LUND UNIVERSITY
School of Economics and Management

Leading indicators: myth or legend?

By Lina Fransson & Måns Hammarbäck

Master thesis

Department of Economics

Supervisor: Fredrik NG Andersson

2012-05-25

Abstract

In this thesis we analyze the predictive power of various traditionally leading indicators of GDP-growth. We apply a dynamic panel data estimation approach and use variables categorized in to two groups; survey indicators and economic variables. The sample is transformed using a wavelet analysis to obtain the short-, medium- and long run impact of the variables. The results show that there are difficulties in estimating the short run fluctuations while there is large explanatory power in the medium run.

Keywords: leading indicators, business cycle, dynamic panel data model, systematic GMM, wavelet transformation

Contents

1. Introduction

Economic decisions made by agents are to a large extent based on forecasts of the future development of macroeconomic variables. Firms' decision to produce, private agents' decision to consume or save and fiscal and monetary policy decisions depend on the future outlook of the business cycle (Camacho & Perez-Quiros, 2010). In this context, the object of an econometrician is to forecast business cycle fluctuations and produce the best possible estimation of economic activity (Dovern & Ziegler, 2008). To predict the economic activity we require variables that move in and out of recessions before the rest of the economy (Camacho & Perez-Quiros, 2002).

The overall economic activity measured as GDP and the national accounts are published with a considerable time lag of about 14 weeks after the end of the respective quarter (Camacho & Perez-Quiros, 2010). The initial publication provides a rough estimate that is often later revised, especially around turning points of economic activity (OECD, 2007; OECD, 2010; Dovern & Ziegler, 2008). The complexity of determining and forecasting the business cycles of economies becomes evident as Hamilton (2010) writes; "[T]he NBER[1] dated the 1990-91 recession as beginning in August 1990 and ending in March 1991. It made the announcement that the recession had begun in April 1991 – one month after the NBER later decided that the recession was already over. The end of the 2001 recession was announced in July 2003, which is 28 months after the recession is now deemed to have ended."

Several institutions, national as well as international continuously publish wide variety of macroeconomic variables and various leading indicators of economic activity (OECD, 2007). The publication of business cycle indicators occurs at a much higher frequency and considerably lower time lag compared to the national accounts (Giannone, et al., 2007). Therefore it is of considerable interest in being able to forecast the economic activity using these indicators and interpreting the change.

A common practice among firms and institutions is to date reference cycles using composite leading indicators. These aggregate time series combine the information of surveys and macroeconomic data to create a reference series leading the business cycle (Stock &

[1] National Bureau of Economic Research

Watson, 2010). The purpose of the composite leading indicator published by OECD[2] is to forecast the business cycle by six to nine months (OECD, 2007). While the ambition of the indicator is to forecast the short run fluctuations it is created in such a manner that the cyclical short- and long run components are removed. Thus, using a composite indicator it is difficult to distinguish between the aggregate components short-, medium- and long run effects on the economy.

As variables are assumed to have leading qualities over different horizons there seem to be no indicator “always working” (Dovern & Ziegler, 2008). Therefore it is of interest to be able to distinguish between time horizons and estimate variables to examine the effects. To overcome the problem of identification between different time horizons the data can be transformed using a wavelet transformation to create a short- medium- and long run time horizon.

Empirical work such as the introduction of the EURO-STING (Camacho & Perez-Quiros, 2010) and real-time work by Giannone *et al.* (2007) has improved the usage of the first releases and real-time data. These authors evaluate the short-term forecasts from a factor model in a truly real time setup with a time horizon of six to nine months. Their focus is on replicability and transparency rather than decisions based on own judgment. Other authors have published articles contributing to the literature on dating reference cycles. Empirical work such as Chauvet and Piger (2005) and Camacho and Perez-Quiros (2010) try to identify turning points using a few or just one aggregate while authors such as Stock and Watson (2010), Dovern and Ziegler (2008) and Babecký *et al.* (2011) analyze the power of various indicators to predict growth rates and the GDP-gap. In this thesis we strive to examine variables that have predictive power over GDP in a short- and medium run horizon. We follow the methodology of Bebecky et al. (2011) and perform a systematic literature review to inherit a group of variables we find likely to be significant in capturing the business cycle movement. We apply a dynamic panel data estimation technique to extract the marginal impact of selected leading indicators and variables.

Our results point towards difficulties in estimating the short run fluctuations while there is large explanatory power in the medium run. The non-significant short run results and very

[2] Organization for Economic Co-operation and Development

significant medium run results can be interpreted as if the data has no predictive power in determining the short run fluctuations while in the medium run; the data can explain much of the fluctuations. For both time horizons the variable that can explain most of the GDP movements is the lagged value of GDP.

The remainder of the thesis is structured as follows. In section two we discuss the empirical approach that we use to assess the different variables and the data included in our analysis. Further we discuss the methodology of our thesis and discuss the econometrics used to test our hypothesis. We continue by discussing the indicators included as regressors and their ability to forecast the GDP-gap. Finally, we present our analysis and discuss our results and the implication our findings. In section three we conclude our results over what horizon our variables have predictive power and finish with a discussion of our research.

2. Empirical analysis

In this section we discuss the data used to examine leading qualities and the econometric model applied for estimation. Further we discuss the wavelet transformation to the data and its implications and the usage of a principal component to avoid multicollinearity in variables.

In order to estimate the GDP-gap a dynamic panel data model is specified. Economic variables expected to have leading qualities over GDP are collected for nine OECD countries and divided into two time horizons using a wavelet transformation. The data is available from 1992q1 to 2011q4 at quarterly frequency and the countries are selected based on data availability. The countries included are Australia, Denmark, France, Finland, Germany, Netherlands, Sweden, United Kingdom and the United States. The variables included need to be comparable between countries and collected under a sufficient long time horizon. In the estimation procedure a systematic GMM is applied to obtain efficient and unbiased estimates.

