	3.1650	3.1619	4.0040

The results for the comparable indicators were somewhat surprising. We see that the Credit growth model, K3, does not Granger cause change in GDP, but GDP has a unidirectional causality to credit growth. This might be an explanation on why we did not get any significant regressors in this model, as we were looking at it the wrong way, according to our causality test.

The OECD CLI model shows indications of bilateral causality. This feedback effect may stem from any of the 6 sub-indices, but we would guess that the sub-index of the general judgment of the future might be affected by the previous GDP number.

For both the Spread and PMI model we found indications of independence. This could possibly explain why we got such odd results for the PMI model where we only saw one significant variable, and this was the second lagged value, also the disappointingly weak results for the Spread-model.

We had three models performing as expected in the Granger causality test, those being the HPI, M1, and OBX model. Here we found indications of unidirectional causality from the indicator to GDP.

7 Discussion

In this section we will discuss some of the models and how we got there. We will also discuss some of the results in context of our hypotheses. We will in the two first subsections keep it in a general level and be more specific when discussing examining our results versus hypotheses.

7.1 Method and models

Here we will discuss the methods we have used in this thesis. We have previously discussed the data quality and validity, and will not go in further detail about our sources. One thing about our data that should be addressed is the handling of it. Historic stock prices for 100+ companies, their historic fundamental data, GDP numbers and our compared data adds up to quite a few numbers that needs to be organized and properly aligned. We have done our utmost to secure that everything is correct, but must recognize the possibility of errors. Though errors may have happened, we believe that because of our efforts of getting everything right, the errors that might still be in the dataset will not severely impact the validity of our results.

The dynamic models we have used for our FI indicators are straight forward, but an obvious discussion point is at what time the forecast should be presented, and also what forecast horizon to choose. We have chosen to forecast one quarter ahead from the end of each quarter.

- Why forecast one quarter ahead?

This goes two ways, why not forecast the GDP that is published approximately 50 days after the quarter end. It would be intuitive to say that the next published GDP numbers is the most anticipated. But as we have noted, changes in fiscal and monetary policy does not always have an immediate impact on the real economy, and it is therefore important to have good future predictions for the policy makers to be able to adjust their policies at an early stage.

Since it is important for policy makers to get an early forecast, why not forecast more than one quarter ahead in time? As with weather forecasts, the predictions become more unreliable the further into the future we try to predict. We do recognize that

others have managed to create relatively good forecasts several quarters ahead, so it should also be possible for our models, but this is out of scope for our thesis. *Table 7.1-1* compares results between the EBITDA based indicator, when forecasting one and two quarters into the future, to exemplify how the credibility may change when forecasting further into the future.

Table 7.1-1

	EBITDA one quarter forecast	EBITDA two quarter forecast
R²	0.3999	0.0914
RMSE	0.0197	0.0240
AIC	-308.30	-277.36

- Why do we forecast from the end of the quarter?

There are at least two alternatives that should be considered. Wait until GDP for the current quarter is published, and do the forecast once the previous GDP is published. We chose not to wait until the current GDP is published, as that would eat 50 days of our forecast. By waiting until end of the quarter we get more updated numbers for our fundamental indices. Choosing end of the quarter is a trade-off between getting the newest numbers and not waiting too long before forecasting. We will come back to this in our suggestions for further research.

As for the peer indicators, it must be noted that they are, in this thesis, used in the same way as our FIs. The indicators based on interest rate spread, house price index, credit, and money supply are all presented as a single indicator through a principle component approach in Vonen (2011). Both PMI and CLI are indicators that vary around a number. CLI was originally intended to indicate trend changes rather than predicting the GDP or its growth. Another point in this discussion is that these other variables might be better to predict GDP for other forecast horizons. There are many different

variations possible, some of them mentioned above, and it is out of the scope of this thesis to find the best model in term of different times and different lead length of the predictions.

7.2 Hypothesis 1

From the results we can see that we get a mixed answer to our first hypothesis

H_0^1 : *A fundamentally weighted stock index is better than a market capitalization weighted stock index as a leading indicator for Norwegian GDP*

H_A^1 : *A fundamentally weighted stock index is not better than a market capitalization weighted stock index as a leading indicator for Norwegian GDP*

The three indicators based on EBITDA, dividends, and the composite index all get a lower RMS(F)E than the OBX based leading indicator, while the indicators based on number of employees, sales, and total assets have all a higher RMS(F)E. Why is this? In this thesis, what the EBITDA, the Dividends, and the Composite index have in common relative to the three others is that their weights contain averages of the weights of the other indices. As explained in the methodology, when EBITDA is negative, it has been replaced by an average of sales and total assets. When dividends are zero, it got the weight equal to the average weight of EBITDA, sales, and total assets, and the composite is an average of all the others except employees. This averaging might have helped to get a better representation of the companies, and be better suited to explain actual size. This has in turn given a more correct prediction of GDP. This then raises the question: “why is not our composite index our top performer?” To investigate this is out of scope for this thesis. As for the other three indicators, they may contain just as much or more inefficiency than the stock market and therefore have worse results than the OBX based indicator, but this is just speculations. This ranking of our models gives the result that our best indicators are also the most complex.

7.3 Hypothesis 2

H_0^2 : A fundamentally weighted stock index is a good indicator for Norwegian GDP

H_A^2 : A fundamentally weighted stock index is not a good indicator for Norwegian GDP

To test this hypothesis we have compared our FI based leading indicators to other known indicators. Our FI based indicators are performing well compared most of the leading indicators we have investigated. Of the 14 indicators we have investigated, including the naïve model, our indicators are ranked number 2-4 and 6-8 in terms of RMS(F)E. The indicator that is splitting the two groups of FI based indicators is, as we said in the previous section, the OBX based indicator.

