ÖREBRO UNIVERSITET

A Coincident Indicator for the Swedish Economy

Master thesis in Applied Statistics

Candidate: Alberto Mulenga **Supervisor: Sune Karlsson**

Orebro, June 2009

A Coincident Indicator for the Swedish economy

Alberto Chicafo Mulenga

Master thesis in Applied Statistics

Swedish Business School, Orebro University SE-701 82 Orebro,

Sweden 2009

Abstract

Economic indicators are important for economic decisions. However, the indicators published by the National Institute of Economic Research (NIER) give information for parts of the whole economy; the Gross Domestic Product from Statistics Sweden is published quarterly, and with great delay. Economic decisions frequently need timely information about the state of the economy. The aim of this study is to construct a coincident indicator for the Swedish economy, which is timely and describe the state of economy. Using 68 monthly and 93 quarterly time series collected from Statistics Sweden and NIER databases I apply the factor model techniques developed by Stock and Watson (2002a), to constructed two composite coincident indicators. The new indicators are close to GDP growth rate with correlation 0.80 and 0.83; 98 and 99 percent of correct signal is predicted respectively by the monthly and quarterly indicators. The RMSFE and MAE are approximately 0.87 percentage point and 0.64 percentage points. The indicators describe the state of economic activity; moreover the monthly composite coincident indicator can provide very timely and useful information about the state of the economy.

Keywords: Composite coincident indicator, Forecast, Gross domestic product, Principal components, Stock and Watson, Time series.

List of Symbols

ADF	Augmented Dickey and Fuller
AERG	Agency for Economic and Regional Growth
AI	Activity index
AR	Autoregressive
ARMA	Autoregressive Moving - Average
CCI	Composite Coincident Indicator
CEI	Coincident Economic Index
CEPR	Centre for Economic Policy Research
CFNAI	Chicago Fed National Activity Index
DI	Diffusion Index
DI_AR	Diffusion Index - Autoregressive
DLS	Dynamic Least Squares
ETI	Economic Tendency Indicator
FHLR	Forni, Halli, Lippi and Reichlin
FSA	Financial Supervisor Authority
GDP	Gross Domestic Product
GPC	Generalized Principal Components
MA	Moving - Average
MAE	Mean Absolute Error
NIER	National Institute of Economic and Research
OLS	Ordinary Least Squares
PC	Principal Components
RMSFE	Root Mean Square Forecast Errors
SW	Stock and Watson
TPI	Turning Point Indicator

List of Tables

List of Figures

I INTRODUCTION

The Swedish National Institute of Economic Research (NIER) report and publish two monthly basic indicators of the tendency of the Swedish economy: the Economic Tendency Indicator (ETI) and the Turning Point Indicator (TPI), with these indicators is also published information related to consumer confidence, households 12 months inflation expectation and perceived inflation. Both the ETI and TPI are based on responses of firms to a number of questions in the Business Tendency Survey. According with NIER the Business Tendency Surveys is published in the series of reports monthly and quarterly. The indicators thus serve as thermometers of the Swedish economy both at present and in the near future, and they are among the many puzzle pieces used by the NIER in preparing economic forecast.

Other indicators used to measure the Swedish economy are the Gross Domestic Product (GDP), Unemployment rate, Consumer Price Index and Activity Index (AI) produced and published by Statistics Sweden. The Activity Index measures the activity in the Swedish economy. It is a regression model with the following explanatory variables: Industrial production index, Hours worked by employees in government authorities. The original series of Activity Index is calibrated to the non adjusted GDP, so that GDP and the Activity Index are equal for all quarters. The GDP, published four times year, gives the value of all goods and services produced in a country during a time period for example quarter or one year. Even though GDP is not a perfect measure its gives an idea of how wealthy a country is. It is also the most common way of describing economic growth in different countries. According to Goodwin et al (2007), the definition of economic growth is in fact an increase in real gross domestic product, however this common indicator is measured quarterly and according to Altissimo et al (2007), GDP is the most comprehensive indicator of real activity but is published with delay and is not free from short – run fluctuations. So the problem is that for business activities, for example a decision for investment and in many sectors of economic activity, the needed timely information about the state of economy is not available.

One possible solution to this problem is to use a coincident indicator for the economy. Gaudreault et al (2003), define the term “business cycle” to refer to co-movements in a broad range of macroeconomic indicators of activity such as output, employment and retail sales. As in Stock and Watson, the coincident economic index (CEI) represents the unobservable “state of the economy” and is coincident with the business cycle, that is move up or down directly with the change in the economy. The coincident indicator alerts companies and governments to whether they should make ad hoc changes in fiscal activities to achieve their long term economic goal.

The main objective of this thesis is to construct a coincident indicator for the Swedish economy using the Stock and Watson approach. To evaluate the adequacy of the model it is proposed to compare the new indicator with gross domestic product series and alternative indexes by calculating correlation between corresponding series in the forecast period. In order to measure the performance of the forecast it is proposed to calculate some measures of accuracy.

The paper has the following organization: section 2, contains a literature review with focus on concepts related to the business cycle, composite indicators and factor model techniques; in section 3, the focus is on data description and the theoretical foundations of the Stock and Watson approach. Results and the corresponding discussion are presented in section 4; finally the main conclusion is presented in section 5.

II LITERATURE REVIEW

2.1 Business cycle and composite indexes

According to Hassler et al (1992), Crone (2006), the popular notion of a business cycle was formulated by Burns and Mitchell (1946) with a focus on reference cycles and on different phases in a given sequence of a cycle. They identified four phases of the business cycle: an expansion followed by recession and contraction and then a revival of economic activity leading to the next expansion phase. These four phases are commonly collapsed into two periods: a period of growth (revival and expansion) and a period of decline in economic activity (recession and contraction).

The main problem with measures of economic activity in different countries is how to provide a good answer to the question, “How is the economy doing?”. According to Crone (2000), it is often not clear which measure to point to. For example if we use indexes such as the unemployment rate job growth to measure economic activity like the change in GDP we get different information about the state of economy and we can not understand the correct signal of the economy, that is the statistics for unemployment insurance, housing permits and others economic series gives specific information of the direction of the economy, which create conflict when we need to know no where we are. This problem can be solved by combining different measures into one index which is called composite index to measure the current or the future economic activity.

An indicator is anything that can be used to measure or predict future financial or economic trends. For example, the social and economic statistics published by institutions such as National Institute of Economic Research (NIER), Statistics Sweden, Swedish Agency for Economic and Regional Growth (AERG), Swedish Financial Supervisory Authority (FSA), publish popular indicators such as unemployment rates, housing starts, inflationary indexes and consumer confidence. The system of economic indicators is designed to signal impending changes in the direction of overall economic activity and to help analyze the short-term prospect of the economy. According to Fukuda and Takashi (2001) economic indicators which are useful for the study of cyclical expansions and contractions in business activities have been grouped in three categories, namely leading, coincident and lagging indicators which indicates the timing of their changes relative to how the economy as a whole changes.