2.1 Model specification

The dynamic panel data model can be specified as:

$$y_{it} = \sum_{j=1}^{p} \alpha_j \, y_{i,t-j} + \sum_{l=0}^{q} x'_{MR,it-l} \, \beta_{l,MR} + \sum_{m=0}^{r} x'_{SR,it-m} \, \beta_{m,SR} + u_{it} \qquad (1)$$

where i denotes country and t denotes time. The GDP-gap is the dependent variable, y_{it} and α_j are scalars where p is the order of lagged GDP values included. The short- and medium run leading indicators are denoted $\boldsymbol{x'_{SR,it-m}}$ and $\boldsymbol{x'_{MR,it-l}}$. They are vectors of $q \; x \; K$ and $r \; x \; K$ where K is the number of leading indicators and q and r is the order of lagged variables of the leading indicators included. $\beta_{l,MR}$, $\beta_{m,SR}$ are defined as $K \; x \; q$ and $K \; x \; r$. The error term is assumed to follow a two-way error component with fixed effects.

$$u_{it} = u_i + \lambda_t + v_{it}$$

Where u_i denotes country specific fixed effects capturing unobserved differences between countries that cannot be explained of the regressors and λ_t denotes time-fixed effects capturing changes over time that are assumed to affect all countries in the same way. This is for example a global shock hitting all countries at the same time, expected to influence all countries GDP equally. u_i and λ_t are the fixed parameters to be estimated and v_{it} denotes the remainder stochastic disturbance term with $v_{it} \sim IID(0, \sigma_v^2)$. The model is a fixed effects model since the sample is a selection of nine OECD countries and not randomly selected (Baltagi, 2005).

Historically, GDP regularly appears to lag the real economy and is often subject to large revisions of the first estimates. This has been shown to be particular evident in turning points where the economy moves into a new phase. The aim of using the econometric model is to find significant variables that have explanatory power over GDP and which can early signal economic turning points. The process where GDP adjusts to changes in other economic variables may depend on the passage of time and argues for lagged versions of these variables as regressors in the model. The process may also depend on the difference between potential GDP and the previous quarters observed GDP since GDP expects to adjust to its potential growth in the long run (Roodman, 2009). The model in (1) captures the dynamic relationship by adding lagged dependent variables among the regressors (Brooks, 2008, p. 291).

To examine how the relationship between the variables changes dynamically and to conduct any meaningful hypothesis testing it requires a sufficient number of observations. To solve the problem with a relative short time series a panel of data is used which combine time series data with cross sectional data. By employing information of many countries at the same time in a set of time series data, the number of degrees of freedom increase and thus the power of the tests increase (Brooks, 2008, p. 488). When working with time series in econometrics it should always be determined whether the time series follow a stationary or non-stationary process. Improperly modeling a non-stationary series as a stationary series could result in spurious regression and misleading results. However, business cycles are by nature stationary and thus, no test for unit root has to be conducted (Harris & Sollis, 2003)

2.2 Method

2.2.1 Determining business cycles

The phenomena of business cycles have long engaged researches and one of the biggest challenges has been the ability to capture the business cycles time changing characteristic (Dovern & Ziegler, 2008). The idea of different time horizons in economic decision-making is well recognized since the relationship between economics variables is expected to vary over time. Autoregressive fractionally integrated moving average, ARFIMA, and error correction models, ECM, are examples of econometric models that allow for a time horizon perspective. The ARFIMA has problems distinguishing the components to the economic time horizons. The ECM is easier to tie to economic theory but can only distinguish between a short- and long run time horizon. A band pass filter is another approach that allows for distinguishing between the two time horizons. Examples of a band pass filters is a simple moving average or the Hodrick-Prescott filter, which among others, OECD use to construct their leading indicator. Both the econometric method and the band pass filter method are classical time domain representations which decompose the series into a cyclical component and trend component (Andersson, 2008).

Instead of representing the data in a time domain the data can move to a frequency domain. The frequency domain is much more appealing when studying the cyclical behavior of a variable and allows the series to be decomposed into more than two time horizons. The time horizons can easily be defined and individual parameter vectors can estimate the economic behavior at each horizon. The Fourier transformation was first described in the beginning of the 19th century and uses a set of sine and cosine functions to represent the time series at one particular frequency (Andersson, 2008). The frequencies can then divide the economy into several time horizons where the low frequencies represent the long run and the high frequencies represent the short run time horizon. The major drawback of the Fourier analysis is that it requires the variables to be stationary since it does not contain a time resolution but only a frequency resolution. In economic data, this is often not the case and many economic variables grow over time. A method to overcome this problem is to use a wavelet analysis, a field that has rapidly expanded in the last decades in a wide variety of disciplines including astronomy, engineering and physics. Despite its success in many other disciplines the application in economic and finance is relatively new (Crowley, 2007).

The wavelet transformation has similar properties as the Fourier transformation and uses the frequency domain to decompose the series into several time periods. In contrast to the Fourier analysis, the wavelet analysis also has the desirable time resolution. The time axis is divided into smaller sub samples, windows, which length depend on the frequency. The window is narrow for high frequencies and wide for low frequencies. The wavelet analysis is then “capable of capturing simultaneously the time-varying nature of low frequency cycles and the frequency distribution of sudden and abrupt shocks in the original time series" (Raihan, et al., 2005).

By definition, wavelets are small waves that begin and die out within a finite time period. In this sense, wavelets are not homogenous over time thus different behavior can be found in different time periods (Crowley, 2007). The tradeoff for both having a frequency and a time resolution is that the wavelet analysis is less detailed in the frequency resolution than the Fourier analysis. This is however a minor problem when modeling business cycles since the number of time horizons are limited and in this study only two out of three time horizons are included (Andersson, 2008).

The wavelet transformation applied is the maximal overlap discrete transformation (MODWT), which does not impose any restrictions on the sample size. The economy is divided into three time horizons, long-, medium- and short run where the low frequencies represent the long run and the high frequencies represent the short run. The time horizon representing short run last up to two years, medium run from two to eight years and long run from eight years and beyond[3].