The best forecaster of the Norwegian GDP, both in, and out of sample, turned out to be the OECDs Composite Leading Indicator for Norway. However, forecasting the change in GDP is not its intended purpose. The CLI provided by OECD is originally a turning point indicator. It is designed to anticipate turning points in the in economic activity relative to trend, and the point of regaining momentum.

Table 7.3-1

	Turning point dates as predicted by CLI	Turning point dates in GDP	Lead (months)
trough	Mar-75	May-75	2
peak	Aug-76	Jul-76	-1
trough	Jul-77	Nov-77	4
peak	Dec-79	Feb-80	2
trough	Sep-82	Jan-83	4
peak	Nov-85	Mar-86	4
trough	Aug-88	May-90	21
peak	Sep-90	May-92	20
trough	Dec-92	Feb-93	2
peak	Dec-94		extra
trough	Nov-95		extra
peak	Jul-97	Nov-97	4
trough	Dec-98	Feb-99	2
peak	Sep-00	Feb-00	-7
trough	Feb-03	Apr-03	2
peak	Sep-07	Oct-07	1
trough	Mar-09	Aug-09	5
peak	Mar-11		provisional

(OECD, 2012)

As we see, the time from when the CLI predicts turning point and when it actually happens ranges from -7 months to 21 months. OECD claims an average of 4 month lead in turning point indication, this average is closer to $4\frac{1}{3}$, not accounting for the lags in the prediction. The average time span between a turning point prediction and the actual turning point has an average of 5.4 months, and then the two predicted turning points that never occurred are excluded. When being, maybe overly, critical, it seems like the CLI just says that there will be a turning point in the economy, but gives little information of when and why. It might have happened 6 months ago. As noted earlier, the CLI indicator has some problems with the Granger causality; it exhibits a bilateral causal relationship to GDP. This might not be surprising as we see that two of the turning points predicted came after the actual event.

In spite of our criticism, we recognize it is a good indicator. We have used the CLI in a different way than for what it is originally designed for, and still it is the best leading indicator of next quarters GDP growth, both in and out of sample.

8 Conclusion

The motivation behind this thesis was to help the policymakers get an early, and accurate indication of the future state of the economy, in order for them to make timely adjustments to fiscal and monetary policies. This because it may take some time before the effects of these changes can be seen in the real economy.

In this thesis we have investigated whether fundamentally weighted indices can be a better leading indicator than market capitalization weighted indices for future change in the Norwegian GDP. The market capitalization weighted indices are represented by the OBX index, which consists of the 25 most traded stocks on the Oslo Stock Exchange Benchmark Index. Furthermore, we investigate whether a fundamentally weighted index is a better predictor than other known leading indicators, such as measures for Money Supply, interest rate spreads, House Price Index, the Composite Leading Indicator for from the OECD, the Purchasing Manager Index from NIMA and Fokus Bank. We also included a measure of Credit Growth as we expected this to have some explanatory power. We have created five different fundamental indices based on different metrics; gross dividends, total sales (revenue), total assets, EBITDA, number of employees, and finally a composite index. The composite index being an equally weighted index based on the four of the other metrics.

Vonen (2011) tried to create the first FCI constructed specially for Norway by using a Principle Component approach. To investigate whether there is something to gain from pooling the information together rather than just forecast GDP based on individual financial series in the Norwegian economy she investigated 13 different financial variables for comparison. Those include most of our comparable indicators, interestingly the results more or less coincide with ours. Vonen (2011) finds that the interest rate spread measure, including 10 year Treasury bond and three month NIBOR, are significant when forecasting 1 period ahead. This is in accordance with our results, although we use different interest rates. Vonen's chosen credit indicator is also significant in the one quarter forecast (K2 for Banks instead of the total K3 that we used). This could indicate that we chose wrong measure in our thesis, however based on our initial reasoning we still believe that we chose the right measures. Another difference is that the HPI model in Vonen (2011) is not significant for a

one quarter forecasts, however it is the best when applied to a two quarter forecast. This seems odd to us, but gives further indications as to why the forecast horizon should be investigated further. Vonen (2011) shows several differences in which models are significant when changing from one quarter to two quarter forecasts. For instance, the stock model is not significant when forecasting for two quarters. We expect this to be due to wrong weighing, according to our hypothesis of FI versus MCAP weighting. Vonen (2011) finds that some individual financial variables occasionally perform better than the investigated FCI, this lends support to our idea that a correctly weighted stock index could be able to predict future growth in GDP. PMI is not mentioned in this paper, not surprising as it is not really recognized as a macroeconomic indicator in Norway.

Our analysis is done via a dynamic regression model using the linear transfer function method to reach the preferred setup for each model. This was done to ensure that our FIs based indicators became as good as possible, and also that they were tested against their best peers.

Regarding our first hypothesis, that an FI based leading indicator is better than a MCAP weighted index leading indicator, our results is as stated, divergent. With three of our FI based indicators outperforming the OBX indicator and three performing worse in terms of RMS(F)E. In addition, it should be noted that in terms of prediction intervals four of our FI indicators outperformed the OBX indicator. This Indicating lower standard deviation in the FI models than in the OBX model. Though the results in the prediction intervals speaks in slightly favor of our FI models, it is still difficult to give a clear conclusion to our first hypothesis..

When comparing our FI based indicators to other known indicators for our second hypothesis the FI based indicators performs well, but the results are somewhat surprising, and very interesting. We find that the Interest rate spread (five years Treasury bond – three months NIBOR) model is not among our top performers, this is surprising as it is one of the well-known leading indicators from the international literature. This might be due to computational differences between our model and that of others, the period in question, and so on. In terms of RMS(F)E the FI based indicators are among the top performers, ranking from 2-4 and 6-8 out of 14, and we would say that we do not reject our null hypothesis. In terms of a narrow prediction interval (at 95 % confidence level) our Dividend

indicator is the best performer. The other FI indicators also perform well with ranking 3-5 and 7-8. Again confirming our hypothesis that a FI leading indicator is in general a good predictor for Norwegian GDP.