Leading - these types of indicators signal future events. Leading indicators change before the economy changes. Stock market returns are a leading indicator, as the stock market usually begins to decline before the economy declines and they improve before the economy begins to pull out of recession. Leading economic indicators are the most important type for investors as they help predict what the economy will be like in the future. According to the Conference Board U.S. Business Cycle indicators review, September 2008, to predict economic trends, the leading indicator take consideration of factors like bond yields, building permits, money supply, production workweek, stock prices, unemployment insurance claims, long and short interest rate spreads, etc. On the other hand the leading indicators gives some incorrect signals for recession and the interval between the time when the information about recession is available and the corresponding occurrence of this event varied widely. That is, the leading indicator sometimes fails in prediction.

Coincident – for Stock and Watson (1989) the coincident index measure the current state of the economy. Gillitzer et al (2005) and Altissimo et al (2007) defines the coincident indicators as those which change at approximately the same time and in the same direction as the whole economy, they give information about the current state of the economy but they don't predict or confirm economic events. However, the coincident indicator can still be useful in confirming a particularly new economic movement or event in its first few weeks. According to Crone (2000) an example of a coincident indicator is personal income: high personal income rates will coincide with a strong economy. For Patnaik and Sharma (2002), "gross national income, real disposable income, real final sales, real manufacturing and trade sales and industrial production and employment", are coincident indicators. For instance an index for coincident indicators for U.S economic business cycle is "industrial production real personal income, manufacturing and trade sales and employment in nonagricultural".

Lagging - a lagging indicator is an economic indicator that reacts slowly to economic changes, and therefore has little predictive value. According to Ahmad (2008) lagging indicators indicate how well an economy has performed in the past few months. That is, they confirm the performances of cyclical moments by the coincident indicator, giving economists a chance to review their predictions and make better forecasts. Logging indicators takes consideration of factors like GDP, unemployment, labor costs, corporate profits, interest rates, etc. The importance of a lagging indicator is its ability to confirm that a pattern is occurring or about to occur for example if the unemployment rate is raising it indicates that the economy has been doing poorly. Unemployment is one of the most popular lagging indicators. Summarizing, it can be said that the leading indicators are designed to predict turning points especially recessions; coincident indicators are used to decide whether a turning point in the business cycle has been reached; while lagging indicators it is a tool used more to confirm the turning points peaks and troughs of the coincident indicator.

2.2 Comparison between the coincident indicator and GDP

For Gillitzer et al (2005), “the business cycle is measuring using gross domestic production or some average of individual economic series from the seminar presented by Burns and Mitchell (1946)”. GDP by definition measures the total output of the economy. Therefore we know that the GDP and many other economic series, is estimated with noise. Gillitzer et al (2005) and Altissimo et al (2007) using ideas of Stock and Watson (1989), suggests to use coincident indicator to measures the state of economy.

For economic decisions an indicator is needed, that is (a) easy to interpret and objective, that is take into account series from different category of economic variables like production, services and the public sector (b) available on time and (c) not affected by high oscillation of the some economic series. Because no available macroeconomic series provides a measure of the state of the economy that fulfills all such above criteria, for example GDP the most comprehensive indicator of real activity, fails to meet (b) and (c), that is, GDP is available only quarterly and with a long delay. Second Altissimo et al (2007), “the gross domestic production is affected by a sizeable short – run component so that, for example, the beginning of a medium - run upswing cannot be distinguished from a transitory upward movement within a basically negative path". While for composite coincident indicator that is a real-time estimate of GDP growth rate, are less affected by short-run oscillations. Gillitzer et al (2005) argue and underline that "a further advantage of coincident indicators is that they can be constructed with monthly data, and if they are produced on an ongoing basis they may be timelier than GDP because many economic series are published with a shorter lag than GDP". With this paragraph we can say that the coincident indicators need few revisions compared with GDP because it's constructed from series not revised or with smaller revisions.

2.3 Factor Model Techniques: SW and FHLR methodologies

In the late 1980s James Stock and Marc Watson cited by Crone (2000) developed an econometric model that estimates changes in the underlying "state of the economy". The "state of the economy" is unobservable but reflected in a number of indicators like production or personal income which are tracked by government agencies or private organizations. In the literature, according to Forni et al (2003) most multivariate forecasting methods are restricted to series of low dimension that is limited in number of variables. Such methods gives little help in large panels of time series, so the challenge for econometricians is to develop alternative techniques that are sufficiently powerful as to overcome this dimensionality problem, yet flexible enough to provide an adequate reality of the state of economy.

In order to solve these problem Stock and Watson (1999, 2002a, 2002b) and Forni et al (2000, 2001) have develop factor model techniques that are tailored to exploit a large dimension of variables. The main assumption for the factor model structure is that, each variable in the panel data is decomposed into two uncorrelated components: the common component which is "strongly correlated" with the rest of the panel and has reduced stochastic dimension, and the residual idiosyncratic component.

So that, we can distinguish essentially two standard methodologies in the literature used for construction and forecast the coincident indicator: the method proposed by Stock and Watson

(2002a, hereafter SW) and the method proposed by Forni, Hallin, Lippi and Reichlin (2001, hereafter FHLR).

To present the general procedure in SW and FHLR techniques lets to consider an ($n \times 1$) covariance stationary process $X_t = (x_{1t},..., x_{nt})'$. We are interested in forecasting some elements x_{it+h} of X by using all the variables $(x_{1t}, ..., x_{nt})'$ as predictors. The best linear forecast is defined by the following linear projection.

$$x_{it+h|t} = \text{proj}\{x_t \mid \Omega\} \tag{2.1}$$

"where $\Omega_t = \text{span}\{X_{t-p}, p = 0,1,2,...\}$ is a potentially large information set at time *t* when the forecasts are made and t = 1, ..., T", D'Agostino and Giannove (2006).

In practice when the number of variables in the panel data is very large equation (2.1) above is unfeasible because the projection requires the estimation of a large number of parameters with considerable degrees of freedom. One solution to this problem is to use few, q, series of common factors $F_t = \{F_{1t},...,F_{qt}\}$ where $q < n$. Here q represents the set of information in panel data taking account of all interactions among variables. Note that for q series there is a limited interaction among the variables, and the dimensionality problem is solved with this proceeding. According to D'Agostino and Giannove (2006) using the description above "the projection of x_{it+h} on the whole information set Ω_t is well approximated by the projection on the small information set including common factors and past values of the variables".

$$x_{it+h|t} = \text{proj}\{x_{it+h|t} \mid \Omega\} \approx \text{proj}\{x_{it+h|t} \mid \Omega_t^i\} \tag{2.2}$$

Where $\Omega_t^i = \text{span}\{F_t\} \cup \text{span}\{x_{it},\ x_{it-1}, ...,\}$ is the parsimonious representation of the information set that exploits the factor structure and i = 1, ..., n variables ; h = -1, 0, 1 forecast horizon.

