M.Sc. Final Project
Financial Economics

A Business Cycle Analysis with Large Factor Model
Construction of a smooth indicator free from short run dynamics

Tómas Michael Reynisson

HÁSKÓLI ÍSLANDS

A Business Cycle Analysis with Large Factor Model

Construction of a smooth indicator free from short run dynamics

Tómas Michael Reynisson

Final project towards M.Sc. degree in Financial Economics

Instructor: Dr. Helgi Tómasson, Professor at the University of Iceland

Department of Economics

School of Social Science, University of Iceland

February 2012

A Business Cycle Analysis with Large Factor Model.

This thesis equals 30 ECTS credits towards M.Sc. degree at the School of Social Science, Universisty of Iceland.

Preface

This paper is written under the instruction of professor Helgi Tómasson and counts for 30 ECTS credits. I want to thank him for providing me with extra reading material that gave me a better and deeper understanding of the methods used in this paper.

Finally, I am forever grateful for the unconditional support provided by my parents, Reynir and Monica. You guys are amazing.

Abstract

In this paper, a large factor model is utilized with the aim of isolating the medium-to-long-run component (MLRC) from a large data set which could potentially describe the underlying data generating process, in this case gross domestic product (GDP). The central idea is to extract a few common factors from all the data set with as much variance as possible. To do so, the use of a particular kind of principal components analysis is employed, which is specifically designed to extract from the data set the common information of the data. More precisely, the linear combinations of the data. Removal of the short-run dynamics from a stationary time series to isolate MLRC can be done with a simple bandpass filter.

Quarterly GDP is not timely, it suffers from heavy short-run component and is under constant revision. The main objective of this paper is to construct a timely indicator which could better describe the current economic state than GDP does. For this to be reliable, the data set is made contemporaneous and synchronized in a way that all time series in the data set are describing the same process at the same time.

Further, the data set is on a monthly basis which makes it even more timely than the quarterly GDP. Therefore, the Icelandic Coincident Indicator (ICI) is on a monthly basis which is an estimate of the MLRC of GDP, in real time.

The main findings are that ICI captures the underlying data generating process of GDP rather nicely, it is smooth with less volatility but still manages to describe the average growth of GDP.

Contents

List of Figures

List of Tables

1   Introduction

The importance of finance on modern macroeconomics has been widely and broadly studied for decades. Despite of the large literature and research, financial crises continue to occur, and there is some evidence that they have become more frequent and more severe in recent years (Caprio and Klingebiel, 2002). Kroszner (2007) analyzed 38 countries experiencing a banking crisis over the last 25 years and found that each country, on average, experienced 1.6 % greater contraction in GDP growth. Financial development and economic uncertainty also have contributed to the increase in economic volatility and the number of crises around the world. Economic volatility or fluctuations can be measured through so-called business cycle analysis which is a measures the state of the economy.

The earliest theories of business cycles go as far back as 1936 where one of the greatest economists of the twentieth century John Maynard Keynes and his theory of *General Theory of Employment, Interest And Money* from 1936 emerged (Keynes, 1937). It was truly the first macroeconomic model and made way for models emphasizing the effects of finance on macroeconomic variables and fluctuations. Its explanation of macroeconomic volatility and its implied involvement for active government in the economy was probably the biggest influence at the time. "Keynes general theory is the first model to make a real distinction between microeconomics and macroeconomics and look at the interaction between the goods, labor, money and financial asset markets simultaneously." (Knoop, 2008)

As the years passed, it became increasingly apparent that the Keynesian framework did not sufficiently explain what happens during a business cycle nor did it seem capable of providing answers to what happens when sudden changes in monetary or fiscal policy occur (Dungey, 2009). For example, the 1970s was an era of economic expansion with high rates of inflation in which Keynesian methods simply could not provide theoretical answers. This huge inflation season was better predicted by other models than Keynesian models (Knoop, 2008). So, the need for producing accurate and reliable forecasts of macroeconomic variables became the driving force for empirical research. Much effort was put down in developing forecasting models that could better predict the economy utilizing data with higher frequency, such as monthly and weekly data. The family of univariate time series models like Box and Jenkins (1970), multivariate time series models like vector autoregressive (VAR) (Sims, 1980) models and cointegration systems (Engle and Granger, 1987) are the fruit of these efforts.

Two big papers from Milton Friedman (1970) and Robert Lucas (1976) changed the aspects of macroeconomics, especially the aspects of the old and out-dated methods of Keynes. Friedman, one of the most influential economists in the 20th century, implemented the *Monetarism* with his theories. Monetarists believe that fluctuations in aggre-

gate demand can have real effect on output and drive business cycles. Friedman stated that money only serves as a medium of exchange and not as financial asset, like Keynes did, and only used to facilitate trade in the present time. He also believed that business cycle declines or accelerations were a monetary phenomenon and a direct consequence of misleading central bankers, so calling it a business cycle would be a misnomer.

Monetarists feel that central banks can avoid these policy swings and eliminate business cycles by changing their goal from economic stabiliziation to monetary stabilization. To sum up, business cycles in Keynesian models are caused by fluctuations in aggregate demand and monetary policy and fiscal policy also influence aggregate demand. Hence, there should be an opportunity for a policy maker to use these tools to offset aggregate demand fluctuations and stabilize growth. So, during periods of slow growth, the central bank should increase the money supply, and during perids of faster growth or rising prices, the central bank should reduce inflation by cutting money growth. Thus, central banks believe monetary policy should be consistent with the Philips curve. Later, Keynes himself admitted that there was significant practical problems in using monetary policy, and stabilization policy may not work in theory nor in practice. An important question about stabilization policy and expectations about the economy was raised by Robert Lucas (1976). He argued that expectations may be impossible to predict in any given circumstance and any forecasts of the future based on data from the past will be unreliable. This statement is now known as the *Lucas Critique.*

This critique raised important questions about stabilization policy for several reasons. First, there are lags in the monetary policy during the decision making process, the money multiplier process, and the monetary transmission process that can take an uncertain amount of time. As a result, accurate forecasts of downturns are critical so that the policy can be enacted in a timely manner before a downturn actually begins. Second, the effects from changes in policy must be fully understood because if the impact of the policy is not known then it may be impossible to predict the impact of the stabilization policy on the economy (Knoop, 2008).

Impacts on economic policy with econometric methods have risen dramatically since the 80s. Sargent (1977; 1978; 1979) made way for business cycle analysis with econometric analysis by employing linear models. These models were simplistic but made way for more advanced models of how modern macroeconomics works and gave a deeper understanding of how financial systems are built.

With the seminal work of Kydland and Prescott (1982) there was a breakthrough in business cycle analysis where they employed nonlinear models. The model economy was perfectly competitive and frictionless, with prices and quantities immediately adjusting to their optimal levels after shock. The model was largely adopted by macroeconomists who introduced several sophistications over the years, exploring its theoretical and empirical

possibilities. This became known as the *Real Business Cycle (RBC)* approach to macroeconomic modelling, constituting a crucial advance in modern macroecconomics, by firmly establishing a dynamic and stochastic models as the new paradigm of macroeconomic theory (Almeida, 2009). Econometric analysis of business cycle models was not really an option for Kydland and Prescott. But over time, computers have become more powerful, theory has become richer, datasets have become larger and most but not least, econometric methods have become more advanced. Later in 2004, Kydland and Prescott received the Nobel Memorial Prize in Economics for their contributions to dynamic macroeconomics: The time consistency of economic policy and the driving forces behind business cycles (Levinovitz and Ringertz, 2001).

Commonly used estimation procedures for business cycle analysis are principal component methods (Stock and Watson, 1998), state space models ((Harvey, 1989) and (Stock and Watson, 1998)) and cointegration frameworks (Gonzalo and Granger, 1995)). Recent empirical studies, however, take a different route and directly examine the evolution of comovement properties of the main macroeconomic aggregates over time. The results of these studies indicate that differences in country coverage, sample periods, aggregation methods used to create country groups, and econometric methods employed could lead to diverse conclusions about the temporal evolution of business cycle synchronization (Kose et al., 2008).

Stock and Watson (2005) use a structural VAR model to analyze the importance of international factors in explaining business cycles in the G-7 countries since 1960. They conclude that comovement has fallen due to diminished importance of common shocks. In contrast, other studies document that business cycle linkages have become stronger over time (Kose et al., 2003).

One recently developed approach, utilized in this paper, is a factor model analysis of business cycles as proposed by Forni and Reichlin (1998), Forni et al. (2000) and Altissimo et al. (2010). Over the last years, there have been a growing number of forecasts of macroeconomic variables which rely on factor models and dynamic factor models (see e.g. (Stock and Watson, 1998), (Gosselin and Tkacz, 2001), (Artis et al., 2005)). Dynamic factor models will not be considered in this research.

This paper uses the factor model framework which accommodates a large cross-section of macroeconomic time series. The rational behind the use of this model is to extract few common factors that drive all series included in a given dataset, and construct a smooth indicator with less volatility but is still able to capture the average growth of GDP. Also, the forecasting properties of the indicator will be analyzed.

This study is organized as follows. Chapter 2 reviews the concepts concerning business cycles and gives overview of previous analysis of the Icelandic business cycle. Chapter 3 defines some topics about the dataset and how it is manipulated. The section also defines

our target — the medium-to-long-run component of GDP and discusses its spectral interpretation. The central point of this paper is in Chapter 4 where the model attributes are discussed in detail along with a short chapter about the forecasting properties. Chapter 5 presents the results and chapter 6 discusses the main findings. Chapter 7 concludes and presents future research.

2 Business cycle definitions and notations

The term *business cycle* refers to fluctuations in production or overall economic activity over time, also referred to aggregate output. Because the business cycle is related to aggregate economic activity, a popular indicator of the business cycle is the real gross domestic product (GDP). These fluctuations move up or down around a long term growth trend and typically shift over time between periods of economic growth or stagnation. Economic growth involves periods of expansion or boom, but periods of stagnation involve contraction or recession (Forni and Reichlin, 1998). Business cycles were thought to be regular in nature, with predictable durations, but today they are widely believed to be irregular, varying in frequency, magnitude and duration. There are countless statistical and mathematical models that have been constructed for business cycle analysis trying to capture its dynamics. Even though they are named cycles, these fluctuations do not follow a mechanical or predictable periodic pattern, hence the wide literature studies of business cycle analysis.

Over time, GDP will change for many reasons, economic and non-econonmic. To name a few, economic reasons involve changes in government policies such as taxes and interest rates but the non-economic reasons involve exogenous factors such as natural disasters, wars, or even nuclear meltdowns.

There are four stages in the business cycle:

1. Peak - When the economy is growing at its maximum.

2. Contraction - When the economy starts slowing down.

3. Trough - When the economy hits bottom, usually in a recession.

4. Expansion - When the economy starts growing again.

The exact end of an expansion is called the *peak* of the business cycle and the exact end of a recession is called the *trough* of the business cycle. Figure 1 shows how a typical business cycle moves from one stage to another.

2.1 Cause and impact

Increase in economic volatility and the number of crises around the world are mostly due to financial development and economic uncertainty. Economic crises have become more frequent over the past two decades, which is a common perception amongst the public, including world leaders and politicians (Knoop, 2008).

It is true that financial crisis arounde the world have been increasing over the last two decades in which the financial systems have played a key role. A good example of this financial instability is the East Asisan crisis between 1997 and 1999, in which currency crisis. asset market crashes, and banking collapses precipitated deep recessions in
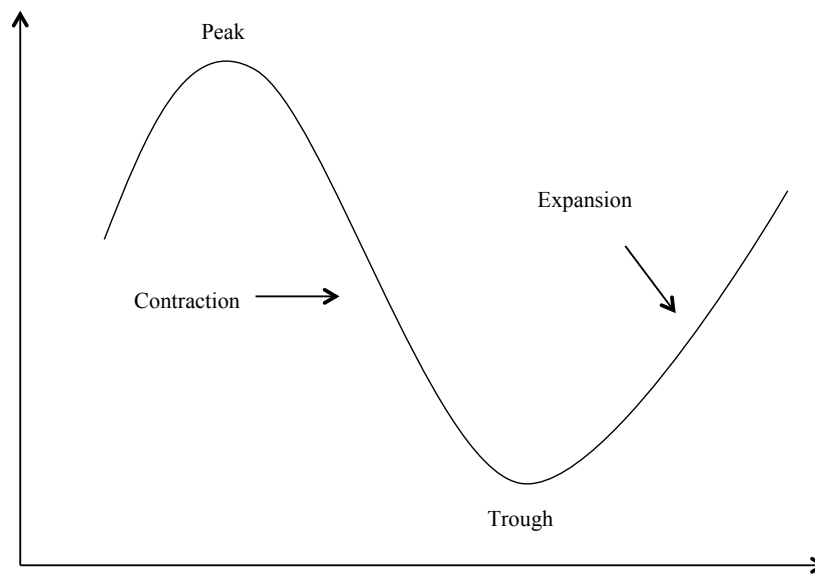
Figure 1: The four stages of the business cycle.

South Korea, Thailand, Indonesia, and Malaysia (Knoop, 2008). For smaller emerging economies, financial volatility has also been a major influence for its economy. For example, financial markets crashed in Chile 1982 and in Mexico 1994 because they issued a substantial amount of debt denominated in foreign currency. Because of uncertainty about the future value of the domestic currency, it led to great financial instability in these emerging markets with overwhelming debt burdens for firms which led to major financial crisis (Mishkin, 2000). Financial volatility was an even greater influence in 2000/2001 when the world experinced a boom/bust cycle in the stock market and a steep decline in investment yielding a huge recession. An important fact about the business cycle is that the worst economic contractions are typically associated with severe and short periods of financial instability (Mishkin, 2000)

It is clear that finance affects macroeconomics on a large scale and the understanding of how financial systems work and what impact they have on macroeconomics is of great importance. The value of developing theories of business cycles or developing policies to combat macroeconomic intstability is important for the well-being of every economy. Over the last 20 years or so, economists have increasingly turned their head to models of market failure to explain business cycle theory and other macroeconomic phenomena. "These *New Institutinal Theories of Finance* are reshaping macroeconomics in general, finance and business cycle theory in particular." (Knoop, 2008)

2.2 Why study them?

The importance of studying business cycles in general is of great value for every economy. There are several reasons. Firstly, if an economy is suffering from a severe recession, it will be extremely costly which means higher rates leading to lower income for everybody. Secondly, recessions follows higher suicide rates, higher crime rates, more divorces, tax evasions etc. Over 100 years ago, Durkheim (1952) recognised that changes in social relations, such as economic crises or periods of war, contribute to changing patterns of suicide. Countless research has been conducted and for example, Gunnel (1999), Gunnel (2003), Lester (1997), and Neeleman (1997) have confirmed and refined Durkheim's theories.