To test the robustness of our results we have tested for structural breaks and causality. Our FI based indicators, in our opinion, pass these robustness tests. We also compared all of our models' RMS(F)E to a naïve model. In-sample the naïve model was outperformed by all of our FIs, as well as by all the comparable leading indicators. This was not surprising as the naïve model is nested in all of the models. Out of sample, the naïve model outperformed K3 and PMI, meaning that the inclusion of these variables only increase the variance compared to a model consisting of only lagged dependent variables.

8.1 Further research

When writing this thesis we have encountered several interesting topics which were out of scope for this thesis. The ones we encourage most for further research is; the discussion of exogeneity and the investigation of the possible spurious causality.

To verify or contradict the findings in this thesis, we suggest that the methodology is applied on all of the companies on the Oslo Stock Exchange, and not only those who are the most traded. As GDP tries to give a complete picture of the economy, all companies on the stock exchange should be included. The method should also be applied to other countries to see if the results also are significant outside of Norway.

We also recommend testing other forecasting horizons and the time of forecast. Another interesting expansion would be to see if this method can provide accurate forecasts more frequently. That is, what happens if the FIs quarterly return is updated monthly, weekly, or even daily?

As discussed, and seen in the findings, this thesis produced the peculiar result of having three FI based indicators better than the OBX based indicator and three worse. We have provided a hypothesis in the discussion saying that FI's containing averages of different fundamental data, gives a better

measure of the firms' fundamental size. But then, why is not the composite index our top performer?
We would recommend further investigation of why this result appears.

9 Bibliography

Aastveit, K. A. & Trovik, T. G., 2008. Nowcasting Norwegian GDP: The role of asset prices in a small open economy. *Working Paper Norges Bank*.

Achen, C. H., 1982. *Interpreting and Using Regression*. Beverly Hills, California: Sage Publications.

Arnott, R. D., Hsu, J. & More, P., 2005. Fundamental Indexation. *Financial Analysts Journal*, 61(2), pp. 83 - 99.

Barnes, R., u.d. www.investopedia.com. [Internett]

Available at: <http://www.investopedia.com/university/releases/gdp.asp#axzz1uZKkiMcE>

[Funnet 20 February 2012].

Beaudry, P. & Portier, F., 2006. Stock Prices, News, and Economic Fluctuations. *The American Economic Review*, pp. 1293 - 1309.

Bernstein, W. J., 2006. www.efficientfrontier.com. [Internett]

Available at: <http://www.efficientfrontier.com/ef/0adhoc/fi.htm>

[Funnet 6 March 2012].

Bloomberg, u.d. www.bloomberg.com. [Internett]

Available at: <http://www.bloomberg.com/quote/BFCIUS:IND>

[Funnet 7 March 2012].

Chiarella, C. & Gao, S., 2007. *Type I Spurious Regression in Econometrics*, s.l.: s.n.

Commision on the Measurement of Economic Performance and social Progress, 2008. SURVEY OF EXISTING APPROACHES TO MEASURING SOCIO-ECONOMIC PROGRESS.

D`Antonio, P., 2008. Appendix, pages 26-28, in DiClemente, R. and K. Schoenholtz (September 26,2008) "A View of the U.S. Subprime Crisis". *EMA Special Report, Citigroup Global Markets Inc..*

Deutsche Bank Securities inc., 2010. Global Economic Perspectives.

Diebold, F., 2001. *Elements of Forecasting*. s.l.:South Western Publishing.

European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development(OECD), United Nations, and World Bank, 2008. System of National Accounts.

Finansdepartementet, u.d. www.ungokonomi.no. [Internett]

Available at: http://www.ungokonomi.no/index.php?option=com_content&task=view&id=12&Itemid=16

[Funnet 22 February 2012].

Fokus Bank, u.d. *Norsk PMI*. [Internett]

Available at: <http://www.fokus.no/nb-no/Bedrift/Store-bedrifter/Markets/Pages/Norsk-PMI.aspx>

[Funnet 22 April 2012].

Gerdrup, K. R., Hammerslad, R. & Naug, B. E., 2006. Finansielle størrelser og utviklingen i realøkonomien. *Penger og Kreditt*, Issue 2.

Gerdrup, K. R., Hammersland, R. & Naug, B. E., 2006. Finansielle størrelser og utviklingen i realøkonomien. *Penger og Kreditt* 2.

Guichard, S., Haugh, D. & Turner, D., 2009. QUANTIFYING THE EFFECT OF FINANCIAL CONDITIONS IN THE EURO AREA, JAPAN, UNITED KINGDOM AND UNITED STATES. *ECONOMICS DEPARTMENT WORKING PAPERS*, Volum 667.

Guichard, S. & Turner, D., 2008. Quantifying the Effect of Financial Conditions on US Activity. *OECD Economics Department Working Papers*, Issue No. 635.

Gujarati, D. N. & Porter, D. C., 2009. *Basic Econometrics*. Fifth Edition red. New York: McGraw-Hill.

Gyomai, G. & Guidetti, E., 2012. OECD System of Composite Leading Indicators. April.

Hain, J., 2010. Comparison of Common Tests for Normality.

Hakkio, C. S. & Keeton, W. R., 2009. Financial Stress: What Is It, How Can It Be Measured, and Why Does It Matter?. *FEDERAL RESERVE BANK OF KANSAS CITY ECONOMIC REVIEW*, Issue SECOND QUARTER 2009.

Hatzius, J. et al., 2010. Financial Conditions Indexes: A Fresh Look after the Financial Crises. *NBER Working Paper*, Issue 16150.

Hendry, D. F. & Richard, J.-F., 1983. The Econometric Analysis of Economic Time Series. *International Statistical Review*, 51(2), pp. 111-148.

Husebø, T. A., Wilhelmsen, B.-R. & Bank, N., 2005. Norwegian Business Cycles 1982-2003. *Staff Memo Economics Department*, Issue 2.