According to Altissimo et al (2007), Eickmeier and Zeigler (2006) and Gillitzer et al (2005), most procedures assume that each series in the panel have an approximate factor representation, that is by the sum of two stationary, mutually orthogonal unobservable components: the common component, called χ_{it}, and the idiosyncratic component, ξ_{it}. Suppose $X_t = (x_{1t}, ..., x_{nt})'$ is the standardized variables, in addition if X_t is described by a factor model, each x_{it} can be written as the sum of components:

$$x_{it} = \chi_{it} + \xi_{it} \tag{2.3}$$

It is assumed that the small number q < n for common components of common shocks f_{it}, i = 1, ..., q, have the same information obtained from all variables in the set, but the q common factors are in general lagged with different coefficients and lag structure:

$$\chi_{it} = b_{1t}(L) f_{1t} + b_{2t}(L) f_{2t} + ... + b_{qt}(L) f_{qt} \tag{2.4}$$

For the second component "idiosyncratic component" it's assumed that the shocks are weakly correlated over time with each variable in the dataset. So we can write equation (2.3) as:

$$x_{it} = b_t(L) f_t + \xi_{it} = \left(B_{i0}, ..., B_{is}\right)\begin{pmatrix} f_t \\ ... \\ f_{t-s} \end{pmatrix} + b_t F_t + \xi_{it} \tag{2.5}$$

Where f_t is a $(q \times 1)$ vector on dynamic factors, $B(L) = B_0 + B_1 L + ... + B_s L^s$ is an $(n \times q)$ matrix of filters of finite length *s*, ξ_t is the $(n \times 1)$ vector of idiosyncratic component, F_t is the $(r \times 1)$ vector of the stacked factors with $r = q(s+1)$. From equation (2.3) where the basic model is represented we can note that the equation (2.3) "is a restricted version of the dynamic factor model" proposed by Forni, Halli, Lippi and Reichlin (2000), in vector notation the model become:

$$X_t = B(L) f_t + \xi_t = \left(B_0, ..., B_s\right) \times \begin{pmatrix} f_t \\ ... \\ f_{t-s} \end{pmatrix} + \xi_t = BF_t + \xi_t \tag{2.6}$$

Note that we continue to assume that the vector of dynamic factors (f_t) and idiosyncratic components are mutually orthogonal stationary process; we can thus be defined $\chi_t = B(L) f_t$ as the common component.

According to D'Agostino and Giannove (2006), Altissimo et al (2007) given the orthogonality assumption between common factors and idiosyncratic component, the spectral density matrix of

X_t at each frequency $\theta \in [-\pi, \ \pi]$ can be decomposed into the sum of the spectral densities of the common and the idiosyncratic component:

$$\sum(\theta) = \sum_{\chi}(\theta) + \sum_{\xi}(\theta) \tag{2.7}$$

The second term in the equation (2.7) is the spectral density matrix of χ_t calculated by $\Sigma_{\chi}(\theta) = B(e^{-\iota\theta})\Sigma_f(\theta)B(e^{-\iota\theta})'$ and the last term $\Sigma_{\xi}(\theta)$ is the corresponding spectral density matrix of the idiosyncratic component ξ_t. Using the sum structure, the covariance matrix of X_t can be decomposed as:

$$\Gamma_k = \Gamma_k^{\chi} + \Gamma_k^{\xi} \tag{2.8}$$

"Where $\Gamma_k^{\chi} = B\Gamma_k^F B'$, and Γ_k^F is the covariance matrix of F_t at lag k and Γ_k^{ξ} is the covariance matrix of idiosyncratic component at lag k", D'Agostino and Giannove (2006).

From equations (2.7) and (2.8) it is easy to verify that for the common component with spectral density $\Sigma x(\theta)$, the number of dynamic factors q is equal to the rank of the matrix while for covariance matrix Γ_k^{χ} the corresponding rank is equal to r static factors.

Using the factor model structure, we can forecast the ith variable h-head steps ahead by the sum of two components where each component is forecasted separately, that is we can start with forecasting the common component and forecast the idiosyncratic component then we sums the resultant series. The forecast considered here is true if the dynamic interaction among variables is captured by the common component and the idiosyncratic component can be obtained only from past value of the dependent variables.

$$\mathrm{x_{it+h|t}} \approx \mathrm{proj}\{\chi_{\mathrm{it+h}} \mid \mathrm{F_t}\} + \mathrm{proj}\{\xi_{\mathrm{it+h|t}} \mid \mathrm{x_{it}}, \mathrm{x_{it\text{-}1}}, ...\} \tag{2.9}$$

The equation (2.9), describe two forecast components, again the sum is not feasible because the common factors are unobserved. However, by factor model techniques this is not a hard problem, thus using dynamic factor model the common factors F_t can be consistently estimated by appropriate averages. Building on Chamberlain and Rothschild (1983), Forni, Hallin, Lippi and Reichlin (2000) and Stock and Watson (2002a), the appropriate average is the principal component of the normalized observations of the variables. Its important to distingue that, SW use r principal components and FHLR use q dynamic principal components to approximate the common component. From equation (2.8) and information reported above, consistent estimator of the auto covariance matrix of standardized data $\hat{X}_t = (\hat{x}_{it},...,\hat{x}_{nt})'$ is given by equation (2.10)

$$\hat{\Gamma}_k = \frac{1}{T-k-1}\sum_{t=k}^{T}\hat{X}_t\,\hat{X}'_{t-k} \tag{2.10}$$

And the correspondent spectral densities matrix will be estimated by averaging a given number *m* of auto covariance's:

$$\Sigma(\theta) = \frac{1}{2\pi}\sum_{k=-m}^{m} w_k \Gamma_k e^{-i\theta k} \tag{2.11}$$

Where w_k are weights satisfying the conditions: w(0)=1 and $0 \le w(k) \le 1,\ \forall k \le m$. Using this weights the estimates of the spectral density are consistent taking account that $m \to \infty$ and $\frac{m}{T} \to 0$ as $T \to \infty$. In the literature it is suggested to use $m = \sqrt{T}$, which satisfies the asymptotic requirements above.