The focus in the analysis is to determine which variables that have a predictable power over GDP in the short- and medium run. The phenomenon of achieving long-run growth is one of the cornerstones in classic growth theory and includes well-known models such as the Solow- and the Romer model (Jones, 2002). Although the literature presents a wide range of different models and factors driving the economy, economists broadly agrees that the main factors responsible for long run growth is increased population, increased capital stock and developments in technology (Hamilton, 2010). These factors are not included in this analysis since the focus is not to determine long-run growth; therefore the long run time horizon will not be included in the model. The desired outcome is instead to find variables that can help explain GDP-growth in a time horizon not too far away into the future but still within a horizon where changes have time to be implemented. The transformation to short- and medium run enables the separate analysis of what effects the leading indicators have on GDP. Some variables affect GDP both in short- and medium run while others have only a short- or a medium run effect on GDP. It is also possible to find variables that have different

[3] Given quarterly data the long run is represented by the frequencies 0 to 1/32, medium run by 1/32 to 1/8 and short run by 1/8 to 1/2.

effects in the short- and medium run (Andersson, 2011). The wavelet transformation can then help to form a better understanding of how the economy will evolve over time or what impact changes in relevant variables will have on GDP (Kim & In, 2003).

Combining the wavelet-transformed data with the dynamic panel data model makes it possible to discover an economic relationship between the set of leading indicators and economic growth at different time horizons. In theory, model (1) is correctly specified and the short- and medium run variables can be estimated in the same regression. The problem with model (1) is that it includes “too many” variables to estimate given the pre-selected variables and the sample size. The model is therefore specified for each time horizon separately where GDP is divided into short - and medium run:

$$y_{SR,it} = \sum_{j=1}^{p} \alpha_j \, y_{SR,it-j} + \sum_{m=0}^{r} x'_{SR,it-m} \, \beta_{m,SR} + u_{SR,it} \tag{2}$$

$$y_{MR,it} = \sum_{j=1}^{p} \alpha_j \, y_{MR,it-j} + \sum_{m=0}^{r} x'_{MR,it-m} \, \beta_{m,MR} + u_{MR,it} \tag{3}$$

The separation of the two time horizons into two models will result in wrong error components and consequently an estimate of $var(\hat{\beta})$ cannot be obtained. The correct error term, $\hat{u}_{it}$ is found when including both short- and medium run in the model but the model specified as in (2), only $\hat{u}_{SR,it}$ is estimated. Under the assumption of white noise, which implies that each frequency has the same length, $\hat{u}_{SR,it}$ can be used to obtain an estimate of $var(\hat{u}_{MR,it})$ (Andersson, 2011). Therefore the ex post standard errors need to be adjusted to yield the correct t-statistic. This is done by using the frequency interval for each time horizon and multiplies $var(\hat{u}_{SR,it})$ with $6/8$ and $var(\hat{u}_{MR,it})$ with $32/6$. The correct t-statistic can then be calculated as:

$$t - stat = \frac{\widehat{\beta}_{SR}}{var(\hat{u}_{SR,it}) * \frac{6}{8}} \tag{4}$$

$$t - stat = \frac{\widehat{\beta}_{MR}}{var(\hat{u}_{MR,it}) * \frac{32}{6}} \tag{5}$$

Post estimation testing for normality and heteroscedacity indicate no problem with incorrect error components and the short- and medium run residuals can be directly used. However, testing for autocorrelation will no longer be appropriate. This can be seen when the frequency goes to zero and eventually only becomes a straight line, the correlation between residuals will approach one. In the case when the frequencies are divided into different time horizons the medium run model will be without the short run fluctuations. This implies that when we estimate an AR-process for medium run, the error terms will be correlated even though the error terms are white noise. The model cannot distinguish between different autoregressive parameters and testing for autocorrelation becomes meaningless (Andersson, 2011).

2.2.2 Estimator

By construction, the models are characterized by persistence over time (Baltagi, 2005, p. 135). The data generating process is as already stated as a dynamic process, with the current value of GDP influenced by past values of GDP (Roodman, 2009). Since the current value of GDP is a function of the fixed effects it directly follow that past values of GDP, entered as regressors in the model, are also functions of the fixed effects and therefore correlated with the error term. This will make the OLS estimator biased and inconsistent. The within transformation wipes out the fixed effects but the average GDP will still be correlated with the error term and the estimators consistency will depend on T being large (Baltagi, 2005, pp. 34, 135). Another problem with OLS is that the variables are assumed to be endogenous. The variables are endogenous since the error term is expected to have some feedback on the realizations of the regressors. This can be assumed since unforecastable errors today and in the future are expected to affect changes in the leading indicators (StataCorp, 2009, p. 32).

Therefore a systematic GMM proposed by Blundell and Bond (1998) will be used as estimator. The Blundell-Bond estimator builds on the Arellano-Bond first difference GMM estimator, which constructs instruments by taking the first difference and then using lagged variables in level as instrument for the first difference model. For example y_{t-2} becomes an instrument for $(y_{t-1} - y_{t-2})$ since y_{t-2} is not correlated with $(v_t - v_{t-1})$. Arellano-Bond use variables in levels as instrument for variables in first differences and Blundell and Bond show that this estimator "become weak as the autoregressive process becomes too persistent or the ratio of the variance of the panel-level effects v_i to the variance of the idiosyncratic error e_{it} becomes too large" (StataCorp, 2009, p. 97). The Blundell-Bond system estimator increase the efficiency by adding a moment condition where lagged differences are used as instruments in addition to the level instruments used by Arellano-Bond. Blundell and Bond show that the system GMM increase precision and reduce the finite sample bias present in the first-difference GMM (Blundell & Bond, 1998).