Koop, G., 2000. *Analysis of Economic data*. New York: John Wiley & Sons.

Leamer, E. E., 1983. Model Choice and Specification Analysis. I: Z. Grilich & Intriligator, red. *Handbook of Econometrics vol. 1*. Amsterdam: North Holland Publishing Company, pp. 300-301.

Macroeconomic Advisers, 1998. From the US Economic Outlook, Technical Notes.

Makridakis, S., Wheelwright, S. C. & Hyndman, R. J., 1998. Forecasting Methods and Applications. I: s.l.:John Wiley & Sons, Inc, pp. 403-418.

- Murdoch University, 2009. <http://www.murdoch.edu.au>. [Internett]
Available at: <http://www.cms.murdoch.edu.au/areas/math/statsnotes/samplestats/ggplot.html>
[Funnet 10 March 2012].
- Nette, W., 2010. A New Stock Index to Better Predict the United States` Real GDP.
- Norges Bank, 2011. *Ord og uttrykk*. [Internett]
Available at: <http://www.norges-bank.no/no/ord-og-uttrykk/#pengemengden>
[Funnet 28 March 2012].
- OECD, 2012. OECD Composite Leading Indicators, Country Reviews. March.
- OECD, u.d. www.oecd.org. [Internett]
Available at: http://www.oecd.org/document/12/0,3746,en_2649_34249_1891148_1_1_1_1,00.html
[Funnet 11 May 2012].
- Oishi, S., Kesebir, S. & Diener, E., 2011. Income Inequality and Happiness. *Psychological Science*, Volum 22(9), pp. 1095 - 1100.
- Ørbeck, A. S. & Torvanger, M., 2011. A Financial Conditions Index for Norway: Can financial indicators preict GDP?.
- Oslo Børs, 2010. www.oslobors.no. [Internett]
Available at: http://www.oslobors.no/ob_eng/Oslo-Boers/About-us/Press-room/Press-releases/Oslo-Boers-to-change-the-rules-for-OBX-and-establish-OBXI-I-International
[Funnet 3 March 2012].
- Pankratz, A., 1991. *Forecasting with dynamic regression models*. New York: John Wiley and Sons.
- Research Affiliates, 2007-2012. www.rallc.com. [Internett]
Available at: <http://www.rallc.com/rafi/index.htm>
[Funnet 20 January 2012].
- Riiser, M. D., 2005. House prices, equity prices, investment and credit– what do they tell us about banking crises? A historical analysis based on Norwegian data. *Economic Bulletin 05 Q3*.
- Riiser, M. D., 2008. Asset prices, investment and credit - what do they tell us about financial vulnerability?. *Economic commentaries*, Issue 6.
- Riiser, M. D., 2010. Asset prices, investment, credit and financial vulnerability. *Economic commentaries*, Issue 4.
- Rosenberg, M. R., 2009. Financial Conditions Watch. *Bloomberg*.

Statistics Norway, 2012. *Verdipapirer, kredittindikatorer, pengemengde og renter*. [Internett]
Available at: <http://www.ssb.no/emner/11/01/k3/index.html>
[Funnet 22 April 2012].

Statistisk Sentral Byrå, 2012. *Generelt om sesongjustering*. [Internett]
Available at: <http://www.ssb.no/metadata/metode/sesongjustering.pdf>
[Funnet 24 February 2012].

Statistisk Sentral Byrå, u.d. *www.ssb.no*. [Internett]
Available at: <http://www.ssb.no/metadata/conceptvariable/vardok/1743/en>
[Funnet 22 February 2012].

The University of British Colombia, u.d. *www.ubc.ca*. [Internett]
Available at: <http://shazam.econ.ubc.ca/intro/garch1.htm>
[Funnet 20 February 2012].

United States Census Bureau, u.d. *www.census.gov*. [Internett]
Available at: <http://www.census.gov/srd/www/x12a/>
[Funnet 24 February 2012].

Vonen, N. H., 2011. A financial conditions index for Norway. *Norges Bank - Staff Memo*, Issue 07.