Amongst other reasons are the opportunities for economists and other policymakers to learn from what happens to the economy when recession and depression occurs. During a contraction, market functions begin to break down and, much like any piece of machinery, it is often easier to identify exactly how underlying mechanism operate when they are not functioning properly than when they are. As a result, recessions and depressions present unique learning opportunities for economists, offering them the chance to "peak under the hood" of an economy. (Knoop, 2008).

The situation that shocked the Icelandic economy in October 2008, should then have been a valuable opportunity for policy and monetary makers and even the Icelandic economy as a whole. It is important for every economy to know how to react when it faces recessions or booms and provide reliable forecasts of its path and proper solutions. Therefore, the value of reliable business cycle analysis is huge and the importance of a *"nice"* macroeconomic policy that can control recessions or make it less volatile has an even greater value and the opportunity to make human welfare more pleasurable.

2.3 Characteristics of business cycles

As mentioned before, GDP is the most commonly used indicator when it comes to business cycle analysis. But GDP has its flaws. Firstly, GDP is only available on a quarterly basis which reduces the frequency and datapoints. Secondly, it is under constant change and revision which makes GDP rather unreliable. Thirdly, GDP can sometimes be available with irregular or long intervals when trying to acces GDP back in time. This could be a serious issue when for example someone wishes to asses the impact of earlier crises.

However, there is a way of mimicking the cyclical behavior of GDP by choosing variables with higher frequency , e.g. on a monthly basis, which have high correlation with GDP. Those variables are called *coincident variables*. Variables that are available on a monthly basis that mimic this cyclical behavior can be selected and classified as coincident, leading or lagging variables. Or more formerly:

- High correlation between a variable X_t and GDP at some lag $k > 0$, is a potential leading variable.

- High correlation between a variable X_t and GDP with lag $k = 0$, is a potential coincident variable.

- High correlation between a variable X_t and GDP with lag $k < 0$, is a potential lagging variable.

Where k represents lag(s).

As the name suggests, leading variables mimic the cyclical behavior of GDP but precede it, but lagging variables do not contain any early information of the present or future economic state of the economy. Other methods of selecting variables that mimic GDP involve consistent timing, economic significance, spectral methods, and lead time in turning points. "Most variables of interest in growth theory and business cycle theory, coincide" Levinovitz and Ringertz (2001), so selecting "nice" coincident variables can be a succesful proxy for GDP.

2.4 Short real business cycle literature

The idea behind real business cycle (RBC) models is their emphasis on the role of real shocks, particularly technology shocks, in driving business fluctuations as first presented by Kydland and Prescott (1982). Other known shocks are monetary, fiscal, and oil price shocks but Prescott (1986) argues that technology shocks account for more than half the fluctuations. The idea that technology shocks are the central driver of business cycles is controversial. Prescott (1986) computes total factor productivity (TFP) and treats it as a measure of exogenous technology shocks. But as Hall (1988) and Evans (1992) show, TFP, as computed by Prescott (1986), is not a pure exogenous shock, but has some endogenous components (Rebelo, 2005) [1]. RBC models also became a point of departure for many theories in which technology shocks do not play a central role (Rebelo, 2005).

Later, RBC models came to be used for policy analysis in general and for the study of optimal fiscal and monetary policy in particular [2].

2.5 The Icelandic business cycle

In figure 2 we can see the quarterly GDP growth or business cycle fluctuations in Iceland for the period 1997 - 2011, including shaded areas that represent periods associated

---

[1] An extensive survey about real business cycle models is available in Rebelo (2005) and King (1999).

[2] See Chari and Kehoe (1999) for a review of the literature on optimal fiscal and monetary policy in RBC models.

with recessions. A recession is typically considered as declines for two quarters in a row (Petursson, 2000).
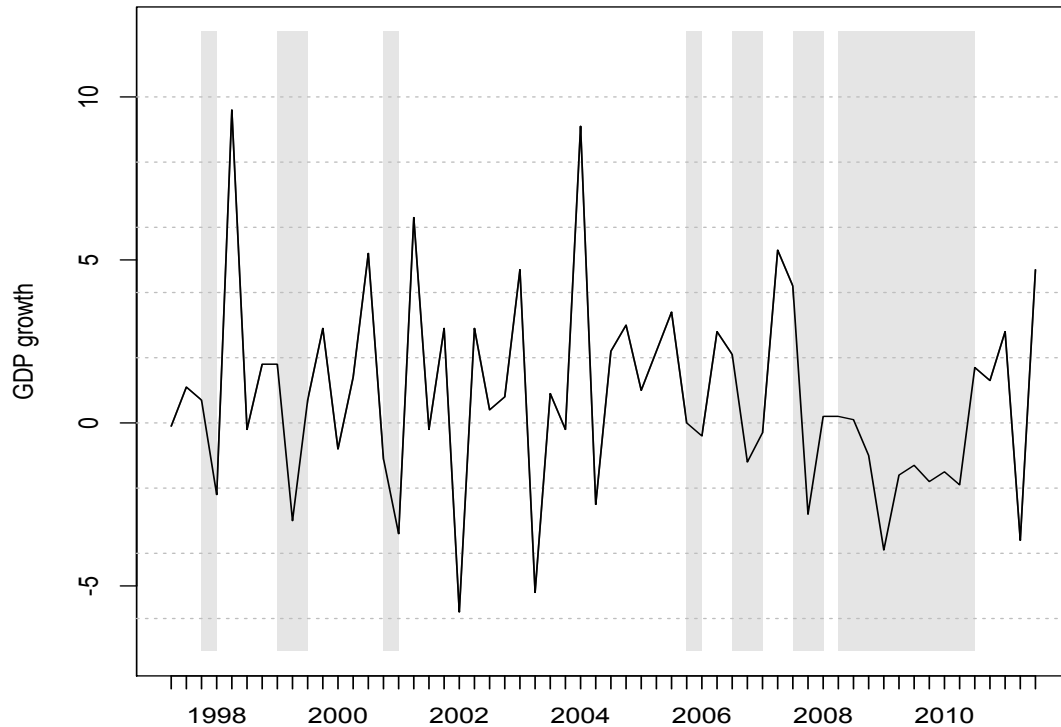
Figure 2: Quarterly GDP growth 1997 - 2011, seasonally adjusted. Shaded areas represent recessions. Source: Statistics Iceland and Datamarket

According to figure 2 there are seven expansions and seven recessions over the period analysed here. Business cycles have certain "stylized facts". One of those facts is that business cycles have a 2 - 6 year periodicity attached to them (Canova, 1998). That certainly seems to be the case for the Icelandic one. Another fact is that periods of expansion usually are longer than contractions and the Icelandic business cycle seems to have significantly longer expansions than contractions. The third fact is that aggregate fluctuations do not seem to exhibit any simple regular or cyclical patterns and business cycle movements seem to be assymetric (Petursson, 2000).

If we focus on the downturns of the economy, then according to table 1 the average duration of a recession is about 3.8 quarters which means the average duration of an expansion is about 4.9 quarters as is typical for business cycles. This suggests assymetry.

Looking at the average growth rate in the downturn period then it is clear that the most serious recession occurs from late 2000 to early 2001 where the economy is probably suffering from the aftershock of the boom in early 2000.

Table 2 shows some descriptive statistics for real GDP growth for the whole period and two subperiods. The whole period returns an average growth of 0.74% and a standard deviation of 2.43% implying that aggregate fluctuations or volatility are severe. Table 2

Table 1: Periods of downturns for the Icelandic business cycle

Period	Duration	Average growth (%)
1997.Q4 - 1998.Q1	2	-0.75%
1999.Q1 - 1999.Q3	3	-0.50%
2000.Q4 - 2001.Q1	2	-2.25%
2006.Q3 - 2007.Q1	3	0.20%
2007.Q3 - 2008.Q1	3	0.53%
2008.Q2 - 2010.Q3	10	-1.10%

also shows that fluctuations are more severe in the first subperiod and average growth rate is higher as well. Gudmundsson and Sighvatsson (2000), and Petursson (2000) also report that volatility and growth rate are decreasing with time.

Table 2: Descriptive statistics for the Icelandic business cycle for the whole sample and two subperiods.

	1997.Q2 - 2011.Q3	1997.Q2 - 2004.Q1	2004Q2. - 2011.Q3
Mean	0.77%	0.98%	0.55%
Standard deviation	3.07%	3.62%	2.44%
Minimum	-5.77%	-5.77%	-3.92%
Maximum	9.64%	9.64%	5.30%
T (Obs.)	58	29	29

2.5.1 Overview of Icelandic business cycle analysis

Business cycle analysis is of key interest to central banks and other market participants and comes as no surprise that business cycle analysis has grown considerably over the last years. The Icelandic business cycle has been studied on a few occasions. To name a few and focusing from the year 2000, Petursson (2000) modelled the business cycle dynamics using a Markov switching model for yearly GDP for the period 1946 - 1998. There he suggests that the Icelandic business cycle dynamics and forecasting ability is better captured with Hamilton's (1989) Markov model than by an autoregressive representation.

Eklund (2007b) predicted recessions with leading indicators and provided probabilities for recession and expansion over a twelve month forecast horizon using monthly data based on the classic Stock and Watson (1989). The same year Eklund (2007a) tries to model and forecast the Icelandic business cycle with monthly variables, coincident and leading, using vector autoregressive (VAR) model. There he described a simple method

of selecting coincident and leading indicators based on their correlation structure towards GDP. VAR model was then specified under general least squares method for the 16 variables chosen, and the Icelandic business cycle was forecasted. VAR models provides one of the simplest model based frameworks for understanding the relationship between coincident and leading variables. Danielsson (2008) discusses accuracy in forecasting of macroeconomic time series in Iceland and analyzes forecast errors in macroeconomic variables and finds that the performance of forecasting for these variables has improved on some measures. He also mentions that there is clear and statistically significant bias in economic forecasting of GDP in Iceland among institutions like the Central Bank of Iceland and National Economic Institute.

There is also extensive literature about the Icelandic business cycle and real business cycle publications in Working Papers from the Institute of Economic Studies of University of Iceland. One of them written by Einarsson and Marquis (2001), made it to the Journal of Money, Credit, and Banking. There they developed a theoretical model in which bank intermediation of working capital in finance is countercyclical and show that this model is consistent with U.S. data. The paper further analyzes various shocks to the model, as is custom to real business cycle models and theories.

Helgi Tómasson (1989) was one of the first researchers in Iceland that employed spectral analysis on the business cycle. He analyzed the cyclical properties of economic growth in OECD countries for the period 1961-1989 with an AR(2) model. There he finds out that the Icelandic business cycle is among the most unstable in the whole OECD countries for the early parts, but at the end it is around average. Most of the countries show a decreasing level of growth in that period.

3 Data treatment

There is no doubt that the real GDP is the most important business cycle indicator that measures the aggregate state of an economy, and estimation of the current state of the economy with variables other than GDP, needs a significant amount of data. The National Bureau of Economic Research (NBER) is the official institution in the U.S. that closely follows economic expansions and contractions, or more precisely, conducts business cycles analysis on a regular basis. A number of business cycle analysis has been conducted over the last 20 years. NBER analyzes economic indicators to determine the phases of the business cycle and use quarterly GDP growth rates as the primary indicator of economic activity. It also uses monthly figures, such as unemployment, personal income, production and retail sales.

That is precisely the main objective of the Icelandic Coincident Indicator (ICI), that is, to make an asessment of the Icelandic economic activity that is:

a) comprehensive and easy to interpret

b) non-dependent

c) timely

d) smooth and free from short-run fluctuations

Requirement *a)* and *b)* are obvious and required. Regarding requirements *c)* and *d)*, it is important for economic policymakers to decide whether it is a beginning of a long recession or just a short-lived phenomenon. None of the available macroeconomic time series provide the necessary requirements listed above. As mentioned before, GDP is probably the most used indicator of the aggregate state of every economy. But it has its flaws. GDP is only available quarterly with a long delay and hence is not timely [3]. For example, the first quarter estimate of GDP does not beome available until mid May. Moreover, the GDP is affected by a considerable short-run component which needs to be eliminated. Distinguishing between long, medium and short waves of a time series is relatively easy with a band-pass filter. The goal is to remove fluctuations of a period shorter than or equal to one year and leaving behind *the medium- to long-run component* (MLRC) which is our ideal target. That will be the discussion of chapter 3.3.

3.1 Data delay

One of the major changes in business cycle analysis and is introduced in Altissimo et al. (2010), is the concept of *data delay*. Macroeconomic time series have different calendar

---

[3]Yearly records are available from 1946.

releases and lack synchronism, and so, they have a significant end-of-sample unbalance because different series have different delays. The data therefore needs heavy revision and rearrangement in order to get a timely and reliable indicator of current ecomic activity. This problem is important to take care of real-time forecasting since variables with different timeliness would give poor end-of-sample balance and even worse forecasts.

GDP suffers from heavy delay. Thus, to analyze the current state of the economy, data with higher frequency needs to be considered, preferably monthly time series. This concept of data delay is best shown by a simple plot. Figure 3 shows the exchange rate of the ISK and the harmonized index consumer price (HICP) of household goods. Those time series clearly provide contradictory signals as can be seen by the shaded areas.
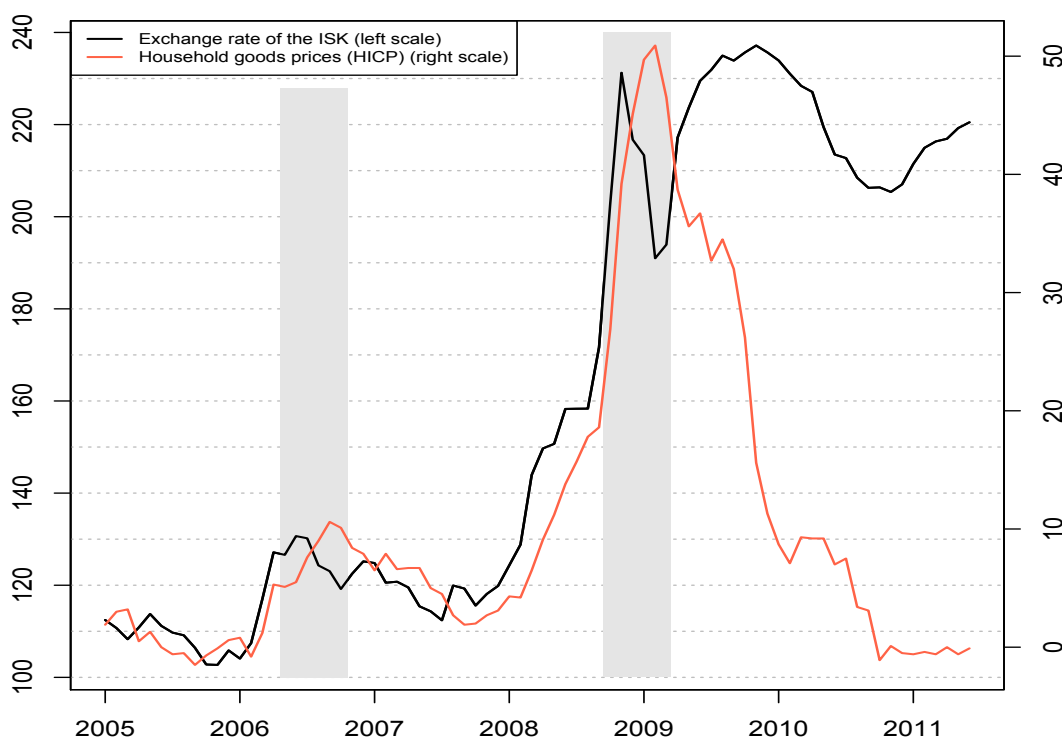
Figure 3: **Data delay of the exchange rate and harmonised price of household goods January 2005 - June 2011. Shaded areas represent examples of obvious delays. Source: Central Bank of Iceland, Eurostat and DataMarket.**

The calendar dates of some macroeconomic series are shown in table 3. As can be seen, even monthly statistics suffer from delay (end-of-sample unbalance).[4]

Financial variables are typically the most timely data with no delay while others suffer from various delays. Figure 3 shows how the exchange rate (no delay) is always couple of steps ahead of HICP price of household goods which suffer from large delays, typically around two months.