These techniques have been used to construct coincident indicators such as: The Chicago Fed National Activity Index – CFNAI, which is index of economic activity developed by Stock and Watson in 1999 for U.S. and Eurocoin index for Europe area published by the Centre for Economic Policy Research - CEPR. Starting at the 2000s the number of countries that use the new techniques to construct their composite indicators instead of traditional methods growth, for example Fukuda and Takashi (2001) used the Stock and Watson approach to construct a new composite index of coincident economic indicators in Japan, Hall and Zonzilos (2003), applied SW approach to estimate composite indicators for Greece, Gaudreault et al (2003) applied Stock and Watson methodology for construct the new coincident, leading and recession index for the Canadian economy, Gillitzer et al (2005) used the SW and FHLR methodologies for construct a coincident indicator for Australia business cycle.

The use of dynamic factor models is increasing when modeling economic indicators; this increase is due to the advantage of the methods to exploit information for a large panel data of variables. According to Forni et al (2001), Stock and Watson (1999) and Quah and Sargent (1993), the extracted unobserved economic indicator from large number of variables capture the changes of economic direction, and by Kubandi (2004), the main assumption for factor model techniques is that the common shocks of the large economic variables can be represented by a small number of common unobserved factors.

Altissimo et al (2001) and Gillitzer el al (2005) classify the models above according to the specification used when computing the common component in static and dynamic factor models. That is the Stock and Watson (2002a) model 'known as static factor model $X_t = BF_t + \xi_t$ and model proposed by Forni, Hallin, Lippi and Reichlin (2005) is said dynamic factor model with representation $X_t = B(L)f_t + \xi_t$. From the above classification F_t is called a vector of static factors while f_t is a vector of dynamic factors. However, there are several variants derived as combinations of these two techniques or by generalization of each method separately which propose to get better performance in appropriate studies.

According to D'Agostino and Giannove (2006) and several other authors reported in the literature the SW and FHLR methods essentially differ in computation and forecast of the common component. Basically differences are founded in the estimation of the factor space and in the form of how the projection on this space is performed. For Stock and Watson the common factors are extracted by standard principal components (PC) of the covariance matrix and the projection of predicted variable on the factors is the forecast of the common component. This proceeding differ with FHLR because for the last authors they propose efficiency improvements by starting with estimation of common factors using generalized principal components (GPC), second they consider constraint's implied by the dynamic factor structure directly on the variable of interest. We can underline that the use of GPC where observations are weighted according to their signal to noise ratio and taking account of the relation between leads and lags by means of principal components in the frequency domain is the main gain of this procedure.

For econometricians it is not clear which method performs better, for example Stock and Watson (2004a) using large dataset of U.S. macroeconomic variables conclude that the methods perform similarly; here its necessary to empathize that the mentioned authors focus only on the weighting of the idiosyncratic component. From investigators like Boivin and Ng(2005) the “SW’s method largely outperforms the FHLR” and using this evidence it can be conjectured that the dynamic restriction implied by the latter method are harmful for the forecast accuracy of the model. Schumacher (2006) with German macroeconomic variables, using forecast results he conclude that FHLR generally outperforms SW’s. D’Agostino and Giannone (2006) find that, both approaches outperform the simple univariate autoregressive model, that is the gain from factor based predictions is substantial, especially in periods of high co-movements, and few factors capture all predictable common components while idiosyncratic dynamic component are negligible. Finally they find also that when factors are estimated by putting less weight to series with large idiosyncratic variance, there is no evidence of improvement in the forecast accuracy.

Summarizing this comparison of performance of the two factor models D’Agostino and Giannove (2006) affirm that, there are three main differences between the SW and FHLR methods. (i) The weighting scheme adopted is different: SW uses standard PC to extract the common factors while FHLR introduce new estimator and use DPC for computing the common factors. (ii) The estimator used in projection on the common factors is also different: SW use ordinary least squares (OLS) while FHLR use a non-parametric method that takes account of restrictions implied by the dynamic factors structure by using dynamic least squares (DLS). (iii) The last difference is the method used to forecast the idiosyncratic component: SW adds lags of the dependent variables in the equation, while FHLR forecasts the two components separately under conditions of orthogonality between common and idiosyncratic components.

In order to develop this study, I suggest to combine the Stock and Watson methodologies and dynamic factor models, thus it attempt to use principal components to extract the factors and in other hand can be consider a restrict dynamic factor model in which the factors are dynamic but the relation between the dynamic factor and the observable variables is static. Moreover in practice for large dimension the generalized principal component is unstable. According to D’Agostino and Giannove (2006), static principal components can be use to overcome this problem.

III MATERIAL AND METHODS

3.1 Data description

Special attention is necessary during the phase of data collection, this care is very important when we propose to use factor model techniques for estimation of the new indicators. For example, if there is no equilibrium in number of series from different categories of macroeconomic variables in the panel data, that is, the panel is dominated by some categories of economy activity; the new indicator tends to be closer to this side of categories instead of the overall economy. Another important point when compiling data is to take care to avoid many similar series and on the other hand ensuring that, as far as possible, series from different categories are include.

The data set used for estimating the Coincident Indicator for the Swedish Economy is divided into two subsets: monthly data series and quarterly data series; that is, I use two panels in this study. All series of economic variables was obtained from the statistical database of Statistics Sweden and from National Institute of Economic Research (NIER). The Statistics Sweden website contains different categories of variables such as business activities, financial markets, labour markets, public finances and so on. For each category several economic series are available with different reference dates, period where official collection starts, etc. However, it was not possible to use all time series available in this study because some of them have too short histories or the series have too many missing observations, also we exclude some series to avoid duplicate and finally we exclude others series in order to have a balanced panel.

In practice, constructing a new indicator using a small number of variables does not imply less accurate estimates, Boivin and Ng (2006) "argue that adding additional series to a panel need not improve the factor estimates if the additional series are noisy or have correlated errors". In earlier studies some large panels have been obtained by disaggregation of series into corresponding components, for example exports of different products or classified by sector activity. If the disaggregated series are likely to contain more idiosyncratic noise and to have positive correlation with the idiosyncratic component, Boivin and Ng conclude that, r factors extracted from small panel with 40 series sometimes produce more accurate forecast compared with factors derived from a large panel of 147 series. Watson (2001) also finds that “the marginal improvement in forecasting performance from using greater than 50 series is very small” and by Inklaar et al (2003), using factor models methods we can construct any indicators closer to the Eurocoin index taking few number of variables from a large panel.