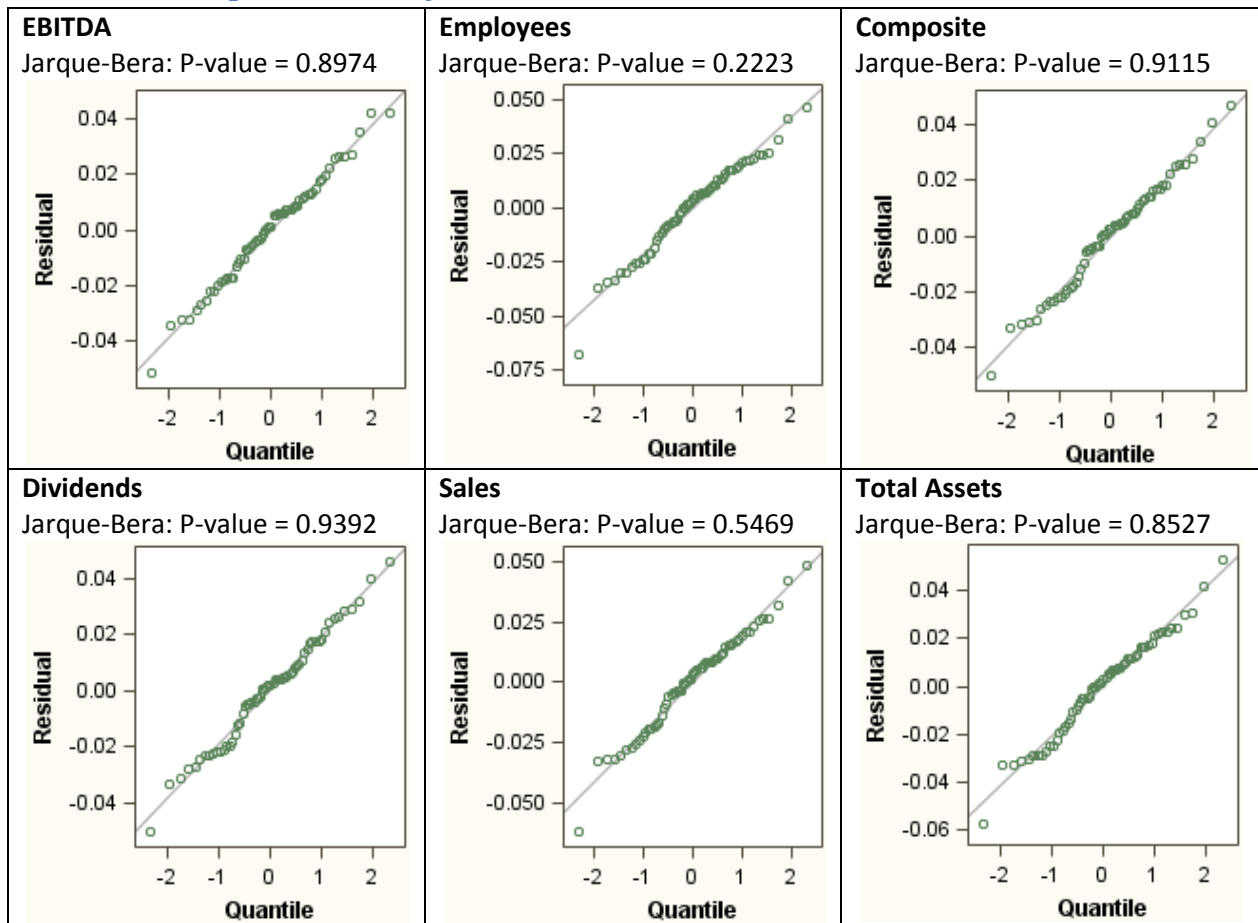
Appendix

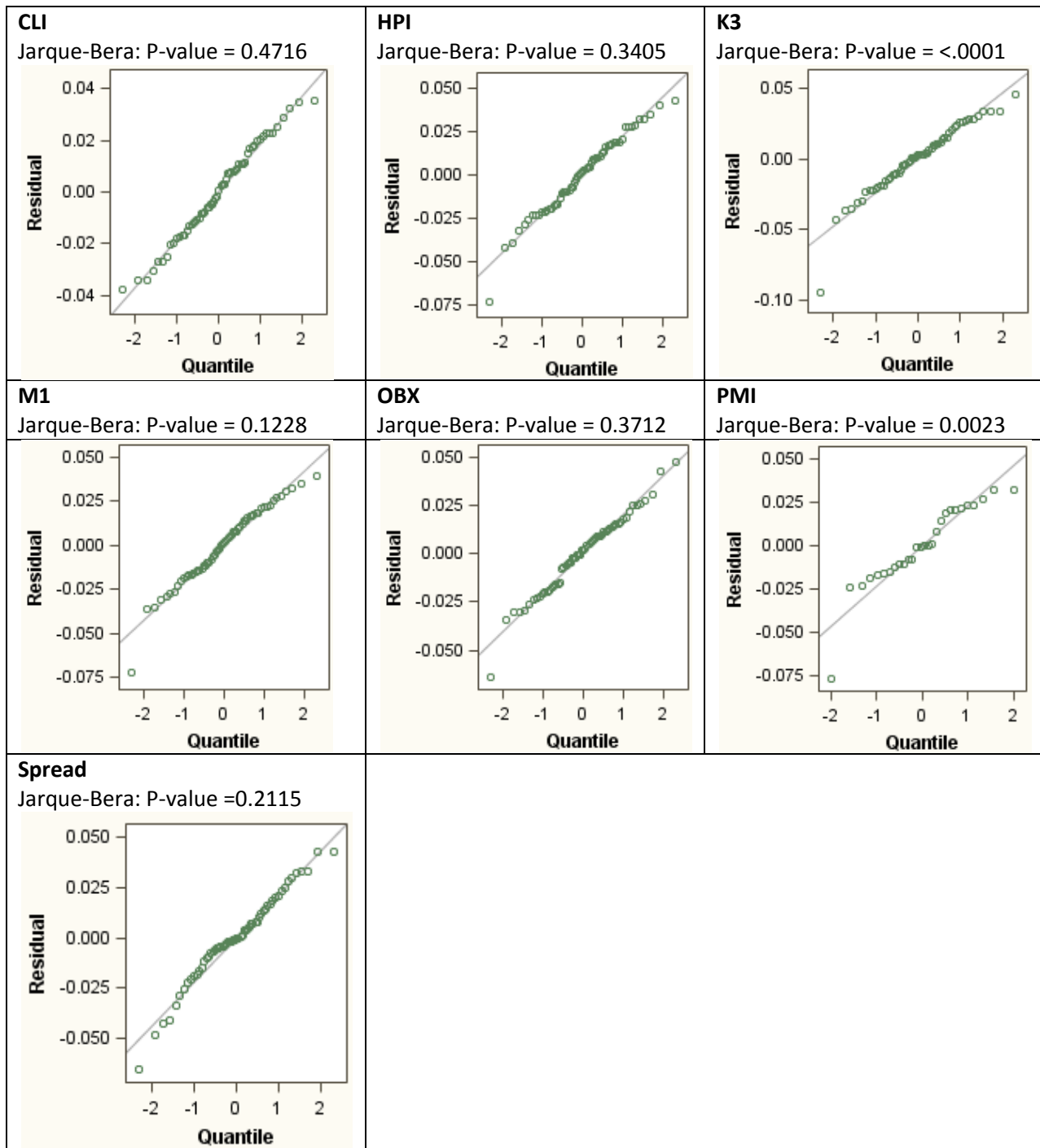
A In-sample autocorrelation

EBITDA <table border="1"> <thead> <tr> <th colspan="3">Godfrey's Serial Correlation Test</th></tr> <tr> <th>Alternative</th><th>LM</th><th>Pr > LM</th></tr> </thead> <tbody> <tr> <td>AR(1)</td><td>0.0599</td><td>0.8066</td></tr> <tr> <td>AR(2)</td><td>1.6549</td><td>0.4372</td></tr> <tr> <td>AR(3)</td><td>1.6576</td><td>0.6464</td></tr> <tr> <td>AR(4)</td><td>4.6002</td><td>0.3308</td></tr> </tbody> </table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.0599	0.8066	AR(2)	1.6549	0.4372	AR(3)	1.6576	0.6464	AR(4)	4.6002	0.3308	Employees <table border="1"> <thead> <tr> <th colspan="3">Godfrey's Serial Correlation Test</th></tr> <tr> <th>Alternative</th><th>LM</th><th>Pr > LM</th></tr> </thead> <tbody> <tr> <td>AR(1)</td><td>0.2628</td><td>0.6082</td></tr> <tr> <td>AR(2)</td><td>0.5027</td><td>0.7777</td></tr> <tr> <td>AR(3)</td><td>0.6885</td><td>0.8759</td></tr> <tr> <td>AR(4)</td><td>3.6798</td><td>0.4511</td></tr> </tbody> </table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.2628	0.6082	AR(2)	0.5027	0.7777	AR(3)	0.6885	0.8759	AR(4)	3.6798	0.4511	Composite <table border="1"> <thead> <tr> <th colspan="3">Godfrey's Serial Correlation Test</th></tr> <tr> <th>Alternative</th><th>LM</th><th>Pr > LM</th></tr> </thead> <tbody> <tr> <td>AR(1)</td><td>0.0641</td><td>0.8002</td></tr> <tr> <td>AR(2)</td><td>1.1754</td><td>0.5556</td></tr> <tr> <td>AR(3)</td><td>1.2723</td><td>0.7357</td></tr> <tr> <td>AR(4)</td><td>4.1721</td><td>0.3832</td></tr> </tbody> </table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.0641	0.8002	AR(2)	1.1754	0.5556	AR(3)	1.2723	0.7357	AR(4)	4.1721	0.3832
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.0599	0.8066																																																						
AR(2)	1.6549	0.4372																																																						
AR(3)	1.6576	0.6464																																																						
AR(4)	4.6002	0.3308																																																						
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.2628	0.6082																																																						
AR(2)	0.5027	0.7777																																																						
AR(3)	0.6885	0.8759																																																						
AR(4)	3.6798	0.4511																																																						
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.0641	0.8002																																																						
AR(2)	1.1754	0.5556																																																						
AR(3)	1.2723	0.7357																																																						