Survey data from Capacent Gallup were considered as a possible time series, as is done

---

[4]Delays are calculated in days.

Table 3: Calendar dates for various macroeconomic time series. Source: Statistics Iceland, (Altissimo et al., 2010) and own work.

Time	Feb 2011	Mar 2011	Apr 2011	May 2011	Jun 2011	Delay (days)
GDP	2010.Q4	2010.Q4	2010.Q4	2011.Q1	2011.Q1	45-90
Financial indicators	Feb 2011	Mar 2011	Apr 2011	May 2011	Jun 2011	0
CPI	Jan 2011	Feb 2011	Mar 2011	Apr 2011	May 2011	20-30
HICP	Des 2010	Jan 2011	Feb 2011	Mar 2011	Apr 2011	45-50
Import/Export	Des 2010	Jan 2011	Feb 2011	Mar 2011	Apr 2011	40-60
Estate market	Des 2010	Jan 2011	Feb 2011	Mar 2011	Apr 2011	60-70

in Altissimo et al. (2010), but unfortunately, the lack of time span made it impossible. Most survey data in Iceland has only been available on a monthly basis from 2004.

As the delays are represented in days, there needs to be some explaining to do on how it is converted to months. As soon as the calender date of the relevant variable exceeds the twentieth of month T, it gets a delay of $T-1$ like the consumer price indexes. Imports and exports get a delay of $T-2$ or $T-3$, and as before, data on financial markets are available up to time T such as exchange rates and some interest rates. "The GDP series is observed quarterly, so that its delay varies with time. For example, on April 20, only data up to the fourth quarter of the previous year are available; thus there is a three-month delay with respect to T, which is March. On May 20, the delay with respect to T is reduced to one month, as a first quarter preliminary estimate is released, and will be two months when T is May. Hence, an average delay of two months $(T-2)$." (Altissimo et al., 2010) In the next section we tackle the end-of-sample unbalance after some introduction to the dataset.

3.2   The dataset and GDP manipulation

The dataset is a panel of 67 monthly time series spanning the time period from January 1997 to September 2011 [5]. The final dataset used in this paper is the result of an extensive search of data regarding the Icelandic economy. Criteria for choice of data are sufficient time span, at least starting January 1997. The reason behind this is simply the fact that quarterly GDP is not available further back according to official records. Moreover, the dataset includes groups of variables that are leading, lagging and coincident with respect to GDP. In particular, the presence of leading variables, which contain information about future values of the GDP, is crucial to obtain a good estimate and to tackle the end of sample unbalance.

All series are transformed to remove outliers, seasonality and non-stationarity in that order. Outliers that are more than five standard deviations away from the mean are re-

---

[5]See Appendix 2 for list of data sources.

moved and replaced with the sample average of remaining observations. If an outlier appeared during the year 2008 it is probably not a good idea to remove them for obvious reasons, hence they were not removed. Seasonal adjustments from a given time series were removed as follows. Let x_t be the observed time series in which we can decompose into a trend component v_t, a seasonal component h_t, and a random component z_t. Then x_t will have the following interpretation:

$$x_t = v_t + h_t + z_t \tag{1}$$

Figure 4 gives an example of how it looks graphically. The observed time series represents the CIF value of car imports.
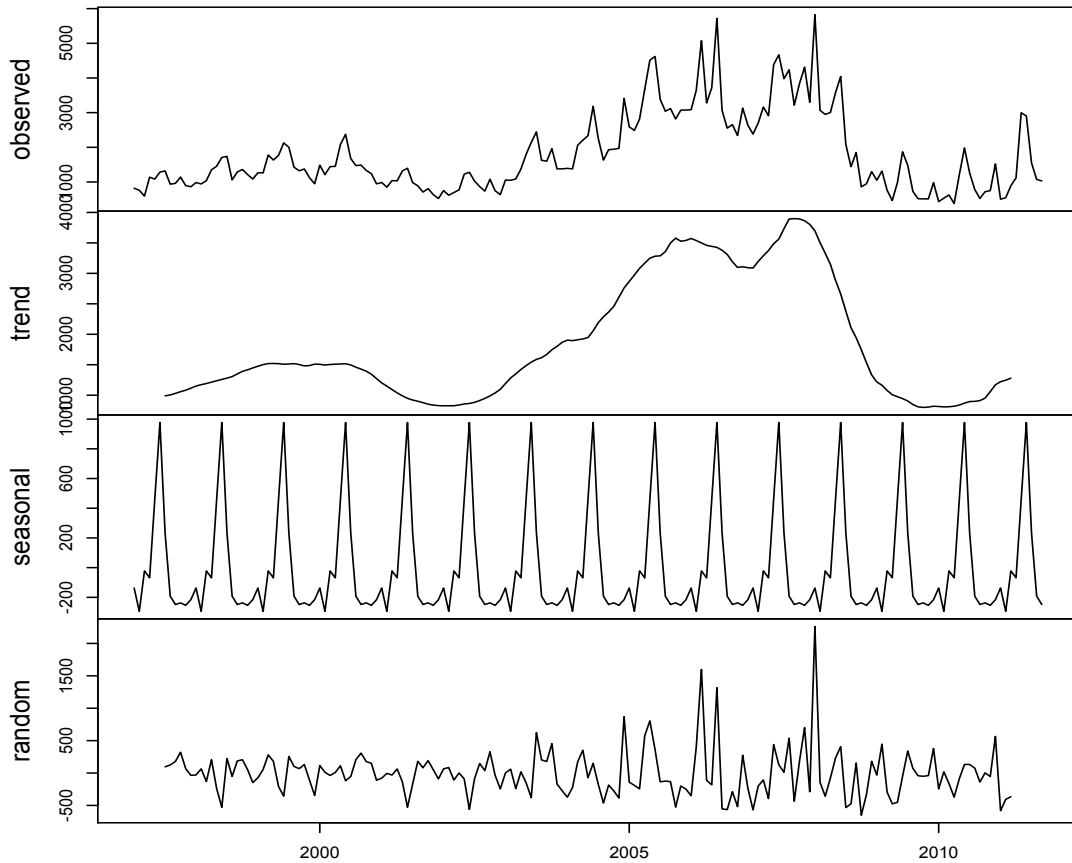
Figure 4: Decomposition of a time series into a trend, seasonal, and random component. The observed time series is the CIF value of car imports in millions of ISK. Source: Statistics Iceland, DataMarket and own work.

The goal is to remove the seasonal part from the observed series so the trend and random part remains, leaving a seasonal free time series. There is no need to resort to other more sophisticated procedures, e.g. Tramo-Seats or X12, since that would revise or shorten our time series too severely[6]. Non-stationarity was removed following an au-

---

[6]See (Findley and Martin, 2006) for a nice overview of those methods. There is also the method of regressing a set of dummy variables on the series.

tomatic procedure in a way that all series in a given category (e.g. demand indicators) were treated in the same way. A unit root test that was run afterwards using Dickey-Fuller test statistics, confirmed the method of categorizing the variables. Finally, the series were normalized subtracting the mean and dividing for the standard deviation as usually is done in the large *factor model literature*. A detailed list of the variables, data sources, and the related transformations are reported in Appendix 2.

Next up is the end-of-sample unbalance. Let x_{it}^* be a series in the panel after outliers and seasonality have been removed and appropriate transformation to achieve stationarity accomplished. Let k_i be the delivery delay in months for variable x_{it}^* so that when we are at the end of the sample, its last available observation is $x_{i,T-k_i}^*$. Then the panel x_{it} can be defined as

$$x_{it} = x_{i,t-k_i}^* \quad \forall i \in i = 1, \ldots, n. \tag{2}$$

which makes the last available observation of x_{it} at T for all i. This realignment and seasonal adjustments of course implies cutting some observations at the beginning and end of the sample for all time series. As a result, the dataset of the panel now goes from June 1997 until December 2010, hence $T = 160$.

Obviously, GDP is not available on a monthly basis and needs to be converted. There are numerous methods available for conversion of quarterly data to monthly data and issues of estimation of aggregate data, missing observations, and outliers have received considerable attention in the literature. The methods of Chow and Lin (1971), Denton (1971), Fernandez (1981), and Litterman (1983) give sophisticated methods of data interpolation and as far as literature on missing observations and outliers is concerned, a selected list of references includes Harvey and Pierse (1984), Kohn and Ansley (1986), Nijman and Palm (1986), and Gomez and Maravall (1994). All these methods are introduced by European Central Bank's Working Paper Series by Angelini, Henry, and Marcellino from 2003 which sooner became available in the Journal of Economic Dynamics and Control (2006). Therein, the method of Monte Carlo simulations, in order to interpolate and backdate data, is introduced[7]. Recently, a new method ermerged that involves Bayesian methods for completing missing data (e.g. GDP) in spatial models. There, Polasek et al. (2010) use Maximum Likelihood (ML) model of the spatial Chow-Lin method and propose a spatial econometrics model in a classical or Bayesian framework. Markov Chain Monte Carlo (MCMC) algorithms and predictions densities are then used for forecasting with encouraging results. Altissimo et al. (2010) simply suggests a simple linear interpolation and shows that there is no significant difference between linear interpolation or the more sophisticated methods of Cow and Lin (1971). This thesis follows their methods of simple

---

[7]See (Angelini et al., 2006) for further information.

linear interpolation.

To use our monthly data in the panel to obtain a timely GDP indicator, it is convenient to think of GDP as a monthly series with missing observations and employ linear interpolation between quarterly datapoints to get a monthly data series of GDP. The datapoint for month t, denoted by η_t , is defined as the aggregate of GDP for months t, $t-1$, and $t-2$, so that there is a two month overlap between two subsequent elements of the series. Obviously the monthly series is observable only for March, June, September, and December.

The monthly GDP growth rate y_t is then:

$$y_t = \log \eta_t - \log \eta_{t-3} \tag{3}$$

and thus defined for all months where quarter 1, quarter 2, quarter 3 and quarter 4 represent March, June, September and December respectively.

3.3 Spectral representation of GDP and its removal of short run fluctuations

As can be seen in figure 2, the GDP growth is rather rigid and not smooth. Identification of turning points (e.g. where the peak starts to contract or trough starts to expand) and forecasting ability of such volatile time series is rather difficult. Figure 2 is an example of a series which exhibits heavy short run fluctuations and might provide contradictory signals. A smoother series would make these calculations more efficient. In particular, GDP growth will be affected both by cyclical and by shorter-run movements, including seasonal and very short-run, high-frequency changes. Before the construction of the Icelandic Coincident Indicator (ICI) begins, higher frequencies should be washed out so the underlying medium- and long-run tendency of the economy can be described.

A popular way to remove the short run component of a time series and defining its medium-to-long-run component (MLRC), is by considering its spectral representation with a band-pass filter. Baxter-King (1999) (BK) filters belong to the category of band-pass filters that remove slow moving components and high frequency noise and are used in (Altissimo et al., 2010), (Forni et al., 2000), (Forni and Reichlin, 1998), and more.

That is precisely the goal of ICI. Defining c_t as our MLRC of the GDP growth which is free from short run component, means that fluctuations of periods shorter or equal to one year, need to be removed. This component, which is of course a smoothing of GDP growth, is our ideal target.

Now focusing on GDP (y_t) growth and assuming stationarity[8], y_t can be represented

---

[8]y_t is stationarity with $p-value < 0.01$.

as an integral of sine and cosine waves with frequency ranging between $-\pi$ and π, with respect to a stochastic measure. But, since we want to get rid of waves shorter than one year and since the filter is symmetric, we need to integrate over the interval between $-\pi/6$ and $\pi/6$ [9]. This threshold-value, so called, gets rid of high frequency waves and seasonality (if present).

Our MLRC, c_t, is the following infinite, symmetric, two-sided linear combination of the GDP growth series:

$$c_t = \beta(L)y_t = \sum_{k=-\infty}^{\infty} \beta_k y_{t-k}, \quad \beta_k = \begin{cases} \frac{sin(k\pi/6)}{k\pi} & \text{for} \quad k \neq 0 \\ 1/6 & \text{for} \quad k = 0 \end{cases} \tag{4}$$

More about the elaboration of the BK filter and construction of the ideal filter in Appendix A.