2.3 Pre-selected variables

There are several approaches in determining which variables to use as regressors. Some studies choose a few variables assumed to be leading for the GDP according to economic theory. These studies usually work with a small amount of reference variables. Many times these models are limited by the theory they are to explain due to imbalances between explanatory effect in variables and effects not captured by the theory. Other studies focus on including as many variables as possible finding those variables with the best leading qualities, disregarding the economic theory behind (Babecký, et al., 2011). We have chosen variables based on our own judgment from a wide literature survey. This implies that we reduce the risk of missing important potential explanatory effect. We follow the methodology of Dovern and Ziegler (2008), Stock and Watson (2010), NBER (2008; 2010) and Babecký, et al. (2011) and test different macroeconomic time series and various indicators with their ability to improve and predict the GDP-gap.

2.3.1 Survey indicators

Survey indicators and business opinions provides qualitative information about agents and producers expectations of the economic development. The surveys are typically based on enterprise- and household interviews where respondents are asked about their current situation and their expected outlook. Households are asked for the tendency of major purposes such as future employment, economic endowment and their general economic situation while enterprises are asked about the tendency in production, orders and stock etc. Survey indicators can extract public information about future expectations and enables an increase in the forecast accuracy. The indices created can then be used to forecast the business cycle and has been shown to improve the estimation of turning points. Although some studies (Carroll, et al., 1994) say that surveys not only forecast changes in spending but also cause them. Philip Howrey (2001) concludes that consumer surveys statistically improve the predictive power of future growth rates of real GDP. The business survey indicators included are production tendency, employment future tendency, order books and order inflow tendency, all for the manufacturing industry. To capture households' expectations consumer confidence is included.

2.3.2 Economic variables

Economic activity variables often have the advantage over GDP data that they are published on a timelier basis and are less subject for revision. As discussed, the difficulty to obtain correct estimates of GDP, especially around turnings points is evident. Using economic variables as proxies for GDP can increase the accuracy when determining economic activity and thus better indicate turning points (Dovern & Ziegler, 2008). There is little or no lag in the publication of these statistics as they are often based on count data and therefore reduce the processing time. The economic variables can be categorized into two types of data sets, real activity variables and financial variables.

The real activity variables included are retail trade volume, passenger car registration and real house prices. These variables are included in some countries composite leading indicator constructed by the OECD and are all shown to have leading qualities over GDP (OECD, 2012). All real activity variables measure the households' actual consumption to some extent. Retail trade can be seen as a proxy for daily business and the current state of the economy while passenger car registration and house prices are larger investments where expectations about the future are incorporated in the consumer's decision.

Financial variables capture agent's future expectations of the economic activity since agents base their investment decisions on their beliefs about the future economic development (Dovern & Ziegler, 2008). The financial variables included in the analysis are real share prices, real interest rate, the spread between the ten year government bond and the short-term interest rate, real money supply and real credit growth. The inclusion of share prices reflects agent´s beliefs about future dividend payments and consumption. The short term real interest rate is assumed to capture the state of investment opportunities in the economy and therefore contain information about the future economic outlook. Stock and Watson (1989) and Estrella and Hardouvelis (1991) among others have found that spread has predictive power over GDP. The spread between the long- and short run interest rates captures the structure of the yield curve, which is suggested to have extra predictive power beyond the using only the short term interest rate (Brooks, 2008, p. 303). The relationship between the interest rate and GDP as well as the spread and GDP is expected to be negative. A higher interest rate lowers the investment activity and causes a drag on GDP. An increase in the spread is the result of a higher risk premium on the long-term interest rates and indicates more uncertainty in the future, which also lowers the financial activity (Stock & Watson, 1989; Fama, 1990). Real money supply, M3, and real credit growth measure the liquidity available in the economy and can also give an indication of future investments opportunities.

2.4 Data sample

Many variables can be found in OECDs statistical database where comparable data for the selected countries is available. The data not found in OECDs database is obtained from Thomson Financial DataStream, Reuters Ecowin and BIS. Monthly data is transformed to a quarterly frequency using the average for all variables except for the business tendency surveys where the end value of each quarter is used. All data expressed as indices is transformed into logarithmic values. The data presented in percent and net balance is estimated in its original units. Table one summarizes the data.

Comparable variables measuring credit growth has been difficult to find and lending to households is used as a proxy for credit growth. House prices in Germany are only available on a yearly basis. The data is transformed using linear interpolation. For Sweden, the OECD index for consumer confidence is only available from 1995. To solve the problem, SCBs[4] consumer confidence measured in net balance is used to determine the confidence direction and then used for interpolation. The consumer price index for each country is used to obtain real values for share prices, house prices, interest rate and lending to households. The spread is calculated by subtracting the three-month interbank rate from the ten-year government benchmark yield.

The series business survey indicators in the manufacturing industry are highly correlated. To improperly estimate these series as regressors in the same estimation would result in multicollinearity (Giannone, et al., 2007). To solve the problem a principal component technique is applied. The principal component (BUSSUR) serves as a factor and reduces the number of variables without losing too much information in the covariance matrix (Campbell, et al., 1997, p. 236). The first principal component weights the survey series to maximize the variance explained by the component. To construct an appropriate factor, we use a benchmark where the first principal component should be able to explain at least 50

[4] Statistiska Centralbyrån

percent of the series variation. Most of the countries acquire first principal component able to explain almost a 100 percent. For two countries, Finland and France, the first principal component is just below 50 percent. Even though the two principal components do not accomplish explaining the variation required, the problem will be ignored since only two components are affected and the divergence is small. Table one summarize the variables included.