AR(4)	4.1721	0.3832																																																						
Dividends <table border="1"> <thead> <tr> <th colspan="3">Godfrey's Serial Correlation Test</th></tr> <tr> <th>Alternative</th><th>LM</th><th>Pr > LM</th></tr> </thead> <tbody> <tr> <td>AR(1)</td><td>0.0127</td><td>0.9102</td></tr> <tr> <td>AR(2)</td><td>1.1032</td><td>0.5760</td></tr> <tr> <td>AR(3)</td><td>1.1252</td><td>0.7710</td></tr> <tr> <td>AR(4)</td><td>3.7350</td><td>0.4431</td></tr> </tbody> </table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.0127	0.9102	AR(2)	1.1032	0.5760	AR(3)	1.1252	0.7710	AR(4)	3.7350	0.4431	Sales <table border="1"> <thead> <tr> <th colspan="3">Godfrey's Serial Correlation Test</th></tr> <tr> <th>Alternative</th><th>LM</th><th>Pr > LM</th></tr> </thead> <tbody> <tr> <td>AR(1)</td><td>0.1525</td><td>0.6961</td></tr> <tr> <td>AR(2)</td><td>0.8649</td><td>0.6489</td></tr> <tr> <td>AR(3)</td><td>0.9881</td><td>0.8041</td></tr> <tr> <td>AR(4)</td><td>4.2197</td><td>0.3771</td></tr> </tbody> </table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.1525	0.6961	AR(2)	0.8649	0.6489	AR(3)	0.9881	0.8041	AR(4)	4.2197	0.3771	Total Assets <table border="1"> <thead> <tr> <th colspan="3">Godfrey's Serial Correlation Test</th></tr> <tr> <th>Alternative</th><th>LM</th><th>Pr > LM</th></tr> </thead> <tbody> <tr> <td>AR(1)</td><td>0.6193</td><td>0.4313</td></tr> <tr> <td>AR(2)</td><td>1.6161</td><td>0.4457</td></tr> <tr> <td>AR(3)</td><td>2.0641</td><td>0.5592</td></tr> <tr> <td>AR(4)</td><td>5.6481</td><td>0.2270</td></tr> </tbody> </table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.6193	0.4313	AR(2)	1.6161	0.4457	AR(3)	2.0641	0.5592	AR(4)	5.6481	0.2270
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.0127	0.9102																																																						
AR(2)	1.1032	0.5760																																																						
AR(3)	1.1252	0.7710																																																						
AR(4)	3.7350	0.4431																																																						
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.1525	0.6961																																																						
AR(2)	0.8649	0.6489																																																						
AR(3)	0.9881	0.8041																																																						
AR(4)	4.2197	0.3771																																																						
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.6193	0.4313																																																						
AR(2)	1.6161	0.4457																																																						
AR(3)	2.0641	0.5592																																																						
AR(4)	5.6481	0.2270																																																						