The data is organized on two panels samples: panel 1990:01 – 2008:12 with monthly series corresponding the period January 1990 to December 2008 and panel 1993:Q1– 2008:Q4 with quarterly series from first quarter 1993 to fourth quarter 2008; and the indicator is constructed from balanced panels containing 68 monthly series reported in Table 3.1 and 93 quarterly series reported in Table 3.2. The macroeconomic time series is grouped into different categories according to Statistics Sweden database and National Institute of Economic Research. Many of the used time series variables report information relating to business activity, financial market, housing construction and building, labour market, national accounts including gross domestic production, prices and consumption, public finances, trade in goods and services and so on.

Table 3.1 Number of monthly series in each category of economic activity

Categories	Monthly 1990:01– 2008:12
Business activities	9
Energy	6
Financial markets	13
Labour market	3
Prices and consumption	20
Public finances	5
Trade in goods and services	6
Transport and communication	6
Total	68

Most original series for this type of study are not stationary; so the first activity is the transformation of the series to stationary series. In general according to Altissimo et al (2007), before making any analysis of time series we should transform the time series to remove outliers, seasonal factors and non – stationary, this is usual preliminary activity

Table 3.2: Number of quarterly series in each category of economic activity

Categories	Quarterly 1993:Q1–2008:Q4
Business activities	7
Housing construction and building	8
Labour market	13
National accounts	34
Prices and consumption	9
Trade in goods and services	22
Total	93

If the series contains some outlier's observations, it's necessary to remove these observations and replacing them with the average of the remaining observations. The criteria adopted to detect outlier observation vary, here we can emphasize one and define an observation as outlier if this point is more than 5 standard deviations away from the mean. Note that the number 5 is not fixed; however, we consider this a reasonable choice for this study. The seasonal adjustment can be done by regressing each time series on dummy variables; to obtain stationary series, we can often use regular differences of the logarithm of the series; In order to support the hypothesis of regular differences in the series the Augmented Dickey and Fuller (ADF) unit root tests can be used; One important transformation is the normalization of the series, which consists of subtracting the mean and dividing by standard deviation: $z = (x - \bar{x}) / s_x$. This standardization is necessary to avoid over weighting series with large variance when estimating the spectral density.

In order to process the structure of factor model suggested by Stock and Wilson (2002a) and consequent estimation of coincident indicator its assumed that the data have the most important requirements necessary for an analysis using factor representation and the transformations conserve these requirements, because its known from D'Agostino and Giannone (2006), the main motivation behind the factor representation structure is the strong co-movement observed in macroeconomic time series, which is possible only if there are few underlying common driving factors and the simplest statistics to describe co-movements among series is the percentage of the variance of the panel accounted for by common factors estimated by principal components. For panel data with strong co-movements, a small number of principal components is enough to explain a relevant percentage of the variation in the panel, while cases where panel data have relatively less co-movement among variables more principal components are needed to explain the relevant percentage of variability. In both cases the adequate number of principal components is denoted by r such that the remaining (r+1), (r+2) and so on principal components have small marginal contribution.

3.2 The Stock and Watson approach

Starting from equation (2.3) where each standardized variable of X_i is described by the factor model, and rewrite the equation has the sum of common component and idiosyncratic component as static factor model gives:

$$x_{it} = \chi_{it} + \xi_{it} \quad or \quad X_t = BF_t + \xi_t \tag{3.1}$$

Following D'Agostino and Giannove (2006), Eickmeier and Ziegler (2006) to forecast using the method proposed by Stock and Watson it is necessary to first estimate the common factors from principal components which are calculated from standardized data variables. Obviously the common factors serve as predictor's. According to Patnaik and Sharma (2002), "the principal components analysis is statistical technique that linearly transforms an original set of variables into a substantially smaller set of uncorrelated variables that represents most of the information in the original set of variables". By definition the set of PC's are uncorrelated, easy to interpret and to use for future analysis than working with large set of correlated variables. This advantage of factor model technique helps researches in the construction of a proxy for a large set of variables. As explained by Patnaik and Sharma (2002), "a mathematical process resolves the data on all components of the group into a set of linear combinations of weighted indices, where each weighted accounting for a stated proportion of the total variation in the data set. The index chosen explains the highest proportion of this variance. This suggests that it is the most convincing surrogate or index, for all the constituent variables".

The equation (3.2) is typically known as an eigenvalue problem for the sample covariance matrix.

$$\hat{\Gamma}_o V_r = V_r D_r \tag{3.2}$$

Where $\hat{\Gamma}_o$ is the sample covariance matrix (equation 2.10 for k = 0); The diagonal matrix contains only the first r largest eigenvalue is denoted by D_r; the matrix of the correspondent eigenvector is define by $V_r = [v_1, ..., v_r]$. Note that, the matrix D_r is calculated from the sample covariance matrix and the dimension of V_r is $(n \times r)$. Using matrixes V_r and X_t we can obtain the first r common factors (equation 3.3):

$$\hat{F}_t = V'_r \hat{X}_t \tag{3.3}$$

The covariance matrix of the common component is estimated by:

$$\Gamma_o = V_r D_r V_r^{'} \tag{3.4}$$

Note that the first r principal components define the number of common factors that should be use to estimate the common component. Another stage in the SW procedure is the determination of adequate number of common factors, precisely number of PC. This problem can be solved by applying some criteria's used in the factor model literature, for this study we can focus on four criteria's: (1) the Kaiser Criteria defined by Kaiser (1960), he propose first to retain only factors with eigenvalues greater than one. In essence this is like saying that, unless a factor extracts at least as much as the equivalent of one original variable; (2) Johnson and Wichern (2007), suggests that a useful technique to determine an appropriate number of components is the scree plot test proposed by Cattell (1966) where the number of PC to take correspond to the point at which the remaining eigenvalues are relatively small and all about the same size; (3) Authors like Altissimo et al (2001), Forni et al (2000, 2001) and Inklaar et al (2003) has used the marginal explanatory power of each factor included in the model; in practice we can look for the percentage increased on R-square by adding one more factor. Based on experience with different datasets the authors suggest to exclude all factors from the point where the marginal contribution is less than 5 percent. (4) The criteria of maximum variability explained by the r PC should be 80 or 90 percent.