Table 1

Variable	Name	Source	Units	Transformation
Survey indicators				
Production tendency	PROD	OECD	Net balance	
Employment future tendency	EMPL	OECD	Net balance	
Order books	BOOK	OECD	Net balance	
Order inflow tendency	INFLOW	OECD	Net balance	
Business survey factor	BUSSUR		Net balance	First principal component
Consumer confidence index	CONF	OECD & SCB	Index	Log
Economic variables				
Gross Domestic Product	GDP	OECD	Index	Log
Retail trade volume	TRADE	OECD	Index	Log
Passenger car registration	CARS	OECD	Index	Log
House prices	HOUSE	Reuters Ecowin & BIS	Index	Deflated & Log
Share prices	SHARE	OECD	Index	Deflated & Log
Real interest rate	IR	OECD	Percent	Deflated
Spread 10yr - 3mth	SPREAD	OECD	Percent	
Money supply, M3	M3	OECD & Thomson Financial Datastream	Values	Deflated & Log
Lending to households	LENDING	Reuters Ecowin	Values	Deflated & Log

2.5 Results

2.5.1 Cross correlation analysis

An early examination of the data to analyze the variables leading qualities is to calculate the cross-correlation. Table two summarizes the results using the data for the whole time horizon, not divided into short- and medium run. The cross correlation is presented for ± 3 time periods.

Table 2

Cross correlation	-3	-2	-1	0	1	2	3
Production tendency	-0.14	-0.09	-0.06	-0.01	0.02	0.03	0.01
Expected employment	0.04	0.10	0.17	0.23	0.23	0.22	0.17
Order books	-0.11	-0.05	0.01	0.06	0.07	0.07	0.04
Order inflow	-0.09	-0.04	0.00	0.04	0.07	0.07	0.04
Consumer confidence	0.00	0.08	0.15	0.21	0.24	0.26	0.25
Retail trade volume	0.80	0.84	0.88	0.92	0.88	0.84	0.79
Car registration	0.23	0.27	0.32	0.36	0.36	0.36	0.35
House prices	0.72	0.75	0.78	0.81	0.78	0.75	0.71
Share prices	0.42	0.49	0.55	0.61	0.62	0.61	0.59
Real interest rate	-0.44	-0.51	-0.57	-0.62	-0.58	-0.53	-0.48
Spread 10yr - 3mth	-0.21	-0.16	-0.12	-0.07	-0.05	-0.03	-0.03
Money supply, M3	0.79	0.82	0.85	0.87	0.82	0.75	0.69
Lending to households	0.17	0.17	0.17	0.17	0.16	0.14	0.12

As can be seen in table two there is mixed results. The best performing series would be the one with a sufficiently high correlation with lagged GDP. The business survey series has leading qualities over GDP since the highest correlation is between the current value of the survey series and GDP one or two quarters ahead. It is only expected employment that has a correlation worth mentioning. Consumer confidence leads GDP by two quarters and has a correlation of 0.26. Share price index has a correlation with one-quarter ahead GDP of 0.62. The spread between the 10-year government benchmark yield and the 3-month interbank rate has the highest correlation when GDP leads the spread with three quarters. This result is notable since the study is anticipating the opposite relationship. A similar relationship can be found in lending to households. Real interest rate coincides with GDP at a negative correlation of -0.62. Real house prices, real trade and real money supply are coincident with GDP and have a correlation of 0.81, 0.92 and 0.87 respectively. Passenger car registration leads GDP by two quarters with a correlation of 0.36. The leading qualities of the cross correlation analysis is questionable for some series, one reason being that the examination is undertaken on the panel data where some countries are missing a relationship among variables.

To analyze the cross correlation for different time horizons the ex post transformed results are compared on a short- and medium run horizon. The correlation between GDP and the variables for the short run is presented in table three.

Table 3

Short run	-3	-2	-1	0	1	2	3
Production tendency	-0.02	-0.03	0.01	0.13	0.05	-0.13	0.00
Expected employment	0.00	-0.01	0.05	-0.05	0.01	0.07	-0.05
Order books	-0.08	-0.02	0.03	0.15	-0.01	-0.06	-0.02
Order inflow	-0.01	-0.08	0.04	0.05	0.07	-0.06	0.00
Consumer confidence	-0.04	-0.01	0.05	0.03	0.07	-0.03	-0.09
Retail trade volume	-0.07	-0.10	-0.01	0.23	-0.07	-0.10	0.04
Car registration	0.04	0.02	-0.08	0.04	0.02	-0.03	0.07
House prices	-0.03	-0.04	0.01	0.03	0.04	-0.02	0.00
Share prices	0.02	-0.02	-0.02	0.00	0.02	0.00	0.00
Real interest rate	-0.09	0.09	0.09	-0.04	-0.03	-0.05	-0.02
Spread 10yr - 3mth	0.04	-0.06	-0.05	0.04	0.03	0.02	-0.01
Money supply. M3	-0.01	-0.03	-0.02	0.04	0.00	-0.03	0.01
Lending to households	0.00	0.01	0.01	-0.06	0.02	0.03	0.00

Now instead examining the data expressed as the short run of up to two years instead of the whole time horizon it can be concluded that there is a similar pattern. Looking at the short run it yields mixed results. As in the full time horizon the survey variables have some leading qualities over GDP but the two variables which yield the highest correlation (order books and production tendency) coincides with GDP. As for the correlation in the full time horizon both consumer confidence and share price index lead GDP, now with one quarter instead of two and with a low correlation of 0.07 and 0.02. The spread between the 10-year government benchmark yield and the 3-month interbank rate and real interest rate have a negative correlation of -0.06 and -0.09 and yield lagged qualities over GDP of two and three quarters. Lending to households has a positive correlation of 0.01 when lagged for two quarters. Real house prices and passenger car registration are both leading GDP with one and three quarters respectively. Real house price has a correlation of 0.04 with GDP and new car registration a correlation of 0.07. As in the whole time horizon, real trade and real money supply are coincident with GDP and has a correlation of 0.23 and 0.04.

The results from the medium run correlation measured as effects between two- and eight years are presented in table four.