CLI <table><tr><th colspan="3">Godfrey's Serial Correlation Test</th></tr><tr><th>Alternative</th><th>LM</th><th>Pr > LM</th></tr><tr><td>AR(1)</td><td>0.3101</td><td>0.5776</td></tr><tr><td>AR(2)</td><td>4.1759</td><td>0.1239</td></tr><tr><td>AR(3)</td><td>5.6672</td><td>0.1290</td></tr><tr><td>AR(4)</td><td>6.3698</td><td>0.1732</td></tr></table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.3101	0.5776	AR(2)	4.1759	0.1239	AR(3)	5.6672	0.1290	AR(4)	6.3698	0.1732	HPI <table><tr><th colspan="3">Godfrey's Serial Correlation Test</th></tr><tr><th>Alternative</th><th>LM</th><th>Pr > LM</th></tr><tr><td>AR(1)</td><td>0.8221</td><td>0.3646</td></tr><tr><td>AR(2)</td><td>1.9305</td><td>0.3809</td></tr><tr><td>AR(3)</td><td>2.1947</td><td>0.5330</td></tr><tr><td>AR(4)</td><td>6.1207</td><td>0.1903</td></tr></table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.8221	0.3646	AR(2)	1.9305	0.3809	AR(3)	2.1947	0.5330	AR(4)	6.1207	0.1903	K3 <table><tr><th colspan="3">Godfrey's Serial Correlation Test</th></tr><tr><th>Alternative</th><th>LM</th><th>Pr > LM</th></tr><tr><td>AR(1)</td><td>2.6216</td><td>0.1054</td></tr><tr><td>AR(2)</td><td>2.6409</td><td>0.2670</td></tr><tr><td>AR(3)</td><td>3.8825</td><td>0.2744</td></tr><tr><td>AR(4)</td><td>9.0102</td><td>0.0608</td></tr></table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	2.6216	0.1054	AR(2)	2.6409	0.2670	AR(3)	3.8825	0.2744	AR(4)	9.0102	0.0608
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.3101	0.5776																																																						
AR(2)	4.1759	0.1239																																																						
AR(3)	5.6672	0.1290																																																						
AR(4)	6.3698	0.1732																																																						
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.8221	0.3646																																																						
AR(2)	1.9305	0.3809																																																						
AR(3)	2.1947	0.5330																																																						
AR(4)	6.1207	0.1903																																																						
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	2.6216	0.1054																																																						
AR(2)	2.6409	0.2670																																																						
AR(3)	3.8825	0.2744																																																						
AR(4)	9.0102	0.0608																																																						
M1 <table><tr><th colspan="3">Godfrey's Serial Correlation Test</th></tr><tr><th>Alternative</th><th>LM</th><th>Pr > LM</th></tr><tr><td>AR(1)</td><td>0.0000</td><td>0.9992</td></tr><tr><td>AR(2)</td><td>3.5287</td><td>0.1713</td></tr><tr><td>AR(3)</td><td>3.8755</td><td>0.2752</td></tr><tr><td>AR(4)</td><td>6.0283</td><td>0.1970</td></tr></table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.0000	0.9992	AR(2)	3.5287	0.1713	AR(3)	3.8755	0.2752	AR(4)	6.0283	0.1970	OBX <table><tr><th colspan="3">Godfrey's Serial Correlation Test</th></tr><tr><th>Alternative</th><th>LM</th><th>Pr > LM</th></tr><tr><td>AR(1)</td><td>0.1954</td><td>0.6585</td></tr><tr><td>AR(2)</td><td>2.3148</td><td>0.3143</td></tr><tr><td>AR(3)</td><td>2.4852</td><td>0.4780</td></tr><tr><td>AR(4)</td><td>4.3805</td><td>0.3570</td></tr></table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.1954	0.6585	AR(2)	2.3148	0.3143	AR(3)	2.4852	0.4780	AR(4)	4.3805	0.3570	PMI <table><tr><th colspan="3">Godfrey's Serial Correlation Test</th></tr><tr><th>Alternative</th><th>LM</th><th>Pr > LM</th></tr><tr><td>AR(1)</td><td>0.1408</td><td>0.7075</td></tr><tr><td>AR(2)</td><td>0.8834</td><td>0.6429</td></tr><tr><td>AR(3)</td><td>1.0721</td><td>0.7838</td></tr><tr><td>AR(4)</td><td>2.8243</td><td>0.5876</td></tr></table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	0.1408	0.7075	AR(2)	0.8834	0.6429	AR(3)	1.0721	0.7838	AR(4)	2.8243	0.5876
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.0000	0.9992																																																						
AR(2)	3.5287	0.1713																																																						
AR(3)	3.8755	0.2752																																																						
AR(4)	6.0283	0.1970																																																						
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.1954	0.6585																																																						
AR(2)	2.3148	0.3143																																																						
AR(3)	2.4852	0.4780																																																						
AR(4)	4.3805	0.3570																																																						
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	0.1408	0.7075																																																						
AR(2)	0.8834	0.6429																																																						
AR(3)	1.0721	0.7838																																																						
AR(4)	2.8243	0.5876																																																						
Spread <table><tr><th colspan="3">Godfrey's Serial Correlation Test</th></tr><tr><th>Alternative</th><th>LM</th><th>Pr > LM</th></tr><tr><td>AR(1)</td><td>3.8528</td><td>0.0497</td></tr><tr><td>AR(2)</td><td>4.0756</td><td>0.1303</td></tr><tr><td>AR(3)</td><td>5.9826</td><td>0.1125</td></tr><tr><td>AR(4)</td><td>7.6910</td><td>0.1036</td></tr></table>	Godfrey's Serial Correlation Test			Alternative	LM	Pr > LM	AR(1)	3.8528	0.0497	AR(2)	4.0756	0.1303	AR(3)	5.9826	0.1125	AR(4)	7.6910	0.1036																																						
Godfrey's Serial Correlation Test																																																								
Alternative	LM	Pr > LM																																																						
AR(1)	3.8528	0.0497																																																						
AR(2)	4.0756	0.1303																																																						
AR(3)	5.9826	0.1125																																																						
AR(4)	7.6910	0.1036																																																						