According to Forni et al (2005) and D'Agostino and Giannove (2006), if the data follow an approximate dynamic factor models, $\hat{F}_t$ are consistent estimates of the unknown common factors. From equation (3.3) the common factors is used in equation (3.5). Specifically we rewrite equation (2.2) as:

$$\hat{x}_{it+h|t} = \text{proj}\{x_{it+h} \mid \Omega_t^i\} \tag{3.5}$$

Where $\Omega_t^i = \text{span}\{\hat{F}_t\} \cup \text{span}\{x_{it}, x_{it-1}, ...,\}$, this approach is implemented through the following forecasting equation.

$$\hat{x}^{PC}_{it+h} = \alpha_{ih} + \sum_{i=1}^{r} \beta_{ih} \hat{F}_t + \sum_{i=1}^{s} \gamma_{it}(L) x_{it} + \xi_{it+h} \tag{3.6}$$

In equation (3.6), *i* indicate the order of variables, *h* is the forecast horizon and superscript *PC* denote principal components (= weights are extracted using principal component analysis). The lag polynomial $\gamma_{ih}(L)$ is of length *s* determined using the relation between static and dynamic representation: $r = q(s+1)$, where r is the number of static factors, s the number of lags on dependent variable and ξ_{it+h} is an error term. As explained in previews paragraphs the SW approach assumes that the common component is captured by common factors and idiosyncratic component is captured by lagged values of the dependent variable. Denoting by $\hat{\alpha}_{ih}$, $\hat{\beta}_{ih}$ and γ_{it} the OLS estimates, and setting h = 0 corresponding to forecasting the current forecast value in horizon H, can be forecast $\hat{x}^{PC}_{iT+h|T}$ successively for period T using equation (3.7). The coincident indicator is the series constituted by all current forecast value for h = 0.

$$\hat{x}^{PC}_{iT+h|T} = \hat{\alpha}_{ih} + \sum_{i=1}^{r} \hat{\beta}_{ih} \hat{F}_T + \sum_{i=1}^{s} \hat{\gamma}_{it}(L) x_{iT} \tag{3.7}$$

Note that in this approach the factor structure assumption is exploited only for the extraction of the common factors. The equation (3.6) does not take account of the restrictions implied by the dynamic factor structure, furthermore the OLS do not use orthogonal assumption between common and idiosyncratic component. For comparison, consider the simple autoregressive model AR(p) process which replace the second term in equation (3.7) with residuals.

According to D'Agostino and Giannove (2006), the forecast equation (3.7) using as predictors the common factors and lags for dependent variables, is know as Diffusion index – Autoregressive (DI_AR) forecast. When the forecasting depend only on common factors, that is, equation (3.7) without lags on the dependent variable the equation is called Diffusion index (DI) forecast. According to Stock and Watson (2002b), both forecasts in most cases perform better.

3.3 Goodness fit and forecast performance

When forecasting a time series, the precision of model fit and forecast should be monitored. For this we be consider two types of measurement: the measures for evaluating the goodness of fit typically used as in- sample and the measures of forecast performance that is out-of-sample.

Goodness of fit: A commonly used statistics to measure the *goodness of fit* of a stationary model is R-squared. By Tsay (2005) for stationary time series model with T observations the measure is defined as:

$$R^2 = 1 - \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{\sum_{t=p+1}^{T} (y_t - \bar{y})^2} \text{ ; where } a_t \text{ is the residuals} \tag{3.8}$$

Typically $0 \le R^2 \le 1$ and a larger R^2 indicates that the model provides a closer fit to the data. However, this is only true for stationary time series. For a given data set, it is well – known that R^2 is a no decreasing function of the number of parameters used. To overcome this weakness, an adjusted R^2 is proposed which is defined as

$$Adj\_R^2 = 1 - \frac{\hat{\text{var}}(a_t)}{\hat{\text{var}}(y_t)} = 1 - \frac{\hat{\sigma}_a^2}{\hat{\sigma}_y^2} \tag{3.9}$$

This new measure takes into account the number of parameters used in the fitted model. However it's no longer between 0 and 1.

Forecast performance: the criteria's used for evaluation are the Root Mean Squared Forecast Error (RMSFE), the Mean Absolute Error (MAE) and the Theil's U statistic which corresponding the component bias in the forecast. The RMSFE is one of methods which measure the deviation of the simulated series from the actual values. Following Pindyck and Rubinfeld (1998), this measure of forecast quality is defined as:

$$RMSFE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(y_{t+h} - \hat{y}_t(h)\right)^2} = \sqrt{bias^2 + \mathrm{var}(e_t(h))} \tag{3.10}$$

Where $\hat{y}_t(h)$ is the forecasted value at time t, y_{t+h} is the actual value at time t, and T is the sample size. The bias in second part of equation (3.10) can be determined by Theil's U statistic (3.11):

$$U = \frac{\left(\bar{y}_{t+h} - \bar{y}_t(h)\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\left(y_{t+h} - \hat{y}_t(h)\right)^2} \quad ; \quad 0 \le U \le 1 \tag{3.11}$$

The value of U is an indication of systematic error, since it measures the extent to which the average of the forecasted and actual series deviates from each other. Summarizing the model with a lower value of RMSFE and U has a higher forecasting performance than the other while a large values means that a systematic bias is persistent indicating revision of the model is necessary. The Mean Absolute Error (MAE) measures the average magnitude of the errors in a set of forecasts without considering their direction. The MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. The MAE is a linear score which means that all the individual differences are weighted equally in the average.

IV RESULTS AND DISCUSSION

4.1 Data analysis and characteristics

(a) Release calendar of time series

It was noted in the introduction that GDP fails to meet timeliness criterion because this indicator is produced with delay, in this section are for illustration focuses on some features of data used in this paper. The release date for each time series was collected from Statistical Sweden's website. For simplicity we consider the information required for December 2008 for monthly data and fourth quarter 2008 for quarterly data. In Tables 4.1 and 4.2, the release date according to Statistics Sweden database is reported for some macroeconomic time series.

Table 4.1 Calendar for some monthly macroeconomic time series

Macroeconomic time series (information required for Dec 2008, Jan 2009 and Feb 2009)	Release date		
	December 08	January 09	February 09
Activity index	Feb 10	Mar 10	Apr 09
Consumer price index (CPI)	Jan 13	Feb 13	Mar 12
Financial market statistics	Jan 29	Feb 26	Mar 26
Industrial production index (IPI)	Feb 10	Mar 10	-
Labour cost index, wages and salaries employees	Mar 02	-	-
Labour force survey (LFS)	Jan 22	Feb 12	Mar 19
Net trade in goods	Jan 26	Feb 26	Mar 25
New orders and deliveries in industry	Feb 10	Mar 10	-
New registration of passenger cars and lories	Jan 02	Feb 02	Mar 03
Producer and import price index	Jan 27	Feb 27	Mar 24
Real estate prices	Jan 14	Feb 16	Mar 16
Retail sales	Jan 28	Feb 27	Mar 27
Services production index	Feb 17	Mar 06	-
Shareholders statistics	Feb 26	-	-

From Table 4.1 we can see that, the release delay for the Statistics Sweden data vary between one to two months; New registration of passenger cars, Consumer price index, Real estate prices variables are the most timely data, while Labour cost index, Shareholders statistics and others are usually available with a 2 – 3 months. For quarterly data, the information is usually available with long delays; the Table 4.2 shows release dates of data for information regarding the fourth quarter 2008. Note that, there are not series published in January 2008 and few series are available from February 2009.