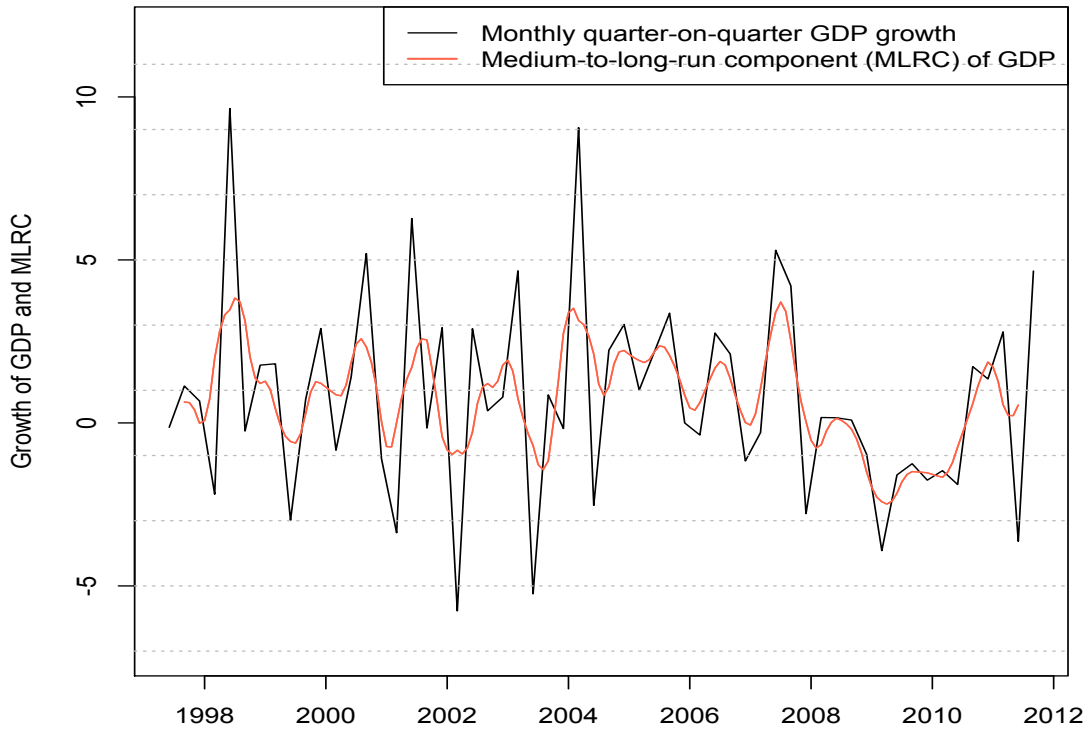**Figure 5: Quarterly GDP growth transformed to monthly frequency by linear interpolation. Also shown is the c_t. Source: Statistics Iceland and own work.**

The time series y_t therefore has the following decomposition.

$$y_t = c_t + s_t = \beta(L)y_t + [1 - \beta(L)]y_t \tag{5}$$

Where s_t includes all the waves of period shorter than 1 year and L is the lag operator

---

[9]$\pi/6 \approx 0.5$ year

where $L = y_t - y_{t-1}$ [10]. Since $\beta(1) = 1$, the mean of y_t is retained in c_t while the mean of s_t is 0. Further, c_t and s_t are orthogonal and the variance of y_t breaks down into short-run variance and medium- to long-run variance. c_t is our ideal target and is best visualized trough a plot as figure 5 shows.

Figure 5 clearly illustrates the smoothing effect of the filter. Now our target, c_t, is free from short-run waves and turning points are much more distinguishable and visual. The main task of this paper is a good estimate of c_t, so turning points can be detected and forecasted in real time. c_t closely tracks the GDP growth as it captures 76% of the total variance of y_t.

By comparing the year-on-year (y-o-y) GDP growth to c_t, some interesting relationship comes to life. Y-o-y growth is sometimes referred to as a measure of medium-to-long-run growth and is calculated as follows:

$$\tilde{y}_t = \frac{y_t + y_{t-3} + y_{t-6} + y_{t-9}}{4} \tag{6}$$

From equation 6, $\tilde{y}_t$ is then just a simple moving average of monthly GDP, and as figure 6 reveals, it is lagging with respect to c_t by several months. This could be a good predictor for future y-o-y growth.
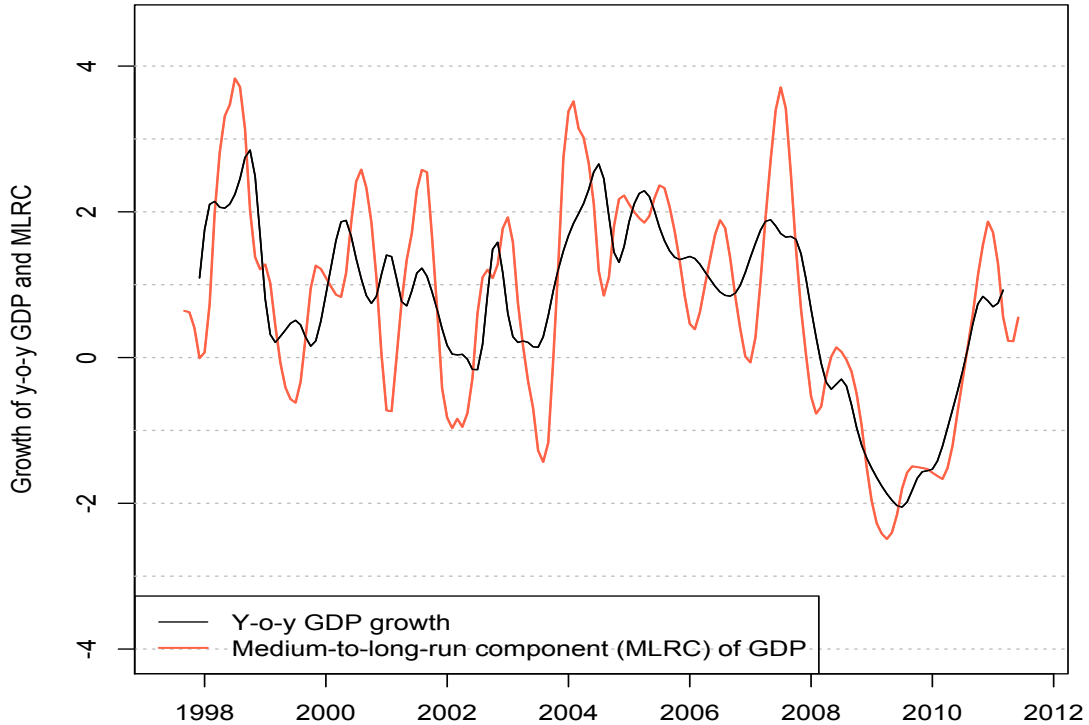

Figure 6: Year-on-year GDP growth and c_t

---

[10]See a note on lag operators in Appendix 1.

4   The Model

Factor analysis originated in psychometrics, and is used in behavioral sciences, social sciences, marketing, product management, operations research, and other applied sciences that deal with large quantities of data (Bartholomew et al., 2008). The idea behind factor analysis of business cycles, is by gathering a reasonable large amount of macroeconomic time series in one panel of data and extract some common factors from all the series that could potentially describe the behaviour of the underlying data generating process, in this case, GDP. Factor analysis has a range of methods for this extraction and probably the most common method is the *Principal Component Analysis (PCA).*

The number of time series in the analysis can be substantial, that is, contains a matrix of high-dimensional *space.* To view a cloud of data points in such high dimensions with the aim of finding a pattern is virtually impossible. PCA tries to find a *subspace* of reasonable dimension (much lower than the original dimensionality) so that the projection onto this subspace retains "as much information as possible" from all the series in the data set. In other words, PCA reduces the dimensions of the clouds of data points but still tries to capture most of the variability between the time series in the panel.

The challenge is then to develop techniques that are sufficiently powerful as to overcome this dimensionality problem, yet flexible enough to provide an adequate picture of economic reality (Forni et al., 2000). One of the main difficulties in multivariate statistical analysis is the "curse of dimensionality." In particular, the number of parameters of a parametric model often increase dramatically when the order of the model or the dimension of the time series is increased. Simplifying methods are often sought to overcome the curse of dimensionality. From an empirical viewpoint, multivariate data often exhibit similar patterns indicating the existence of common structure hidden in the data. Statistical factor analysis is one of those simplifying methods available in the literature. The aim of statistical factor analysis is to identify, from the observed data, a few factors that can account for most of the variations in the covariance or correlation matrix of the data (Tsay, 2005).

To sum, PCA tries to find a linear combination of time series in a given data set such that the maximum variance is extracted from each time series. It then removes this variances and tries to find a second linear combination which explains the maximum proportion of the remaining variance, and so on until the desired number of factors needed to explain the data or enough variance has been extracted. This will result in *orthogonal*, or uncorrelated factors. The number of factors or principal components will always be less than the original data set.

The model utilized in this paper is the factor model which was proposed by Forni and Reiclin (1998), Forni, Hallin, Lippi and Reichlin (2000) and (2005) which was further developed by Altissimo et al. (2010). Infact, they go couple of steps further and use spectral

decomposition of a covariance matrix and extract *dynamic* eigenvalues and eigenvectors with 100 different frequencies, hence the name: Dynamic Factor Model. This paper is only concerned about the eigen decomposition of the same covariance matrix.

There were two competing models around the millenium where static factor model of Stock and Watson (1999) and the fortold dynamic factor model were proposed. In both cases the goal is either estimation of the common component or forecasting of the x_{it}'s. The shortcoming of the static method is that it only exploits the information contained in the sample covariance matrix of x_{it}'s, whereas lagged covariances are ignored which means that the static method is not taking the dynamic structure of factors into account, at least not enough (Forni et al., 2005)

Forni, Hallin, Lippi and Reichlin (2000) fixed this problem but still had problems with forecasting since end-of-sample deterioration caused the filter to give mixed signals. Later, Altissimo et al. (2010) made the data set contemporaneous and aquired consistent estimate of the forecasting with lower Root Mean Square Error (RMSE) than in their previous studies years back.

Factor models can be very useful in adressing different economic issues. For instance, a factor structure is often assumed in both financial and macroeconomic literature to estimate insurable risk where the latter is measured by the variance of idiosyncratic components of asset prices. Moreover, factor models can be used to learn about macroeconomic behaviour on the basis of disaggregated data (Forni et al., 2000). Finally, factor models can be successfully used for prediction (Stock and Watson, 1998).

In the above examples, the number of variables is typically large, possibly larger than observations. Vector autoregressive (VAR) models are therefore ill suited since they estimate too many parameters.

PCA is an important topic in multivariate time series analysis where it studies the covariance structure of the series. For example, the covariance structure of a vector return series plays an important role in portfolio selection for stocks and bonds. Given a k-dimensional random variable $r = (r_1, \ldots, r_k)'$ with covariance matrix Σ_r, a PCA is concerned with using a few linear combinations of r_i to explain the structure of Σ_r. If r denotes the monthly log returns of k assets, then PCA can be used to study the source of variations of these k asset returns. Here the keyword is few so that simplification can be achieved in multivariate analysis (Tsay, 2005).

Principal components will be consistent and independent only if the data in the panel is jointly normaly distributed, therefore the data set needs to be normalized before any estimation takes place (Forni et al., 2005).

It has become acceptable to transform the data to zero mean and unit variance prior to any analysis in this framework in order to keep the analysis simple. This will ensure that the first principal component describes the direction of the maximum variance. If mean

subtraction is not performed, the first principal component might instead correspond more or less to the mean of the data. A mean of zero is needed for finding a basis that minimizes the mean square error of the approximation of the data[11].

Factor analysis is related to principal component analysis (PCA), but the two are not identical. Latent variable models, including factor analysis, use regression modelling techniques to test hypotheses producing error terms, while PCA is a descriptive statistical technique

This approach has been developed to deal with large panel of time series, i.e. when the number of variables becomes large. Each variable in the data set, denoted x_{it}, is assumed to be driven by a small number of shocks that are common to each other called common components, and a variable-specific shocks ususaly called idiosyncratic components. Idiosyncratic means peculiar or individual and can be thougt of as shocks that are unique to each x_{it}. The final goal of this paper is to remove both idiosyncratic and short-run components which will make the indicator (ICI) smooth and common.

4.1 Selecting the number of factors

There is a large literature about selecting factors or principal components in factor analysis such as this one. One is the method proposed by Connor and Korajczyk (1993) where they suggest m to be the proper number of common factors in a $m \times n$ matrix:

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{m,n-1} & a_{mn} \end{bmatrix} \tag{7}$$

where the positive integers m and n are the **row** dimension and **column** dimension of $\mathbf{A}$ respectively. Then there should be no significant decrease in the cross-sectional variance of the specific error ϵ_{it} when the number of factors moves from m to $m+1$.

Another method (and the most common) proposed by Bai and Ng (2002) adopts some information criteria to select the number of factors. This method is based on the observation that the eigenvalue – eigenvector analysis of the sample covariance matrix $\widehat{\Sigma}_{\mathbf{T}}$ solves the least squares problem:

$$\min_{\alpha,\beta,\mathbf{f_t}} \frac{1}{kT} \sum_{i=1}^{k} \sum_{t=1}^{T} (r_{it} - \alpha_i - \boldsymbol{\beta_i}', \mathbf{f_t})^2 \tag{8}$$

There are assumed to be m factors so that $\mathbf{f_t}$ is m-dimensional. By letting $\hat{\sigma}_i(m)$ be

---

[11]Zero mean and unit variance signifies that the mean of each series in a given data set is subtracted and divided by the standard deviation.

the residual variance of the inner regression of the least square problem for series i and obtaining $\mathbf{f_t}$ from PCA analysis, the cross – sectional average of the residual variances is:

$$\hat{\sigma}(m) = \frac{1}{k} \sum_{i=1}^{k} \hat{\sigma}_i(m) \tag{9}$$

The two criteria of Bai and Ng (2002) are then defined as:

$$C_{p1}(m) = \hat{\sigma}^2(m) + m\hat{\sigma}^2(M) \left(\frac{k+T}{kT} \right) ln \left(\frac{kT}{k+T} \right), \tag{10}$$

$$C_{p2}(m) = \hat{\sigma}^2(m) + m\hat{\sigma}^2(M) \left(\frac{k+T}{kT} \right) ln(P_{kT}^2) \tag{11}$$

where M is a prespecified positive integer denoting the maximum number of factors and $P_{kT} = min(\sqrt{k}, \sqrt{T})$. One selects m that minimizes either $C_{p1}(m)$ or $C_{p2}(m)$ for $0 \leq m \leq M$.

Most PCA analysis assume that the number of time series is smaller than the number of time periods, that is, $k < T$. To deal with situations of a small T and large k, Connor and Korajczyk (1988) developed the concept of asymptotic principal component analysis (APCA), which is similar to the traditional PCA but relies on the asymptotic results as the number of time series k increases to infinity.

Connor and Korajczyk propose refining the estimation of $\hat{\mathbf{f_t}}$ as follows:

1. Use the sample covariance matrix Σ_T to obtain an initial estimate of $\hat{\mathbf{f_t}}$ for $t = 1, \ldots, T$.

2. For each k, perform the OLS estimation of the model $x_{it} = \alpha_i + \beta_i^T \hat{\mathbf{f_t}} + \epsilon_{it}$ for $t = 1, \ldots, T$ and compute the residual variance $\hat{\sigma}_i^2$.

3. Form the diagonal matrix $\mathbf{D}$ with residual variance and rescale the time series as $x_{it}* = x_{iy}\mathbf{D}^{-1/2}$

4. Compute Σ_T^* using x_{it}^* and perform eigenvalue – eigenvector analysis of Σ_T^* to obtain a refined estimate of $\mathbf{f}_t$ (Tsay, 2005).

An informal but useful procedure to determine the number of principal components needed in an analysis, is to examine a plot of the eigenvalues λ_i from Σ_T against each component in descending order (from largest to smallest). This is sometimes referred to as a *scree plot*. By looking for an elbow shaped plot of λ_i, one can determine the appropriate number of components. Selecting the first i principal components only provides an approximation to the total variance of the data. If a small i can provide a good approximation, then the

simplification becomes valuable. This method is used by Schneider and Spitzer (2004), and many more.