B In-sample normality





C In-sample Ramsey's RESET test

EBITDA <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>1.9975</td><td>0.1629</td></tr> <tr> <td>3</td><td>1.6223</td><td>0.2064</td></tr> <tr> <td>4</td><td>1.2822</td><td>0.2894</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	1.9975	0.1629	3	1.6223	0.2064	4	1.2822	0.2894	Employees <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>1.6294</td><td>0.2069</td></tr> <tr> <td>3</td><td>1.0536</td><td>0.3554</td></tr> <tr> <td>4</td><td>1.9353</td><td>0.1343</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	1.6294	0.2069	3	1.0536	0.3554	4	1.9353	0.1343	Composite <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>2.7139</td><td>0.1049</td></tr> <tr> <td>3</td><td>2.0400</td><td>0.1394</td></tr> <tr> <td>4</td><td>1.6813</td><td>0.1814</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	2.7139	0.1049	3	2.0400	0.1394	4	1.6813	0.1814
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	1.9975	0.1629																																													
3	1.6223	0.2064																																													
4	1.2822	0.2894																																													
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	1.6294	0.2069																																													
3	1.0536	0.3554																																													
4	1.9353	0.1343																																													
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	2.7139	0.1049																																													
3	2.0400	0.1394																																													
4	1.6813	0.1814																																													
Dividends <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>2.5215</td><td>0.1177</td></tr> <tr> <td>3</td><td>2.4136</td><td>0.0986</td></tr> <tr> <td>4</td><td>1.9384</td><td>0.1339</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	2.5215	0.1177	3	2.4136	0.0986	4	1.9384	0.1339	Sales <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>2.3536</td><td>0.1304</td></tr> <tr> <td>3</td><td>1.3246</td><td>0.2740</td></tr> <tr> <td>4</td><td>1.1896</td><td>0.3220</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	2.3536	0.1304	3	1.3246	0.2740	4	1.1896	0.3220	Total Assets <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>3.3117</td><td>0.0740</td></tr> <tr> <td>3</td><td>2.2734</td><td>0.1122</td></tr> <tr> <td>4</td><td>1.5768</td><td>0.2052</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	3.3117	0.0740	3	2.2734	0.1122	4	1.5768	0.2052
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	2.5215	0.1177																																													
3	2.4136	0.0986																																													
4	1.9384	0.1339																																													
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	2.3536	0.1304																																													
3	1.3246	0.2740																																													
4	1.1896	0.3220																																													
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	3.3117	0.0740																																													
3	2.2734	0.1122																																													
4	1.5768	0.2052																																													

CLI <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>1.0658</td><td>0.3064</td></tr> <tr> <td>3</td><td>0.9909</td><td>0.3779</td></tr> <tr> <td>4</td><td>4.3924</td><td>0.0078</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	1.0658	0.3064	3	0.9909	0.3779	4	4.3924	0.0078	HPI <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>2.0000</td><td>0.1629</td></tr> <tr> <td>3</td><td>1.5736</td><td>0.2167</td></tr> <tr> <td>4</td><td>1.0610</td><td>0.3735</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	2.0000	0.1629	3	1.5736	0.2167	4	1.0610	0.3735	K3 <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>0.6928</td><td>0.4089</td></tr> <tr> <td>3</td><td>1.0639</td><td>0.3523</td></tr> <tr> <td>4</td><td>0.8700</td><td>0.4626</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	0.6928	0.4089	3	1.0639	0.3523	4	0.8700	0.4626
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	1.0658	0.3064																																													
3	0.9909	0.3779																																													
4	4.3924	0.0078																																													
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	2.0000	0.1629																																													
3	1.5736	0.2167																																													
4	1.0610	0.3735																																													
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	0.6928	0.4089																																													
3	1.0639	0.3523																																													
4	0.8700	0.4626																																													
M1 <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>0.6744</td><td>0.4151</td></tr> <tr> <td>3</td><td>0.4454</td><td>0.6430</td></tr> <tr> <td>4</td><td>0.5600</td><td>0.6438</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	0.6744	0.4151	3	0.4454	0.6430	4	0.5600	0.6438	OBX <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>2.7006</td><td>0.1057</td></tr> <tr> <td>3</td><td>1.4637</td><td>0.2400</td></tr> <tr> <td>4</td><td>1.0005</td><td>0.3995</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	2.7006	0.1057	3	1.4637	0.2400	4	1.0005	0.3995	PMI <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>0.6974</td><td>0.4119</td></tr> <tr> <td>3</td><td>0.8487</td><td>0.4409</td></tr> <tr> <td>4</td><td>2.0043</td><td>0.1428</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	0.6974	0.4119	3	0.8487	0.4409	4	2.0043	0.1428
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	0.6744	0.4151																																													
3	0.4454	0.6430																																													
4	0.5600	0.6438																																													
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	2.7006	0.1057																																													
3	1.4637	0.2400																																													
4	1.0005	0.3995																																													
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	0.6974	0.4119																																													
3	0.8487	0.4409																																													
4	2.0043	0.1428																																													
Spread <table border="1"> <thead> <tr> <th colspan="3">Ramsey's RESET Test</th></tr> <tr> <th>Power</th><th>RESET</th><th>Pr > F</th></tr> </thead> <tbody> <tr> <td>2</td><td>0.6216</td><td>0.4340</td></tr> <tr> <td>3</td><td>0.7296</td><td>0.4870</td></tr> <tr> <td>4</td><td>2.7082</td><td>0.0548</td></tr> </tbody> </table>	Ramsey's RESET Test			Power	RESET	Pr > F	2	0.6216	0.4340	3	0.7296	0.4870	4	2.7082	0.0548																																
Ramsey's RESET Test																																															
Power	RESET	Pr > F																																													
2	0.6216	0.4340																																													
3	0.7296	0.4870																																													
4	2.7082	0.0548																																													