Table 4

Medium run	-3	-2	-1	0	1	2	3
Production tendency	-0.25	-0.06	0.15	0.32	0.38	0.33	0.19
Expected employment	-0.07	0.19	0.39	0.48	0.43	0.25	0.02
Order books	-0.15	0.10	0.32	0.45	0.44	0.31	0.11
Order inflow	-0.14	0.04	0.22	0.33	0.33	0.24	0.09
Consumer confidence	-0.11	0.03	0.16	0.24	0.25	0.17	0.04
Retail trade volume	-0.24	-0.08	0.11	0.27	0.35	0.32	0.20
Car registration	-0.30	-0.11	0.11	0.29	0.37	0.34	0.22
House prices	-0.10	-0.03	0.06	0.14	0.13	0.08	0.03
Share prices	-0.20	-0.11	0.01	0.14	0.21	0.21	0.16
Real interest rate	0.22	0.14	0.03	-0.08	-0.17	-0.20	-0.16
Spread 10yr - 3mth	-0.17	-0.13	-0.05	0.03	0.10	0.13	0.12
Money supply. M3	-0.09	-0.06	0.00	0.06	0.07	0.06	0.03
Lending to households	0.07	0.07	0.05	0.02	-0.01	-0.03	-0.05

In the medium run the correlation increase compared to the short run. Employment expectations and order books coincide with GDP and have correlations of 0.48 and 0.45. The other survey variables, order inflow, production tendency and consumer confidence have leading qualities over GDP with one lag and a positive correlation of 0.33, 0.38 and 0.25 respectively. The real share price index is leading in two quarters with a correlation of 0.21. The 10-year government benchmark yield and the 3-month interbank rate have a negative correlation of -0.17 and lagged qualities over GDP in three quarters. A similar relationship can be found in lending to households but with a positive correlation of 0.07. The real interest rate has a negative correlation of -0.20 and in the medium run it has a leading quality over GDP in two quarters. Real house prices coincide with GDP with a correlation of 0.14. New car registration, real trade and real money supply lead GDP by one quarter with a correlation of 0.37, 0.35 and 0.07 respectively.

The mixed results when comparing the whole time horizon with the two frequency domains can be explained as that the whole time domain nests both the short- and medium run. This implies that the correlation of both short- and medium run are weighted in the full time domain where it can be see that the medium run dominates the short run since the correlation is higher. Both time horizons with lower frequency (short- and medium run) yield similar results when comparing leading and lagged effects between variables. The biggest difference between the two frequency horizons is the interest rate which in the short run is negative and yields lagged qualities over GDP while in the medium run it yields leading qualities.

2.5.2 Estimation

To determine the predictive power of the variables we estimate a dynamic panel data model as presented in (2) and (3) for the short- and medium run. The models are estimated using a systematic GMM which yields unbiased and efficient estimates and allow endogenous variables to enter as regressors.

To determine the appropriate lag structure, the Schwarz Bayesian Information Criterion (SBIC) is applied. The lag length is chosen from the model specification that minimizes the value of SBIC. Since GMM is not a maximum likelihood function the likelihood ratio cannot be used to calculate SBIC and the residual variance σ_u^2 is used instead. The SBIC provides a trade-off between goodness of fit and the number of parameters estimated. Following Sims (1980) methodology the choice of lag length will be uniform across all explanatory variables. If variables are chosen with differing lag length to best fit the data this can be seen as a restriction to the model thus reducing the power.

Previously we have argued for a dynamic model with past lags of GDP entering as regressors, however, in the estimation procedure we limit the lag length of GDP to one. This is done with the motivation that the aim of the study is to find leading variables that can determine the development of GDP and not to estimate how previous values of GDP can predict the future. For the predicting variables we limit the lag length to three variables. The inclusion of too many lags would increase the risk of business cycles overlapping each other since the data is quarterly. The estimated coefficients would then be difficult to interpret. The highest SBIC-value is achieved when one lag is included in both the short- and medium-run model.
The estimation results for the short run model are presented in table five.

Table 5

System dynamic panel-data estimation - short run - one lag			
	Coefficient	Std.error	t-dist.
GDP(-1)	-0.212***	0.021	-10.18
BUSSUR	0.019***	0.000	3.68
BUSSUR(-1)	0.005	0.000	1.00
CONF	0.000	0.000	1.60
CONF(-1)	0.000	0.000	2.01
SHARE	-0.005***	0.001	-3.39
SHARE(-1)	-0.002	0.002	-1.07
SPREAD	0.000	0.000	1.63
SPREAD(-1)	0.000	0.000	-0.91
IR	0.006	0.019	0.30
IR(-1)	-0.026	0.019	-1.37
LENDING	-0.004**	0.001	-2.74
LENDING (-1)	-0.001	0.002	-0.44
HOUSE	-0.020**	0.008	-2.60
HOUSE(-1)	0.007	0.007	0.99
CARS	-0.002	0.001	-1.70
CARS(-1)	0.001	0.001	0.59
TRADE	0.120***	0.009	12.68
TRADE(-1)	-0.008	0.010	-0.84
M3	0.009	0.007	1.28
M3(-1)	-0.006	0.007	-0.89
sig^2	0.00	**DF**	21
N	711	**SBIC**	-11.15
Notes: *, **, and *** indicates significance levels of 10%, 5%, and 1%.			

The short run model represents GDP fluctuations that last up to two years. As can be seen in table five the factor variable capturing business survey tendencies, share prices, lending to households, house prices and retail trade are significant in the current value. There is no significance in any lags. The business survey factor is significant in the current value and positive. Since the variable is a factor variable and measured in net balance it is hard interpret the magnitude of the effect but we conclude that an increase in the business sentiment will have a positive effect on GDP. According to the estimation, a one percent increase in share prices will affect GDP negatively by 0.005 percent. The result is rather unexpected since share prices are assumed to have a positive effect on GDP. The effect is so small that no real interpretation or explanation needs to be considered. A one percent increase in lending to households affects GDP negative by 0.004 percent. The effect is unexpected but very small. A one percent increases in house prices affects GDP negatively by 0.02 percent and a one percent increase in retail trade affects GDP positively by 0.12 percent. Retail trade is the only series with a correct sign and with a sufficient effect on GDP. This seems reasonable since retail trade is a measure of households' actual consumption, which is one of the components in the GDP decomposition.