The decision regarding the number of factors to retain is important for a number of reasons. The main reason is that there is conceptual and empirical evidence that both specifying too few factors and specifying too many factors give substantial errors that affect results, although specifying too few is traditionally considered more severe. Both types of misspecifications have been empirically demonstrated to lead to poor factor-loading pattern reproduction and interpretation (Velicer et al., 2000). An extensive overview of factor extraction methods is available in Hayton, Allen, and Scarpello (2004).

Since dimensionality is important in this model, the mathematical concept of the *Hilbert space* needs to be defined[12]. This assumption is an important notion for stationarity, ergodicity and orthogonality. Hilbert space is a space where orthogonality has some meaning. Further, it is important for determining dimensionality of the factor space. It extends the methods of vector algebra and calculus from the two-dimensional Euclidean plane and three-dimensional space to spaces with any finite or infinite number of dimensions.

Generally there are as many dimensions as there are variables. The problem is that with four or more variables, there would be the need of four or more dimensions – which is quite difficult to visualize, but quite possible in mathematics of multidimensional space. This is where the Hilbert Space comes in handy and why it is needed in studies with high dimensions.

4.2 The factor model

Each time series in the data set can be thought of as the sum of two unobservable orthogonal components: the common component and the idiosyncratic component. The common component of the time series is driven by a few (fewer than idiosyncratic ones) underlying uncorrelated and unobservable common factors. If we have only one common factor affecting the data contemporaneously , then such factor can be interpreted as the reference cycle of the data (Stock and Watson, 1989).

The panel consists of i time series that are zero mean and unit variance and follow a stationary process $\{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z} \quad \text{and} \sim N(0,1)\}$[13]. All processes are stationary and which holds for any of the n-dimensional vector processes $\{x_{it} = (x_{1t}, x_{2t} \ldots x_{nt})'; n \in \mathbb{N}, t \in \mathbb{Z}\}$.

Following Greene (2008), a stochastic process x_t is stationary if it satisfies the following requirements:

1. $E[x_t]$ is independent of t.

---

[12]See Appendix 3.

[13]See full treatment of the data in Appendix 2.

2. $Var[x_t]$ is a finite, positive constant, independent of t.

3. $Cov[x_t, x_s]$ is a finite function of $|t - s|$, but not of t or s.

Each series x_{it} in the data set is the sum of the common component χ_{it} and the idiosyncratic component ξ_{it}. The model is then:

$$x_{it} = \chi_{it} + \xi_{it} \tag{12}$$

where χ_{it} and ξ_{it} are vectors, or more precisely $\boldsymbol{\chi_{it}} = (\chi_{1t}, \chi_{2t}, \ldots, \chi_{it})'$ and $\boldsymbol{\xi_{it}} = (\xi_{1t}, \xi_{2t}, \ldots, \xi_{it})'$. The common component is driven by a q-dimensional vector of common factors $\boldsymbol{f_{qt}} = (f_{1t}, f_{2t}, \ldots f_{qt})$, which however are loaded with possible different coefficients and lags:

$$\chi_{it} = b_{i1}(L)f_{1t} + b_{i2}(L)f_{2t} + \ldots + b_{iq}(L)f_{qt} \tag{13}$$

where $\boldsymbol{f_t}$ follows a Vector Auto Regressive (VAR) scheme of the form:

$$\mathbf{A}(L)\boldsymbol{f_t} = u_t \tag{14}$$

and $\boldsymbol{u_t} = \{u_{1t}, u_{2t}, \ldots, u_{qt}, t \in \mathbb{Z}\}$ is vector of *common shocks* which are white noise processes orthogonal to $\{\xi_{it}, i = 1, \ldots, n, t \in \mathbb{Z}\}$ and $b_i(L) = b_{i1}(L), b_{i2}(L), \ldots, b_{iq}(L)$ is a vector of lag polynomials. The common component is then driven by a small number of common shocks and our model then becomes:

$$x_{it} = b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \ldots + b_{iq}(L)u_{qt} + \xi_{it} = \sum_{i=1}^{q} b_{it}(L)u_{it} + \xi_{it} \tag{15}$$

It is assumed throughout that common- and idiosyncratic components are orthogonal at all leads and lags. Moreover, the individual idiosyncratic components ξ_{it} and ξ_{jt} are mutually orthogonal at all leads and lags for $i \neq j$. These assumptions are crucial for the model since they allow for a limited amount of cross-correlation between idiosyncratic components and ensures that the variances explained by the idiosyncratic component vanishes as $N \to \infty$. It further assumes a minimum correlation between the common components.

4.3 Estimation and construction of the regressors

The name *regressors* are derived from the fact that they will be regressed or projected on c_t, which was the medium-to-long-run-growth of monthly GDP as defined in chapter 3.3. They are the first r principal components or factors which are in fact the common component χ_{it}.

The factor model exploits the covariance structure of x_{it} where the estimated factors can be derived by applying an eigen decomposition of the covariance matrix of x_{it}. Let Σ_x be the sample covariance matrix of x_{it}, which has been scaled to zero mean and unit variance as introduced before. Then the $n \times n$ sample covariance matrix Σ_x can be defined as

$$\sum_x = \frac{1}{N-1}\mathbf{X}\mathbf{X}^T, \quad \mathbf{X} = \{x_{it} = (x_{1t}, x_{2t} \ldots x_{nt})'\} \tag{16}$$

Next up is finding and extracting the eigenvalues and eigenvectiors of Σ_x by employing the eigen decomposition which is a useful mathematical concept in analyzing square matrices like Σ_x. Then the eigenvalues are collected and arranged in decreasing order of magnitude and the matching eigenvectors in the same order. By letting $\mathbf{\Lambda}$[14] be the diagonal matrix with eigenvalues $\lambda_1, \ldots \lambda_N$ of Σ_x on the diagonal, arranged by size so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$, the following matrix is formed:

$$\mathbf{\Lambda} = [\lambda_N] = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ & & & \cdots & \\ 0 & 0 & 0 & \cdots & \lambda_N \end{bmatrix} \tag{17}$$

The $N \times N$ symmetric matrix Σ_x has N distinct characteristic vectors or eigenvectors, $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N$ which corresponds to the characteristic roots or eigenvalues $\lambda_1, \ldots \lambda_N$. Eigenvectors of a symmetric matrix are orthogonal to each other which implies that for every $i \neq j$, $\mathbf{c}_i\mathbf{c}_j = 0$ and $\mathbf{C}\mathbf{C}^T = \mathbf{I}$, so the rows as well as columns of $\mathbf{C}$ are orthogonal. $\mathbf{C}$ is then

$$\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N] \tag{18}$$

These unit eigenvectors $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N$ of Σ_x are called the **principal components** of the data in $\mathbf{X}$. The first principal component is the eigenvector corresponding to the largest eigenvalue of Σ_x, the second principal component is the eigenvector corresponding to the second largest eigenvalue, and so on. The first principal component $\mathbf{c}_1$ determines the new variable w_1 in the following way. Let $d_1, \ldots, d_n$ be the entries in $\mathbf{c}_1$. Since $\mathbf{c}_1^T$ is the first row of $\mathbf{C}^T$, the equation $\mathbf{W} = \mathbf{C}^T\mathbf{X}$ reveils that:

$$w_1 = \mathbf{c}_1^T\mathbf{X} = d_1x_1 + d_2x_2 + \ldots + d_nx_n \tag{19}$$

Thus, w_1 is a linear combination of the original variables in $\mathbf{X}$, using the the entries in

---

[14]Λ is the greek capital letter of λ.

the eigenvector $\mathbf{c}_1$ as weights. In a similar fashion, $\mathbf{c}_2$ determines the variable w_2, and so on.

Principal component analysis is potentially valuable for aplications in which most of the variation in the data is due to variations in only a few of the new components $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N$. It can be shown that an orthogonal change of the variables, $\mathbf{W} = \mathbf{CX}$, does not change the total variance of the data which is a very interesting property (Lay, 2006). This means that when $\Sigma_x = \mathbf{CDC}^T$, then:

$$\left\{ \begin{matrix} \text{Total variance} \\ \text{of} \quad x_1, \ldots, x_n \end{matrix} \right\} = \left\{ \begin{matrix} \text{Total variance} \\ \text{of} \quad w_1, \ldots, w_n \end{matrix} \right\} = \text{tr}(\mathbf{\Lambda}) = \sum_{n=1}^{N} \lambda_n = \lambda_1 +, \ldots, +\lambda_n \quad (20)$$

where $\text{tr}(\mathbf{\Lambda})$ is the cumulative sum of the eigenvalues which are located on the diagonal on $\mathbf{\Lambda}$ as shown in equation 17. The variance of w_1 is exactly the same as the first eigenvalue λ_1, and the quotient $\lambda_1/\text{tr}(\mathbf{\Sigma}_x)$ measures the fraction of the total variance that is explained or captured by w_1, and so on. In this case the total variance is naturally the same as the number of variables because of unit variance.

Most often the principal components with the highest variance are selected but the low variance principal components may also be important, in some cases even more important. The low variance components are not of interest here.

Now since all the principal components have been estimated, the selection of the first r components are only of interest since they will contain most of the variance of $\mathbf{X}$. The regressors, denoted by w_{kt}, $k = 1, \ldots, r$, are then contemporaneous linear combinations of such variables and are specifically constructed to minimize the short run component of all the variables. Then, c_t is projecteded on the linear space spanned by the first r principal components. The projection[15], denoted by $\mathcal{K}_r$, can not be done until the regressors $w_t = w_{1t}, w_{2t}, \ldots, w_{rt}$ are filtered and smoothed before any estimation of the final indicator takes place. The regressors are simply filtered the same way according to the filter in equation 4.

The construction of the filter implies cutting of observations at the beginning and end of w_{rt}. This can be solved by simply augmenting the missing observations of each w_r with its mean so the length will be proportional to c_t. Then, a moving average smoother is applied to w_{rt}. This projection of c_t on the first r smoothed factors is the Icelandic Coincident Indicator, or ICI.

That will be the discussion in the next chapter. Further explanation of the model will be done (if needed) along with the results.

---

[15]It is important to note that $\mathcal{K}_r$ is obtained by regressing the the first r components on c_t.

4.4 Forecasting

Most multivariate forecasting methods in the literature are restricted to series of low dimension, and allow for incorporating only a limited number of key variables (or methods based on low dimension models such as VARs). Such methods are thus of little help in large panels of time series, where the cross-sectional dimension is often of the same order as, or even larger than the series lengths (Forni et al., 2005).

This shows the strength of PCA and factor models. In section 5.2, a basic ARMA model is used to predict ICI and c_t where the root mean square errors (RMSE's) are calculated to measure the goodness of fit with respect to GDP growth. RMSE will be calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2} \tag{21}$$

where n is the number of periods being forecasted, y_i is the benchmark, in this case GDP, and $\hat{y}_i$ is the estimated prediction of ICI or c_t. This measure is ofcourse backward looking in that they are computed using the observed data. Several other measures are available such as Mean Absoulte Error (MAE) and Theil U Statistics.

5 Results

Before any results can be estimated, the number of common factors r need to be determined. Several strategies are available in determing the number of factors to retain, as is listed in chapter 4.1. Commonly used method in principal component analysis is the examination of the plot of eigenvalues in descending order. Such a plot is usually elbow-shaped and the optimal number of common components to retain is either determined by the last point before the kink (bottom of elbow) or simply a suitable number of components that explain enough variance of the original data.
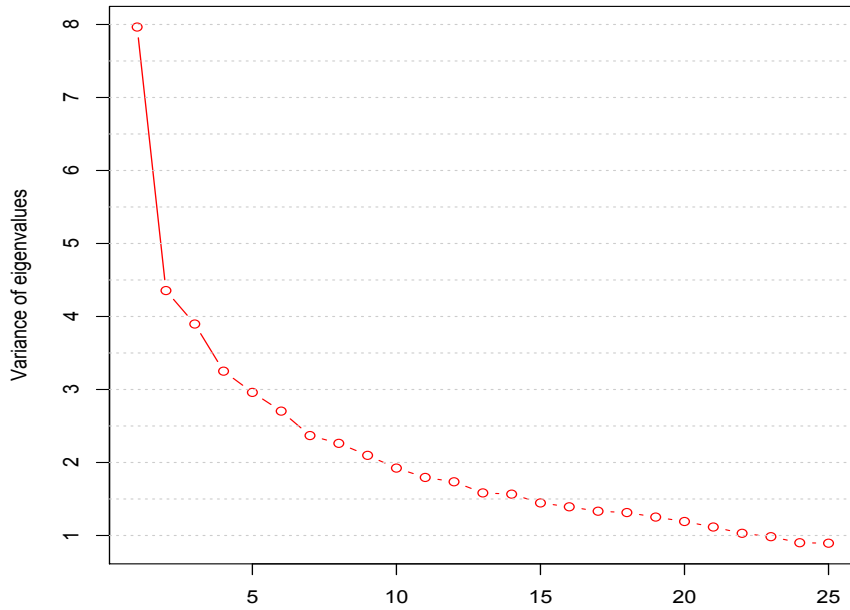

Figure 7: Plot of eigenvalues in descending order. Only the first 25 values are shown.

This approach is informal and is based on the researcher's decision. It has the disadvantage that there is no formal basis or underlying foundation that selects the proper number of common components to retain. Another standard tradition is to keep only the components whose eigenvalue is larger than the average eigenvalue. Formally, this amounts to keeping the rth component if:

$$\lambda_r > \frac{1}{N} \sum_r^N \lambda_r \tag{22}$$

By looking at the screeplot in figure 7 there is no obvious break in the plot which makes it hard to determine the "right" number of r. There is also a slow decrease in variance explained by the eigenvalues. All factor tests mentioned in chapter 4.1 (except APCA) give mixed criteria ranging from $r = 5$ to $r = 25$, the result being $r = 11$. Listed in table 4 are the first eleven components with their cumulative percentage of variance

along with the contribution each r gives. The goal is then to describe c_t by projecting c_t on the first eleven smoothed regressors w_r, which then will be the desired indicator ICI.