Table 4.2 Calendar for some quarterly macroeconomic time series

Macroeconomic time series (information required for 4th quarter 2008)	Release date	
Industrial capacity utilization	Feb 13	-
Labour force survey	Feb 12	-
National accounts	Feb 27	-
Service production price index	Feb 13	-
Short term business statistics on sick pay	-	March 03
Stocks in industry	Feb 13	-
The Swedish economy – statistical perspective	-	March 17
Trade in volume of exports and imports of good	-	March 05

Tables 4.1 and 4.2 illustrate that, some of quarterly time series is published with long delays around 60 or more days, which support that the construction of any composite indicator with monthly data compared with quarterly index where, for example the GDP and other series are available with long delays. The different release dates of the variables in the panel data set, makes it difficult to interpret the composite indicator, because the reference date for the new indicator is unclear. One way to avoid this problem of interpretation is to use lags of the variables in order to take account of the information available in the reference period. For this study we consider a coincident indicator is constructed for December 2008 and fourth quarter 2008. In order to taking account of the different release dates, we use 1, 2 and 3 lags for data available in January, February and March for monthly panel and one lag for data available at the first quarter 2009 for quarterly panel, note that we updating also our dependent variable.

(b) Data transformations

In chapter three the need to work with stationary series was explained, so for the monthly panel 1990:01 – 2008:04 with 68 economic times series, to be stationary, 40 series was transformed using first differences on logarithm, 23 series by first difference and 5 series taking logarithm. From the transformations of the series we can observe that the data set is dominated by unit root series; to verify this idea the Augmented Dickey Fuller (ADF) unit root tests are carried out using auxiliary regression with p lags and a constant term. The results of the unit root test confirm that the data are dominated by stochastic trend so the hypothesis to proceed with first differences seems to be valid and supported by the ADF test. Applying the procedure presented in section 3.1, all series are normalized. For detail you can see the list of variables in appendix A, Table A1.

For quarterly data, panel 1993:Q1 – 2008:Q4 with 93 economic times series, 67 series was transformed using first differences on logarithm series, 17 series by logarithm and taking second differences, 6 by first difference and 3 series taking only logarithm. The results of Augmented Dickey Fuller (ADF) unit root tests confirm that it was necessary to first difference the series. Finally all series were normalized using the same procedure applied for monthly data.

4.2 Monthly Coincident Indicator

From the set of 68 variables, 67 principal components explain total variability of the panel data; that tells us that each component explain a small fraction of total variability, that means that the series are not characterized by strong co - movement in the panel data, which suggests more components to explain the relevant percentage of variation of the panel data. Using the criteria described in section 3.2, for identification of the number of principal components r (common factors) by requiring a maximum amount of variability of panel data explained by the set of common factors, r is found equal to 18 PCs. In this set of principal components the first one explain 11.31 percent of the total variance of the panel of 68 series, two PCs explain 22.19 percent; six PCs explain more than 50 percent precisely 50.96 percent; the last 18th principal component explain 1.49 percent of total variation. The set of 18 PCs explain 80.65 percent of total variation of the data which is more than 80 percent (criteria 4). The remaining eigenvalues are less than one in magnitude, (criteria 1) moreover increasing the number of components to 19 the percentage increase less than 5 percent of explanation of panel variability (criteria 3).

Before estimating the coincident indicator a reference variable is needed. Here monthly GDP is used as a reference variable because it is a single proxy of overall economic activity. Among different alternatives, the following procedure is used to transform quarterly into monthly GDP. First takes average of quarterly GDP by dividing by three, take log for the monthly series and transform to the monthly year-on-year growth rate using the following expression

$Y_t = \ln(GDP_t/3) - \ln(GDP_{t-12}/3)$, second fitting a univariate ARMA to yearly growth rate series with lags and checking the residuals an AR(4) model was selected, without autocorrelations in residuals, that is pass the Ljung-Box test for residuals. The estimates using AR(4) model is presented in Table 4.3.

Table 4.3 AR results for monthly year-on-year GDP growth rate equation

Dependent variable: Y_t = monthly year-on-year GDP growth rate			
Variable	Estimate	Standard error	t- ratio
Constant	0.0160	0.0110	1.47
Y_{t-1}	0.9720	0.3340	2.91
Y_{t-2}	0.0001	0.4720	0.00
Y_{t-3}	- 0.1546	0.3360	- 0.46
Y_{t-4}	0.1162	0.0430	2.71
sigma	0.0097	0.0002	47.66

The in sample fitted values are then used as estimates of monthly year-on-year GDP growth rate. Figure 4.1 show the approximate monthly year-on-year GDP grown rate (y_t) which plays a key identifying role in the compilation of the indicator because it is a reference variable that can be used to evaluate the relative weight of all other variables in the determination of the coincident indicator.

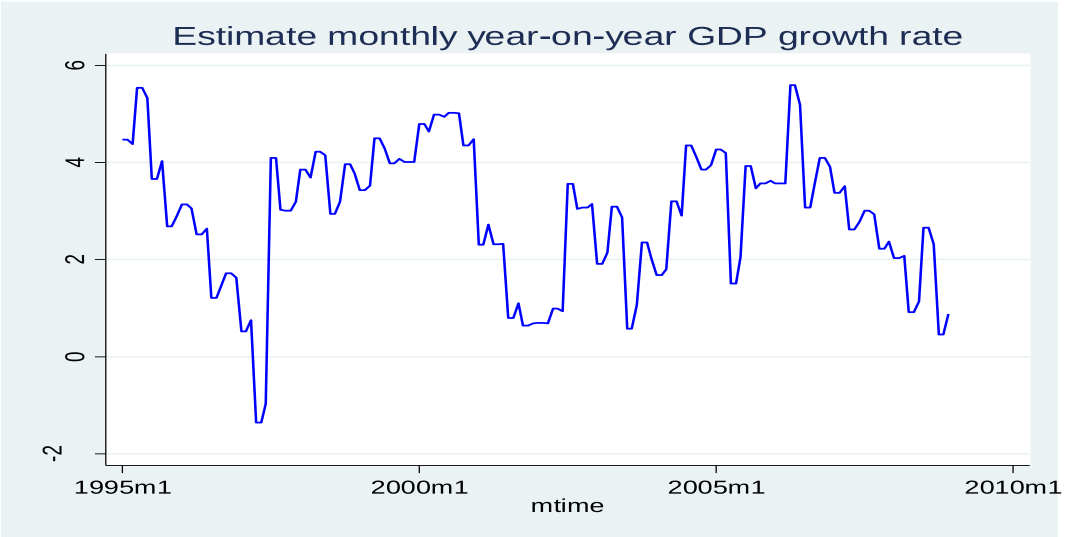

Figure 4.1 Estimated monthly year-on-year GDP growth rate

From equation (3.3) the corresponding 18 common factors was calculated; by relation r = q(s+1), the possible number of lags for dependent variable s vary between 1 and 17. The goodness of fit, i.e, the adjusted R-square of the regression of GDP growth rate on 18 common factors adding dependent variable with s equal to 1, 2, 3, 4, 5, 8 and 17 is 0.83, 0.83, 0.84, 0.84, 0.84, 0.83 and 0.85 for s equal to 17. Checking for residuals autocorrelation using the Ljung - Box test statistics, 3 lags for the dependent variable was selected. Specifically the estimated model using the time period 1990:01 to 2008:12 is reported in Table 4.4.