The results show that most of the variables have poor information in explaining short run fluctuations in GDP. The variable that seems to have the best properties to explain changes in GDP is the lagged values of GDP itself. The effect is negative which can be interpreted as if the short run GDP-gap is positive in current quarter and we can expect a negative correction in the following quarter. The variables estimated for the medium run that captures fluctuations between two years and eight years. The results are presented in table six.

Table 6

System dynamic panel-data estimation - medium run - one lag			
	Coefficient	Std.error	t-dist.
GDP(-1)	0.84***	0.031	27.51
BUSSUR	0.109***	0.000	5.70
BUSSUR(-1)	-0.105***	0.000	-5.43
CONF	0.001***	0.000	3.67
CONF(-1)	-0.001***	0.000	-3.51
SHARE	0.010***	0.006	1.55
SHARE(-1)	-0.004***	0.006	-0.69
SPREAD	-0.003***	0.001	-3.11
SPREAD(-1)	0.003***	0.001	2.69
IR	-0.192***	0.056	-3.45
IR(-1)	0.141***	0.056	2.49
LENDING	0.015***	0.004	3.82
LENDING (-1)	-0.013***	0.004	-2.87
HOUSE	0.031***	0.015	2.06
HOUSE(-1)	-0.010***	0.015	-0.65
CARS	0.010***	0.004	2.28
CARS(-1)	-0.002***	0.005	-0.43
TRADE	0.051***	0.034	1.49
TRADE(-1)	0.000	0.035	0.01
M3	0.006**	0.016	0.38
M3(-1)	0.002	0.017	0.09
sig^2	0.00	**DF**	21
N	711	**SBIC**	-14.92

Notes: *, **, and *** indicates significance levels of 10%, 5%, and 1%.

In the medium run all included variables have some predictive power over the GDP-gap over two- to eight years. The business survey factor is significant in both the current value and the first lag. The current value affects GDP positively while the lagged value affects GDP negatively. The overall effect of the coefficients is positive, but similar to the short run it will be hard to conclude the total impact on GDP. Consumer confidence show similar results as the business survey factor. Both coefficients are significant but have different signs. The overall effect of consumer confidence is positive but small. The share price index is positive in the current value but negative in the lagged value. The overall effect is that a one percent increase in share prices will increase GDP by 0.04 percent. The spread is significant in both variables. The spread is expected to show a negative sign and the coefficient for the current value is negative by 0.003 percent. If the spread would increase by one percentage point, GDP would be decrease by 0.003 percent. The lagged value is positive but the overall effect is negative. The real interest rate has a negative effect on GDP. An increase in the real interest rate is expected to cool the economy and lead to a decrease in GDP. A one percentage point increase in the current value of the real interest is expected to decrease GDP by 0.19 percent. The lagged value has the opposite effect and is expected to increase GDP by 0.14 percent. A one percent increase in lending to households affects GDP positively by 0.015 percent in the current value. Again, the lagged value has a coefficient of the opposite sign and an increase in lending to households affects GDP negatively by 0.013 percent. For house prices the coefficient in the current value is positive and one percent increase affects GDP by 0.03 percent while the lagged value has a negative impact of 0.01 percent. For passenger car registration, the coefficient in the current value has a positive effect on GDP by 0.01 percent while the lagged value has a negative effect of -0.002 percent. Retail trade has a positive effect on GDP in the current value at 0.05 percent but is not significant in the lagged value. Money supply measured as M3 has a small positive effect on GDP in the current value of 0.006 percent but is not significant in the lagged value.

The results in the medium run are much more appealing than the results obtained in the short run. All variables are significant in the current value and most of them also in the lagged value. The coefficients estimated for the current values have the correct sign according to economic theory. The opposite sign obtained in the lagged variables can be interpreted as an adjustment process over time. The overall effect over business cycle will eventually sum to zero. The size of the coefficient and how large part of GDP a change in the explanatory variable will explain depends on the volatility of a variable. A one percentage point increase in the interest rate is a larger fluctuation than a one percent increase in passenger car registration. When the variation is considered the business survey indicator is the indicator able to explain most of the variation in GDP, around ten percent. Still, the lagged value of GDP has most explanatory power and account for around 70 percent of the variation.

In general, in the short run the explanatory power among the suggested predictive variables is poor. Only a few variables are significant and their effect on GDP is small, with retail trade as the only exception. The results are quite controversial and indicate that the money and time spent on collecting and analyzing high frequency business cycle data is unnecessary. The variables do not contain any significant information on the short run fluctuations in GDP. The conclusion is that in the short run, it is hard to do correct estimation and forecasts of the short run GDP. In the medium run the explanatory variables and GDP seems to move in the same direction. In the medium run the variables can be used to achieve an early and correct estimation how GDP will develop in the coming two to eight years.

3. Conclusion

This thesis has addressed the problem of finding variables that can predict the GDP-gap. We have performed dynamic panel data estimation using GMM to identify the predictive power of traditionally leading indicators when estimating the business cycle. We have estimated a data set consisting of several variables over a cross section of nine OECD countries using quarterly data. The data has been transformed using a wavelet transformation to form a short-, medium- and long run horizon where the latter has been disregarded in the estimation. To reduce the risk of multicollinearity, highly correlated series have been aggregated using a principal component computation.

Our results show that there is low explanatory power in the short run while there is large explanatory power in the medium run. The implication of this result is that there are few- or no variables that have explanatory power over the short run GDP-gap ranging up to two years. On the contrary, in the medium we show that all included variables have predictive power over the GDP-gap. For both time horizons, lagged GDP has the most predictive power and contain most information about the future development of the GDP-gap. A conclusion we draw from our results is that collecting data on a high frequency has no forecasting value in the short run.

The low explanatory effect in the variables on the short run can be attributed a causation effect. Following Carrol et al. (1994) we cannot say whether GDP forecast the variables or if the variables forecast GDP. This question is especially raised for the survey indicators which are based on expectations. It is reasonable to assume that expectations are affected by the current economic situation and statistical data such as the tendency variables. The skewness in expectations adds to the discussion of multicollinearity and statistical independency.