Table 4: Total variance explained by the first r components. Also shown is the cumulative sum of the explained variance and contribution to explained variance. Source of layout: (Schneider and Spitzer, 2004) and own work.

r	Cumulative %	Contribution to explained variance (%)
1	18.39%	18.39%
2	24.20%	5.81%
3	29.06%	4.85%
4	33.47%	4.42%
5	37.51%	4.04%
6	41.05%	3.54%
7	44.42%	3.38%
8	47.55%	3.13%
9	50.42%	2.87%
10	53.10%	2.68%
11	55.70%	2.60%

Finally, except for the case in which $\lambda_n = 0$ for $n > r$, selecting the first r principal components only provides an approximation to the total variance of the data. If a small r can provide a good approximation, then the simplification becomes valuable. In this case 11 out of 67 (16%) has been chosen.
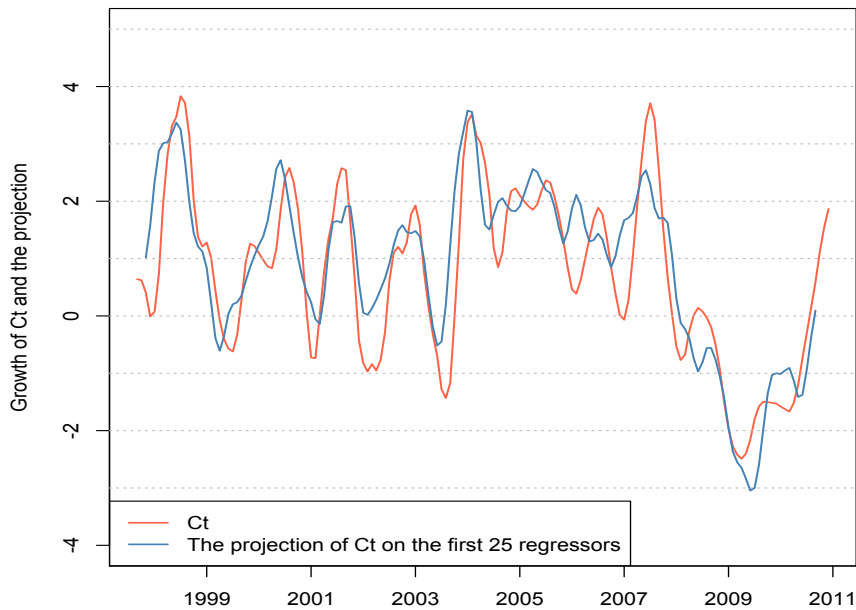

Figure 8: Projection of c_t on the first 25 components. $R^2 = 0.72$ and $F = 14.4$.

Now turning to the estimation of the indicator *ICI* where $\mathcal{K}_r$ is the projection of c_t on the first r components and looking at the most extreme case, i.e. where $r = 25$.

Figure 8 shows the projection of the first 25 components together with c_t, denoted $\mathcal{K}_{25}$. The results indicate that $\mathcal{K}_{25}$ is a fairly good approximation to c_t with R^2 as high as 0.72 and $F = 14.4$. However, as figure 8 shows, $\mathcal{K}_{25}$ contains a sizable short run component. The reason behind this is the huge number of principal components.

Reducing the dimension of the components to the previous chosen $r = 11$ and regressing them on c_t, a smoother version of $\mathcal{K}_r$ is obtained, namely $\mathcal{K}_{11}$. This is the desired indicator ICI which is sought, and shown in figure 9 along with c_t[16]
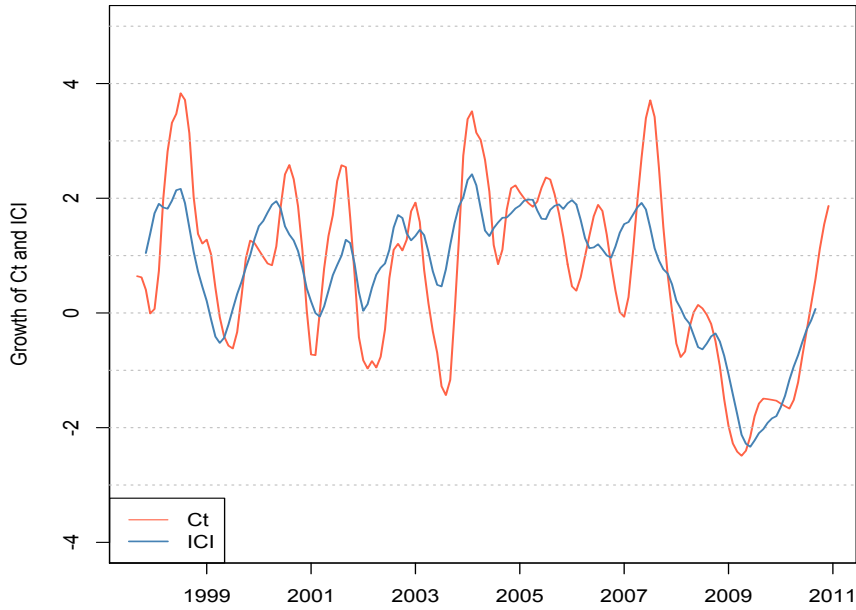

Figure 9: ICI compared to c_t. $R^2 = 0.63$ and $F = 22.42$.

Firstly, the R^2 of the regression falls to 0.63 but the fit as a whole improves since $F = 22.42$ which can be best seen by the increase in smoothness, and secondly, ICI seems to follow the ideal target c_t quite nicely which makes ICI relatively smooth and less volatile than c_t.

It is also important to note that components 12 to 25, which were regressed as in the extreme case in figure 8, are only explaining 9% more with even worse fit. Overfitting, that is, trying to estimate a model that is too large with too many variables will give a huge bias which is a common error. That is precisely the strength of PCA and factor analysis. In stead of regressing all the variables directly, one can use PCA and factor analysis with considerably less variables but still retain the variability of the underlying data generating process.

Even further, the moving average smoothed factors fit rather nicely. By the use of moving averages one hopes to decrease the variation so that long term properties will be more clearly visible. There is however a possible side effect when analyzing cyclical behaviour.

---

[16]The plots of $\mathcal{K}_r$ and c_t with fewer regressors are shown in Appendix 3.

A certain moving average formula might dampen one cycle and amplify another. This is a well known phenomena in time series analysis and is referred to as the Slutsky effect (Tómasson, 1989).

The regression results are shown in table 5.

Table 5: Summary of regression of the factors on c_t. As usual, * for 0.1%, ** for 1%, and * for 5% are significance levels.**

	Estimate	Std. Error	t value	p-value
(Intercept)	0.71	0.076	9.37	0.00***
w_1	0.51	0.13	3.93	0.00***
w_2	3.23	0.59	5.51	0.00***
w_3	-1.90	0.29	-6.52	0.00***
w_4	0.30	0.54	0.55	0.58
w_5	0.60	0.34	1.78	0.08*
w_6	2.28	0.42	5.44	0.00***
w_7	0.16	0.24	0.67	0.50
w_8	0.40	0.20	2.04	0.04**
w_9	-0.61	0.43	-1.44	0.15
w_{10}	0.07	0.27	0.24	0.80
w_{11}	-0.39	0.27	-1.43	0.16

To begin with, the first three regressors are higly significant along with regressors 5, 6, and 8. There are a couple of regressors highly insignificant and removing them produces worse fit and more volatile ICI. Therefore, the construction of ICI is now complete and is plotted along with the original interpolated monthly GDP and our ideal smooth target, c_t in figure 10.

ICI seems to follow the underlying data generating process quite nicely and in some cases, precedes the original GDP which could be helpful in determining which state the the economy as a whole is heading.

Table 6: A simple descriptive statistics of GDP, c_t, and ICI.

Indicator	Standard deviation	Mean
GDP	2.43%	0.74%
c_t	1.49%	0.75%
ICI	1.19%	0.74%

Table 6 lists up a simple but very informative descriptive statistics of the indicators shown in figure 10. An interesting relationship then reveals. The volatility is reducing as we go from the original GDP to the smooth ICI, but the mean or average growth still holds
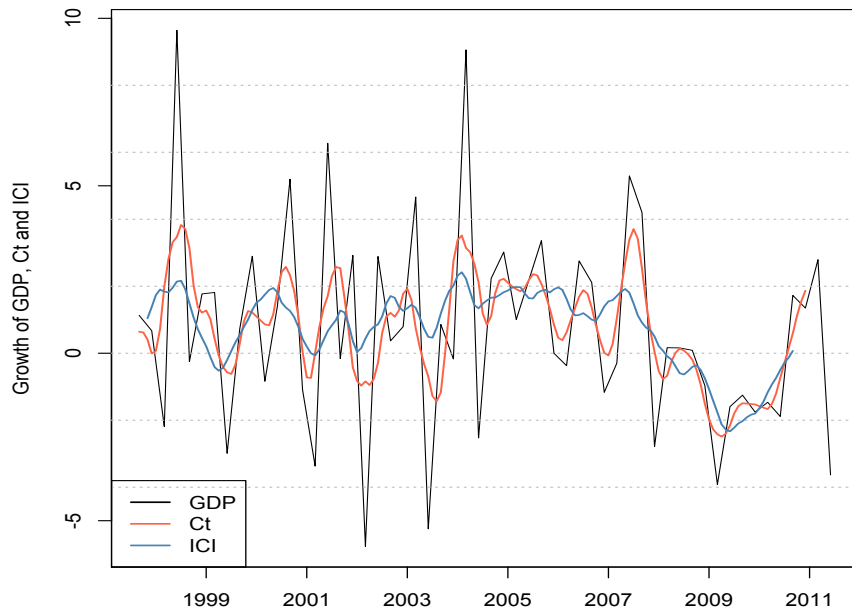
Figure 10: A plot of GDP, c_t, and ICI.

which is an important characteristic of ICI. The goal of constructing an indicator free from short run dynamics but still tracks and describes GDP growth was the main goal of this paper.

5.1 Behavior around turning points

To analyze whether or not ICI is performing well around the turning points, that is, the behavior of signals and the percentage of correct signs, there needs to be a precise defintions of turning points and signals. To begin with, the definition of a *turning point* is the slope change of c_t. There is an upturn (downturn) at time t, if $\Delta c_{t+1} = c_{t+1} + c_t$ is positive (negative). With this definition, 14 upturns and 13 downturns of c_t have been identified for the whole period. ICI is simply calculated in the same way and compared as shown in table 7.

Table 7: Performance of ICI around the true turning points.

Indicator	Upturns	Downturns	% of correct upturn direction	% of correct downturn direction
c_t	14	13	-	-
ICI	12	10	86%	83%

5.2 Forecasting properties of ICI and c_t

This section compares the forecasting properties of c_t and ICI with respect to GDP growth. That is, the idea is to calculate the predictions of c_t and ICI from a simple ARMA model, and compare it to GDP growth. In chapter 3.3, it was mentioned that year-on-year (y-o-y) growth of GDP is sometimes referred to the measure of medium-to-long-run growth of GDP which is just a simple moving average of GDP.

To better see the forecasting abilities of ICI, a comparison is done with a univariate ARMA models of GDP growth. Such models are often used as benchmark in forecasting studies. Root mean square errors (RMSE) of year-on-year (y-o-y), quarter-on-quarter (q-o-q) and month-on-month (m-o-m) growth rates og GDP are reported in table 8.

Table 8: RMSE of ICI and c_t with respect to different growth rates of GDP.

	y-o-y growth	q-o-q growth	m-o-m growth
ICI	1.20	0.91	1.35
c_t	1.90	0.95	1.28

Q-o-q growth rate seems to be the best forecasting method of both ICI and c_t and shown in figure 11 is the original GDP growth along with c_t and ICI.
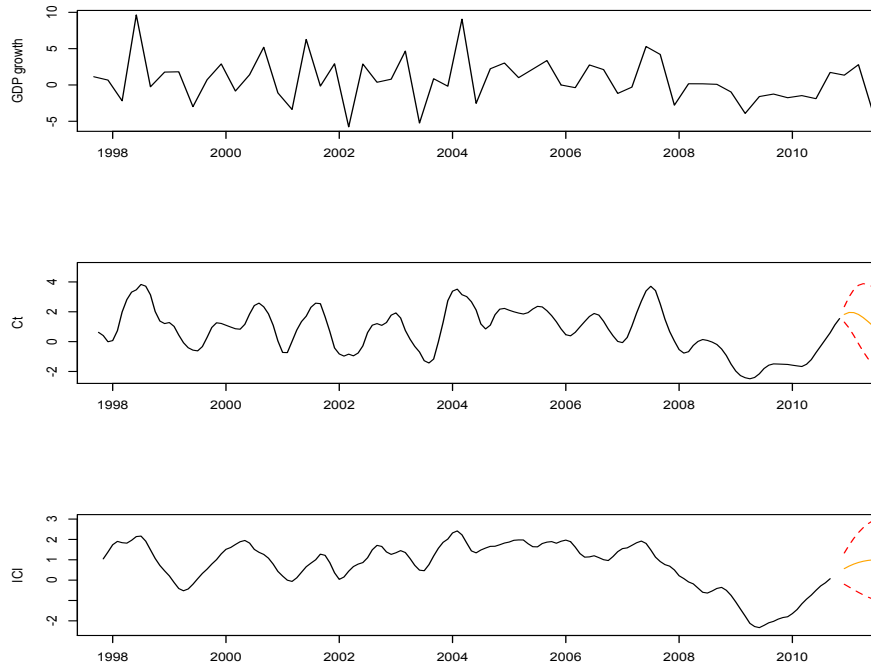

Figure 11: Forecasts of ICI and c_t. Orange lines indicate the actual forecasts and the red dotted lines indicate the standard errors (SE) of forecasts. Standard errors are calculated as: forecast $\pm$ 2 SE

ICI performs slightly better with a RMSE of 0.91 and the reason is that c_t falls rather rapidly and has a big error as shown by the red dotted lines.

6 Conclusion

The aim of this paper was to construct an indicator from a reasonably large number of macroeconomic time series which could describe the economic growth as a whole. This indicator was supposed to be free from short-run dynamics and therefore smooth, and most but not least a reliable indicator which could describe the Icelandic economy. The most commonly known indicator known today in describing the same process is the GDP growth. But as argued before, it is only available on a quarterly basis with heavy short-run component and under constant revision, making it rather unreliable.

From this discussion and the framework of factor models and principal component analysis, the **ICI** an Icelandic Coincident Indicator, was constructed which is an indicator free from short run dynamics, smooth and describes the underlying data generating process, in this case GDP growth.

ICI follows the ideal target, c_t, quite nicely i.e. it captures 63% of the variance of c_t and is relatively easy to interpret. In terms of detection of turning points, it scores quite high, or on average 84.5%. In some cases it precedes the original GDP which could be helpful in determining which state the economy is currently heading.

The downfall of PCA is that the method is very sensitive to the original scaling of the data set and the choice of variables. A scaling of zero mean and unit variance is usually required for the data of interest, especially if the data are not homogenous in nature e.g. interest rates, imports and car registrations. These time series and many more are very different in nature. The scaling of data then could possibly reduce the variability needed. Unfortunately, the slow decrease of the explained variance of the eigenvalues is propably the most serious downfall of this research, that is, the first 11 eigenvalues only explain 56% of the data comparing to 76% of c_t. Also, the short period of available data points and frequency in Iceland, limits the use of factor models and PCA even further since the use of both spectral methods and seasonal adjustments will heavily revise the data set which essentially affects the forecasting properties.

But still, the reduction of dimensions of variables and the projection of c_t on the 11 regressors still provides some useful interpretation. ICI is less volatile but still retains the average growth and follows the c_t's path. Overfitting with too many regressors resulted in worse fit.

The forecasting properties of ICI and c_t was done with the simple ARMA model and compared to the GDP growth. RMSE of ICI compared to q-o-q GDP growth was 0.91 which is relatively high. The reason is the serious turbulance of all time series during the periods of 2008 and 2009. The majority of variables showed serious outliers when the data was analyzed which made some of them too noisy to be included in this research. Even after adjusting for outliers by replacing them with mean or by interpolation, they

still experienced a lot of noise.

The performance of ICI as a real-time estimator of the Icelandic economy has now been presented in details and is smooth and easy to interpret.

7 Discussion and future research

An important empirical question which has not been answered thoroughly, concerns the size and the composition of the optimal data set. Although factor models have been developed to deal with large panels with hundreds of variables, there are some results which indicate that increasing the number of variables over a certain size do not improve or even worsen forecasting results (Boivin and Ng, 2006) and (Watson, 2003).

One direction of future research is the similarity (or unsimilarity) of data, which may be worth further investigation. As shown in Boivin and Ng (2006), preselecting variables for factor estimation may improve the fit of the model, because the data often does not represent a homogenous structure. Following this discussion, data selection in factor models may be, will be and should be an important topic for future research. For example, Schneider and Spitzer (2004), Forni et al. (2005) and Altissimo et al. (2010) all reduce their data sets to get better estimates.

The dataset has been organized in taking into account the calendar of data releases with the aim of reproducing the flow of information available through time and was accomplished by introducing the concept of *data delay*. This made the data more synchronized which means that the data is describing the same underlying process and is crucial in developing a real indicator of current economic state. The concept of synchronizing the data is relatively new and has been neglected in the literature and should be properly researched.

The lack of survey data (or shortness of frequency there of) are another downfall since they are only readily available from 2004. Some survey data are available from 1999 and 2000, but still are too short to be included in this research. Survey data are important because they are quantitative and can succesfully describe the true customers and how they really think. This could be valuable in a research like this. There is a way of backdating survey data as well as other data with Monte Carlo (MC) simulations. The MC procedures could be exploited in the next phase of this research.

Very recent research in this area of interest, is the **dynamic** factor models and bayesian estimation. The dynamic factor models use a spectral density estimation to estimate dynamic eigenvalues and eigenvectors of different frequencies. In these frameworks, data are allowed to be of different time intervals and varying in lenght. Bayesian estimation are more concerned about real business cycle estimation through various shocks.

The need for methods to deal with large data sets is huge. Data is getting bigger, more reliable and with greater frequency. Business cycle analysis through factor models is potentially valuable and should be investigated more thoroughly and the value in producing accurate forecasts of key macroeconomic variables is of great importance for every country.

Appendix 1

The low-pass filter of Baxter & King (1999)

A filter that retains only the slow-moving components of the data is a symmetric low-pass filter, which passes only frequencies $-\hat{\omega} \leq 0 \leq \hat{\omega}$ and has a frequency response function given by $\beta(\omega) = 1$ for $|\omega| \leq \hat{\omega}$ and $\beta(\omega) = 0$ for $|\omega| > \hat{\omega}$. The symmetry of the filter implies $\beta(\omega) = \beta(-\omega)$.

If we let the time-domain representation of this ideal low-pass filter to be $b(L) = \sum_{h=-\infty}^{\infty} b_i L^h$, the filter weights may be found by the Fourier transform of the frequency response function:

$$b_h = \frac{1}{2\pi} \int_{-\pi}^{\pi} \beta(\omega) e^{i\omega h} \, \mathrm{d}\omega$$

Evaluating this integral gives the weights of our ideal filter. The Fourier integral of the ideal low-pass filter $\beta(\omega)$ implies that the filter coefficicents satisfy:

$$b_h = \frac{1}{2\pi} \int_{-\pi}^{\pi} \beta(\omega) e^{i\omega h} \, \mathrm{d}\omega = \frac{1}{2\pi} \int_{-\hat{\omega}}^{\hat{\omega}} e^{i\omega h} \, \mathrm{d}\omega$$

where $\beta(\omega) = 1$ for $|\omega| \leq \hat{\omega}$ and $\beta(\omega) = 0$ for $|\omega| > \hat{\omega}$. Hence, it follows that

$$b_0 = \frac{1}{2\pi} \int_{-\hat{\omega}}^{\hat{\omega}} \mathrm{d}\omega = \frac{\hat{\omega}}{\pi}$$

and

$$b_h = \frac{1}{2\pi} \int_{-\pi}^{\pi} \beta(\omega) e^{i\omega h} = \frac{1}{2\pi} \left[\frac{1}{ih} e^{i\omega h} \right] = \frac{1}{\pi h} \sin(\hat{\omega} h)$$

The ideal weights for the filter are $b_0 = \frac{\hat{\omega}}{h}$ og $b_h = \frac{\sin(h\hat{\omega})}{\pi h})$ for $h = 1, 2, \ldots$

The weights converge to zero when h grows larger which means that we need an infinite order moving average to construct this filter. Hence, the ideal filter needs a finite moving average:

$$b(L) = \sum_{h=-k}^{k} b_h L^h$$

The lag and difference operators

A covenient way of manipulating lagged variables is the **lag operator**,

$$Lx_t = x_{t-1}.$$

A related operation is the first difference,

$$\Delta x_t = x_t - x_{t-1}.$$

Then:

$$\Delta x_t = (1 - L)x_t$$

A regression model can be written as:

$$y_t = \alpha + \sum_{i=0}^{\infty} \beta_i L^i x_t + \epsilon_t = \alpha + B(L)x_t + \epsilon_t$$

where $B(L)$ is a polynomial in L, $B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \dots,$. A polynomial in the lag operator that reappears in many context is

$$A(L) = 1 + \alpha L + (\alpha L)^2 + (\alpha L)^3 + \dots = \sum_{i=0}^{\infty} (\alpha L)^i$$

The Hilbert space

The data in the dataset is assumed to be stochastic and belong to the Hilbert Space $L_2(\Omega, \mathcal{F}, P)$ where $(\Omega, \mathcal{F}, P)$ is a given probability space where the double sequence $\{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ is under study. $\mathcal{F}$ is a σ-algebra of subsets of Ω.

Appendix 2

Data description

Class	Data description	Unit	Type of treatment	Data delay	Source
Indexes	Salary Index	Index	d log	T	STAT & DM
Indexes	Price Index	Index	d d log	$T-2$	STAT & DM
Indexes	Index of pensionfund debts	Index	d d log	$T-2$	STAT & DM
National accounts	Government revenue	Million ISK	d	$T-3$	FR & DM
National accounts	Government costs	Million ISK	d	$T-3$	FR & DM
Labour	Unemployment	%	d log	$T-2$	STAT & DOL
Import/Export	Cement sales	Index	d log	$T-2$	STAT & DM
Import/Export	Import FOB	Million ISK	d log	$T-2$	STAT & DM
Import/Export	Import CIF	Million ISK	d log	$T-2$	STAT & DM
Import/Export	Export FOB	Million ISK	d log	$T-2$	STAT & DM
Import/Export	Import - cars CIF	Million ISK	d log	$T-2$	STAT & DM
Import/Export	Import - Operation goods	Million ISK	d log	$T-2$	STAT & DM
Import/Export	Import - Durables	Million ISK	d log	$T-2$	STAT & DM
Import/Export	Import - Gasoline	Ton	None	$T-2$	STAT & DM
Import/Export	Import - Fuel	Ton	None	$T-2$	STAT & DM
Import/Export	Import - Gas	Ton	None	$T-2$	STAT & DM
Import/Export	Import - Jet fuel	Ton	None	$T-2$	STAT & DM
Demand indicators	Registered cars, used and new	Number	d	$T-2$	STAT & DM
Demand indicators	Hotelvisits, foreigners	Number	d	$T-2$	STAT & DM
Demand indicators	Arrivals to Keflavik airport	Number	d	$T-2$	STAT & DM
Demand indicators	Poultry meat sales	Ton	d	$T-2$	STAT & DM
Demand indicators	Sheep meat sales	Ton	d	$T-2$	STAT & DM
Demand indicators	Beef meat sales	Ton	d	$T-2$	STAT & DM
Demand indicators	Pork meat sales	Ton	d	$T-2$	STAT & DM
Demand indicators	Milk production	Thousands of liters	d	$T-2$	STAT & DM
Industry	Electricity consumption, industry	Gigawatt hours	d log	$T-1$	STAT & DM
Industry	Electricity consumption, general	Gigawatt hours	d log	$T-1$	STAT & DM
Industry	Aluminium price, USD/ton	$	d log	T	STAT & DM
Industry	Oil price (UK Brent 38), USD/unit	$	d log	T	STAT & DM
Financial indicators	Exchange rate index of ISK	Index	d log	T	CBI & DM
Financial indicators	Real exchange rate index	Index	d log	T	CBI
Financial indicators	Exchange rate index of trade, wide	Index	d log	T	CBI & DM
Financial indicators	Exchange rate index of trade, narrow	Index	d log	T	CBI & DM
Financial indicators	Broad merchandise index	Index	d log	T	CBI & DM
Financial indicators	Narrow merchandise index	Index	d log	T	CBI & DM
Financial indicators	Special drawing rights	Index	d log	T	CBI & DM
Financial indicators	Broad money (M3)	ISK	d log	$T-1$	CBI
Financial indicators	Money and sight deposits (M2)	ISK	d log	$T-1$	CBI
Financial indicators	Money supply (M1)	ISK	d log	$T-1$	CBI
Financial indicators	Base money (M0)	ISK	d log	$T-1$	CBI
Funds	Total assets of pension funds	ISK	d	$T-1$	CBI
Funds	Total assets of private funds	ISK	d	$T-1$	CBI
Funds	Total equities of private funds	ISK	d	$T-1$	CBI
Funds	Unindexed treasury bonds	ISK	d	$T-1$	CBI
Funds	Indexed treasury bonds	ISK	d	$T-1$	CBI
Funds	Municipal bonds	ISK	d	$T-1$	CBI
Funds	Bonds of deposit money banks	ISK	d	$T-1$	CBI
Funds	Bonds of other financial institutions	ISK	d	$T-1$	CBI
Funds	Corporate bonds	ISK	d	$T-1$	CBI

Data description (continued)

Class	Data description	Unit	Type of treatment	Data delay	Source
Interest rates	Penalty rates	%	d d	T	CBI & DM
Interest rates	Monetary rates	%	d	T	CBI
Estate indexes	Apartments index	Index	d d log	$T-2$	STAT & DM
Estate indexes	Detached apartments index	Index	d d log	$T-2$	STAT & DM
Estate indexes	Apartment price index	Index	d d log	$T-2$	STAT & DM
HICP Prices. Base=2005	Acommodation services	%	d	$T-2$	EURO
HICP Prices. Base=2005	Actual rentals for housing	%	d	$T-2$	EURO
HICP Prices. Base=2005	Beer	%	d	$T-2$	EURO
HICP Prices. Base=2005	Financial services n.e.c.	%	d	$T-2$	EURO
HICP Prices. Base=2005	Fish and seafood	%	d	$T-2$	EURO
HICP Prices. Base=2005	Household goods, non-durable	%	d	$T-2$	EURO
Consumer price index by categories	Agriculture products, less vegetables	Index	d log	$T-1$	STAT & DM
Consumer price index by categories	Vegetables(domestic and imported)	Index	d log	$T-1$	STAT & DM
Consumer price index by categories	Imported food and beverages	Index	d log	$T-1$	STAT & DM
Consumer price index by categories	Cars and spare parts	Index	d log	$T-1$	STAT & DM
Consumer price index by categories	Alcohol and tobacco	Index	d log	$T-1$	STAT & DM

List of data sources [17]

Acronym	Source	Ratio of data
STAT	Statistic Iceland	52.2%
CBI	Central Bank of Iceland	34.3%
EURO	Eurostat	9.0%
FR	Fjársýsla ríkisins	3.0%
DOL	Directorate of Labour	1.5%

Out of all the data, 67% of them was fetched through **DataMarket (DM)**.

---

[17]Where ever DataMarket is named as source, the data is fetched with a code written in **R**. DataMarket recently released a package specifically written for **R** to fetch data through their site to be imported and viewed. This thesis exploits this code extensively. See Appendix 3 for further information on how to fetch data through **R** via DataMarket.

Appendix 3

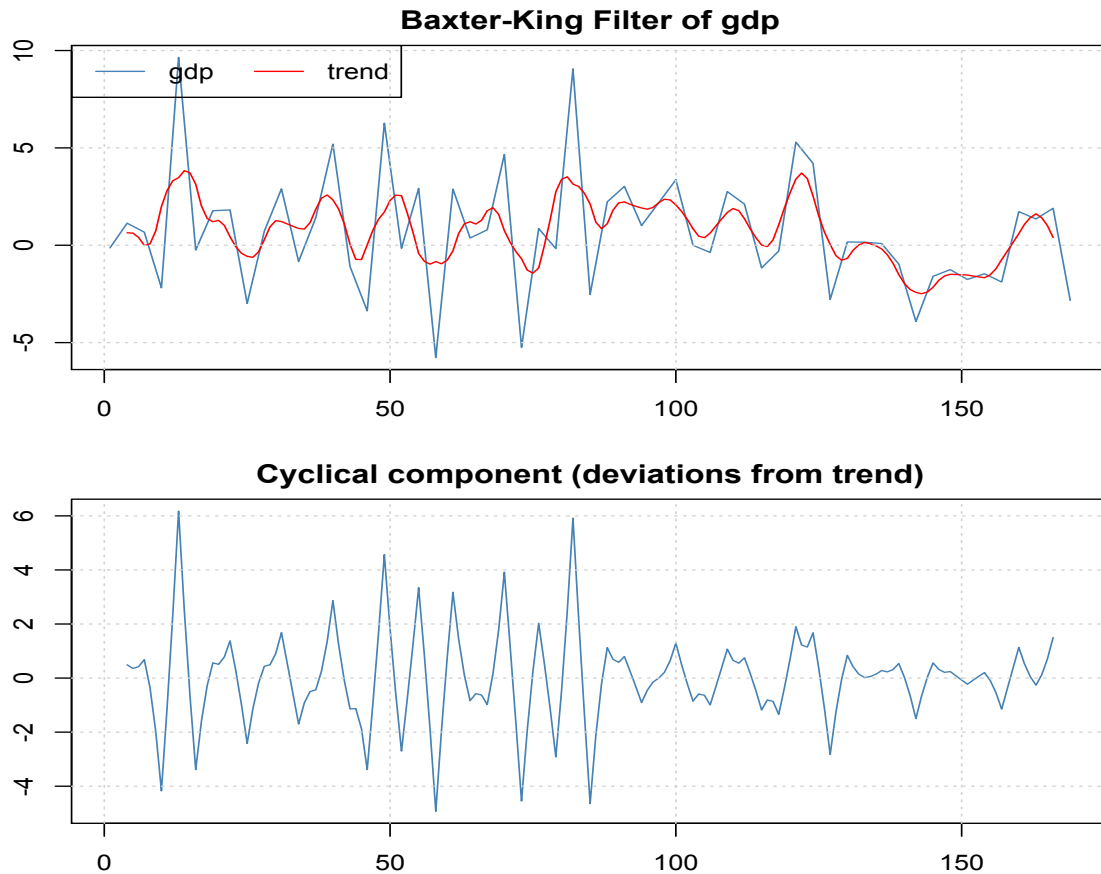## Output from Baxter-King filtration in R

Figure 12: Output from the filter.