Concluding our results we acknowledge the complexity of determining and forecasting the business cycles that has been stressed by several authors. The lag in the publication of official statistics, revision of data and short run fluctuations makes predictions about the short run future uncertain. A drawback of the panel data model is that we need to have comparable data over countries. This implies that instead of using the optimal set of variables for each individual country we use variables that are comparable across countries. Often there is no uniform way of processing collected data thus it can be difficult to find comparable variables across countries. Institutions such as OECD produce comparable data over several countries but these might not be the best variables leading the GDP-gap in each country. Composite indicators constructed for each individual country can therefore better lead the GDP-gap rather than a uniform measure constructed by OECD. The leading qualities of the variables constructed by the OECD capture the general movements of countries but there may be better explanatory power when investigating leading variables for specific countries.

Using wavelet analysis to estimate business cycles and the usage of dynamic panel data models is a rather new field in macroeconomics. There is great potential in the usage of these methods and we recommend research trying to forecast different time horizons. In our view, it would be interesting to perform research using larger data sets including more countries in the analysis. An analysis looking at differences in explanatory effect between developed- and developing countries could possibly find other explanatory variables. Further, we suggest that more research should be performed finding differences among the leading properties of indicators. Some indicators may show better results only predicting troughs as opposed to showing both peaks and troughs. Therefore it would be very interesting to analyze if there are any non-linear relationships among indicators.

4. Bibliography

Andersson, F. N. G., 2008. *Wavelet Analysis of Economic Time Series.* Lund: Lund Economic Studies.

Andersson, F. N. G., 2011. *Band Spectrum Regressions using Wavelet Analysis,* Lund: Lund University, Department of Economics.

Babecký, J. et al., 2011. *Early Warning Indicators of Crisis Incidence: Evidence from a Panel of 40 Developed Countries,* Charles University Prague: Working Papers IES 2011/36.

Baltagi, B. H., 2005. *Econometric analysis of panel data.* 3rd ed. Chichester: John Wiley & Sons Ltd.

Blundell, R. & Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics,* 87(1), pp. 115-143.

Brooks, C., 2008. *Introductory econometrics for finance.* 2nd ed. Cambridge: Cambridge University Press.

Camacho, M. & Perez-Quiros, G., 2002. This is what the leading indicators lead. *Journal of applied econometrics,* pp. 61-80.

Camacho, M. & Perez-Quiros, G., 2010. Introducing the EURO-STING: Short-term indicator of the Euro area growth. *Journal of applied econometrics,* pp. 663-694.

Campbell, J. Y., Lo, A. K. & MacKinlay, C., 1997. *The econometrics of financial markets.* 2 ed. Chichester: Princeton University Press.

Carroll, C. D., Fuhrer, J. C. & Wilcox, D. W., 1994. Does Consumer Sentiment Forecast Household Spending? If So, Why?. *American Economic Review,* Volume 84(5), pp. 1397-1408.

Chauvet, M. & Piger, J., 2005. *A Comparision of the Real-Rime Performance of Business Cycle Dating Methods,* St.Louis: Federal Reserve Bank of St.Louis.

Crowley, P. M., 2007. A GUIDE TO WAVELETS FOR ECONOMISTS. *Journal of Economic Surveys,* 21(2), pp. 207-267.

Dovern, J. & Ziegler, C., 2008. *Predicting growth rates and recessions, assessing U.S leading indicators under real-time conditions,* Kiel: Kiel Institute for the World Economy.

Estrella, A. & Hardouvelis, G. A., 1991. The Term Structure as a Predictor of Real Economic Activity. *Journal of Finance, American Finance Association,* 46(2), pp. 555-576.

Fama, E. F. 1., 1990. Stock Returns, Expected Returns, and Real Activity. *Journal of Finance, American Finance Association,* 45(4), pp. 1089-1108.

Giannone, D., Reichlin, L. & Small, D., 2007. *Nowcasting: The Real-Time Informational Content of Macroeconomic Data,* s.l.: Money Macro and Finance Research Group.

Hamilton, J. D., 2010. *Calling Recessions in Real Time,* Cambridge MA: National Bureau of Economic Research.

Harris, R. & Sollis, R., 2003. *Applied Time Series Modelling and Forecasting.* 2 ed. Chichester: John Wiley & Sons Ltd.

Howrey, P. E., 2001. The Predictive Power of the Index of Consumer Sentiment. *Brookings Papers on Economic Activity,* Volume 32(1), pp. 175-216.

Jones, C. I., 2002. *Introduction to economic growth.* 2 ed. New York: W. W. Norton & Company Inc.

Kim, S. & In, F. H., 2003. The Relationship Between Financial Variables and Real Economic Activity: Evidence From Spectral and Wavelet Analyses. *Studies in Nonlinear Dynamics & Econometrics,* 7(4).

OECD, 2007. *Current period performance of OECD composite leading indicators (CLIs),* s.l.: OECD.

OECD, 2008. *OECD system of composite leading indicators,* s.l.: OECD.

OECD, 2012. *OECD Composite Leading Indicators: Reference turning points and Component series.* [Online] Available at: http://www.oecd.org/document/29/0,3746,en_2649_34249_35725597_1_1_1_1,00.html [Accessed 18 05 2012].

Raihan, S. M., Wen, Y. & Zeng, B., 2005. *Wavelet: a new tool for business cycle analysis,* St. Louis: Federal Reserve Bank of St. Louis.

Roodman, D., 2009. How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal,* 9(1), pp. 86-136.

Sims, C. A., 1980. Macroeconomics and Reality. *Econometrica,* 48(1), pp. 1-48.

StataCorp, 2009. Stata: Realese 11. In: *Longitudinal Data/Panel Data.* College Station, TX: StataCorp LP.

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

Stock, J. H. & Watson, M. W., 2010. *Estimating turning points using large data sets,* Cambridge, MA: National Bureau of Economic Research.