R code for fetching data through DataMarket

An example on how to fetch consumer price index trough DataMarket by simply pasting the url to the code.

```
    library(rdatamarket)
cpiA <- dmlist("http://data.is/pOwjs8")
cpiB <- cpiA$Value                       # Extract the values for CPI.
cpiC <- cpiB[105:278]                     # Values for the period 1997 - 2011
cpiD <- ts(cpiC, frequency=12, start=1997)  # Function to make CPI a timeseries which starts Jan 1997
```

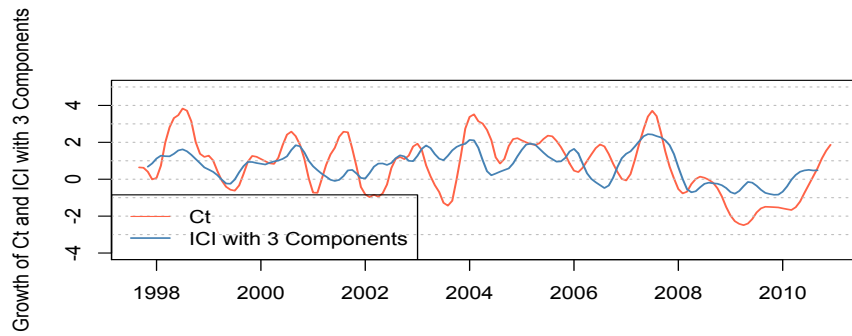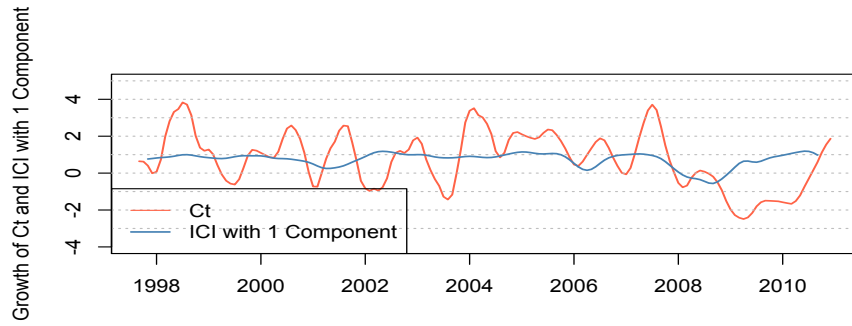# Plots of ICI with different number of components
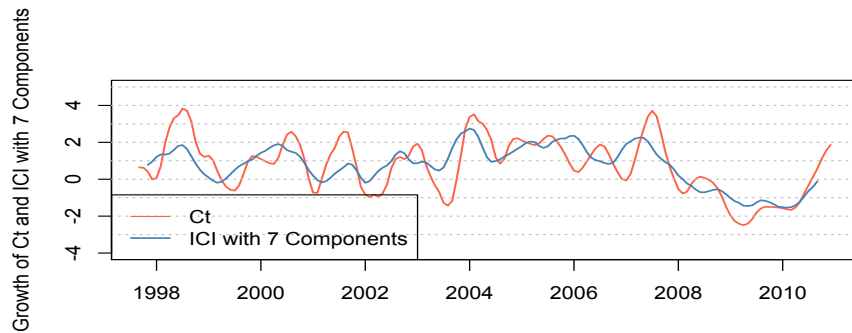

Figure 13: c_t vs ICI with 1 and 3 components.
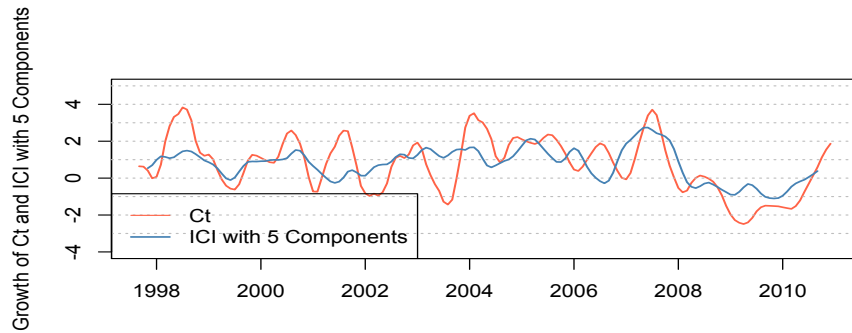

Figure 14: c_t vs ICI with 5 and 7 components.

References

Almeida, V. (2009). Bayesian estimation of a DSGE model for the Portuguese economy. *Bank of Portugal: Economic Research Department, Working Papers.*

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., and Veronese, G. (2010). New euro-coin: Tracking economic growth in real time. *The Review of Economics and Statistics*, 92, 1024–1034.

Angelini, E., Henry, J., and Marcellino, M. (2006). Interpolation and backdating with a large information set. *Journal of Economic Dynamics and Control*, 30, 2693–2724.

Artis, M., Banerjee, A., and Marcellino, M. (2005). Factor forecasts for the UK. *Journal of Forecasting*, 24, 279–298.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70, 191–221.

Bartholomew, D., Steele, F., Galbraith, J., and Moustaki, I. (2008). *Analysis of multivariate social science data (2nd edition)*. Boca Raton: CRC Press.

Baxter, M. and King, R. (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics*, 81, 575–593.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132, 169–194.

Box, G., Jenkins, G., and Reinsel, G. (1970). *Time series analysis*. San Francisco: Holden Day.

Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics*, 41, 475–512.

Caprio, G. and Klingebiel, D. (2002). Episodes of systemic and borderline banking crises. *World Bank Discussion Paper*, 428, 31–49.

Chari, V. and Kehoe, P. (1999). *Optimal fiscal and monetary policy: Handbook of macroeconomics, edited by John B. Taylor and Michael Woodford*, volume 1C.

Chow, G. and Lin, A. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53, 372–375.

Connor, G. and Korajczyk, R. (1988). Risk and return in an equilibrium: Application of a new test methodology. *Journal of Financial Economics*, 21, 255–289.

Connor, G. and Korajczyk, R. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance*, 48, 1263–1291.

Danielsson, A. (2008). Accuracy in forecasting macroeconomic variables in Iceland. *The Central Bank of Iceland, Working Papers No. 39*.

Denton, F. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, 46, 99–102.

Dungey, M. (2009). Monetary policy, inflation, and the business cycle: An introduction to the new keynesian framework. *Economic Record*, 85, 493–493.

Durkheim, E. (1952). *Suicide: A study in sociology. Translated by John A. Spaulding and George Simpson*. London: Routledge and Kegan Paul. (Originally published 1897).

Einarsson, T. and Marquis, M. (2001). Bank intermediation over the business cycle. *Journal of Money, Credit and Banking*, 33, 876–899.

Eklund, B. (2007a). Forecasting the Icelandic business cycle using vector autoregressive models. *Central bank of Iceland, Working Papers No. 36*.

Eklund, B. (2007b). Predicting recessions with leading indicators: An application on the Icelandic economy. *Central bank of Iceland, Working Papers No. 33*.

Engle, R. and Granger, C. (1987). Cointegration and error correction: representation, estimation, and testing. *Econometrica*, 55, 251–276.

Evans, C. (1992). Productivity shocks and real business cycles. *Journal of Monetary Economics*, 29, 191–208.

Fernandez, R. (1981). A methodological note on the estimation of time series. *The Review of Economics and Statistics*, 63, 471–476.

Findley, D. and Martin, D. (2006). Frequency domain analysis of SEATS and X-11/12-ARIMA seasonal adjustment filters for short and moderate length time series. *Journal of Official Statistics*, 22, 1-34.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics*, 82, 540–554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model. *Journal of the American Statistical Association*, 100, 830–840.

Forni, M. and Reichlin, L. (1998). Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies*, 65, 453–473.

Friedman, M. (1970). A theoretical framework for monetary analysis. *The Journal of Political Economy*, 78, 193–238.

Gomez, V. and Maravall, A. (1994). Estimation, prediction and interpolation for nonstationary series with the kalman filter. *Journal of the American Statistical Association*, 89, 611–624.

Gonzalo, J. and Granger, C. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics*, 13, 27–35.

Gosselin, M. and Tkacz, G. (2001). Evaluating factor models: an application to canadian inflation. Technical report, Mimeo.

Greene, W. (2008). *Econometric analysis (6. edition)*. New Jersey: Prentice Hall.

Gudmundsson, M., Petursson, T. G., and Sighvatsson, A. (2000). Optimal exchange rate policy: The case of Iceland. *Central bank of Iceland, Working Papers No. 8.*

Gunnell, D., Lopatatzidis, A., Dorling, D., Wehner, H., Southall, H., and Frankel, S. (1999). Suicide and unemployment in young people. Analysis of trends in England and Wales, 1921-1995. *The British Journal of Psychiatry*, 175, 263–270.

Gunnell, D., Middleton, N., Whitley, E., Dorling, D., and Frankel, S. (2003). Why are suicide rates rising in young men but falling in the elderly? A time-series analysis of trends in England and Wales 1950-1998. *Social Science & Medicine*, 57, 595–611.

Hall, R. (1988). The relation between price and marginal cost in US industry. *Journal of Political Economy*, 96, 921–947.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57, 357–384.

Harvey, A. (1989). *Forecasting, structural time series models and the kalman filter*. Cambridge: University Press.

Harvey, A. and Pierse, R. (1984). Estimating missing observations in economic time series. *Journal of the American Statistical Association*, 79, 125–131.

Hayton, J., Allen, D., and Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7, 191–205.

Keynes, J. (1937). The general theory of employment, interest and money. *The Quarterly Journal of Economics*, 51, 209–223.

King, R. and Rebelo, S. (1999). *Resuscitating real business cycles: Handbook of macroeconomics, edited by John B. Taylor and Michael Woodford*, volume 1B.

Knoop, T. A. (2008). *Modern financial macroeconomics - Panics, crashes, and crises*. Oxford: Blackwell Publishing.

Kohn, R. and Ansley, C. (1986). Estimation, prediction, and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association*, 81, 751–761.

Kose, M., Otrok, C., and Whiteman, C. (2008). Understanding the evolution of world business cycles. *Journal of International Economics*, 75, 110–130.

Kose, M., Prasad, E., and Terrones, M. (2003). How does globalization affect the synchronization of business cycles? *American Economic Review*, 93, 57–62.

Kroszner, R., Laeven, L., and Klingebiel, D. (2007). Banking crises, financial dependence, and growth. *Journal of Financial Economics*, 84, 187–228.

Kydland, F. and Prescott, E. (1982). Time to build and aggregate fluctuations. *Econometrica*, 50, 1345–1370.

Lay, D. (2006). *Linear algebra and its applications (3. edition)*. New York: Addison-Wesley.

Lester, D. (1997). An empirical examination of Thomas Masaryk's theory of suicide. *Archives of Suicide Research*, 3, 125–131.

Levinovitz, A. W. and Ringertz, N. (2001). *The first 100 years*. London: Imperial College Press.

Litterman, R. (1983). A random walk, Markov model for the distribution of time series. *Journal of Business & Economic Statistics*, 1, 169–173.

Lucas, R. (1976). Econometric policy evaluation: A critique. *The Phillips Curve and Labor Markets*, 1, 19-64.

Mishkin, F. (2000). Inflation targeting in emerging market countries. *National Bureau of Economic Research, Working paper No. 7618*.

Neeleman, J., Halpern, D., Leon, D., and Lewis, G. (1997). Tolerance of suicide, religion and suicide rates: An ecological and individual study in 19 western countries. *Psychological Medicine*, 27, 1165–1171.

Nijman, T. and Palm, F. (1986). The construction and use of approximations for missing quarterly observations: A model-based approach. *Journal of Business & Economic Statistics*, 4, 47–58.

Petursson, T. G. (2000). Business cycle forecasting using regime switching. *Central bank of Iceland, Working Papers No. 7*.

Polasek, W., Llano, C., and Sellner, R. (2010). Bayesian methods for completing data in spatial models. *Review of Economic Analysis*, 2, 194–214.

Prescott, E. (1986). Theory ahead of business cycle measurement. *Carnegie-Rochester Conference Series on Public Policy*, 25, 11-44.

Rebelo, S. (2005). Real business cycle models: Past, present and future. *Scandinavian Journal of Economics*, 107, 217–238.

Sargent, T. (1978). Estimation of dynamic labor demand schedules under rational expectations. *The Journal of Political Economy*, 86, 1009–1044.

Sargent, T. (1979). A note on maximum likelihood estimation of the rational expectations model of the term structure. *Journal of Monetary Economics*, 5, 133–143.

Sargent, T. and Sims, C. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *Federal Reserve Bank of Minneapolis, Working Paper No. 55*.

Schneider, M. and Spitzer, M. (2004). Forecasting Austrian GDP using the generalized dynamic factor model. *Austrian Central Bank, Working Papers*.

Sims, C. (1980). Macroeconomics and reality. *Econometrica*, 48, 1–48.

Stock, J. and Watson, M. (1989). New indexes of coincident and leading economic indicators. *National Bureau of Economic Research, NBER Book Series Macroeconomics Annual*, 4, 351–409.

Stock, J. and Watson, M. (1998). Diffusion indexes. *National Bureau of Economic Research, Working Paper No. 6702*.

Stock, J. and Watson, M. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44, 293–335.

Stock, J. and Watson, M. (2005). Understanding changes in international business cycle dynamics. *Journal of the European Economic Association*, (5, 968–1006).

Tómasson, H. (1989). Analysis of the stability and cyclical properties of economic growth in OECD countries: A description of the period 1961-1989. *University of Iceland, Working Papers No. 14*.

Tsay, R. (2005). *Analysis of financial time series, (2nd edition)*. New Jersey: John Wiley & Sons.

Velicer, W., Eaton, C., and Fava, J. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. *In R. D. Goffin & E. Helmes (Eds.), Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy. Norwell, MA: Kluwer Academic.*, 41–71.

Watson, M. (2003). Macroeconomic forecasting using many predictors. *Econometric Society Monographs, Cambridge University Press*, 37, 87–114.