Table 4.4 OLS results for monthly GDP growth rate equation

Dependent variable: y_t = estimate monthly year-on-year GDP growth rate			
Variable	Estimate	Standard error	t - ratio
F_1	0.002	0.021	0.07
F_2	0.011	0.022	0.51
F_3	0.019	0.023	0.83
F_4	0.012	0.025	0.51
F_5	0.003	0.027	0.10
F_6	0.063	0.031	2.02
F_7	-0.005	0.034	- 0.14
F_8	0.054	0.037	1.46
F_9	0.022	0.039	0.55
F_{10}	-0.124	0.042	-2.95
F_{11}	-0.045	0.042	-1.06
F_{12}	0.078	0.046	1.68
F_{13}	0.079	0.049	1.61
F_{14}	0.008	0.052	0.16
F_{15}	-0.065	0.053	-1.23
F_{16}	-0.113	0.054	-2.09
F_{17}	-0.012	0.055	-0.21
F_{18}	0.001	0.057	0.02
y_{t-1}	0.967	0.068	14.05
y_{t-2}	-0.243	0.094	-2.59
y_{t-3}	0.166	0.067	2.45
constant	0.241	0.091	2.66

The forecast period is 1995:01 to 2008:12 with horizon 4 observations. That is, starting on point 1995:01 was calculating forecasts for horizon 1-4 and continues in the same way by adding one observation at the time and re - estimate until the end of the data. The composite coincident indicator series was obtained by taking the first forecast value for each forecast horizon; Figure 4.2 shows the constructed coincident indicator.

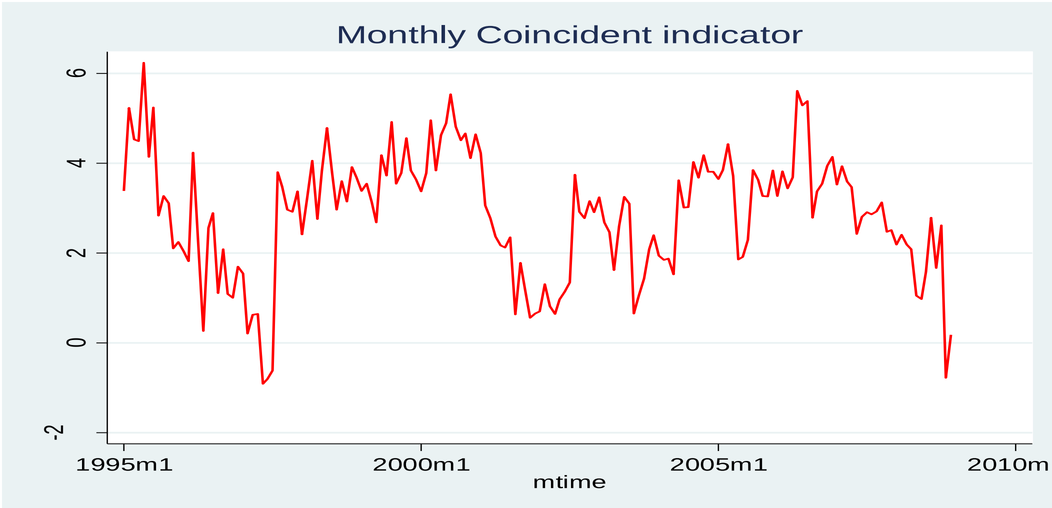

Figure 4.2 Monthly Coincident indicator

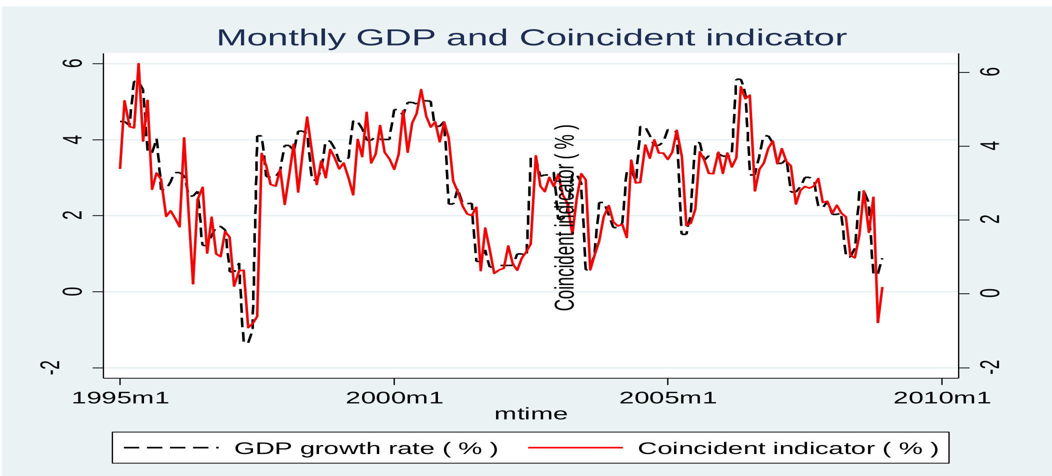

Figure 4.3 Estimated monthly GDP and Coincident indicator

The constructed monthly CCI is compared with estimated monthly year-on-year GDP growth rate and with the quarterly GDP growth rate. In order to compare with estimate monthly year-on-year GDP growth rate, Figure 4.3 shows the composite coincident indicator (solid line) along with the GDP growth rate (dashed line); To compare with the quarterly GDP growth rate the sum of the CCI over the three months in each quarter is calculated and the series from 1996:Q1 to

2008:Q4 for comparison (see Figure 4.4). The patterns are quite similar in both Figures 4.3 and 4.4, from these figures it can be seen that, the major part of fluctuations in estimated monthly GDP (reference variable) and quarterly GDP is captured by the constructed indicator, indicating that the general message of CCI, estimated GDP and quarterly GDP is essentially the same. So the CCI report in summary co - movement of all economic time series and gives us the state of the economy as a whole.

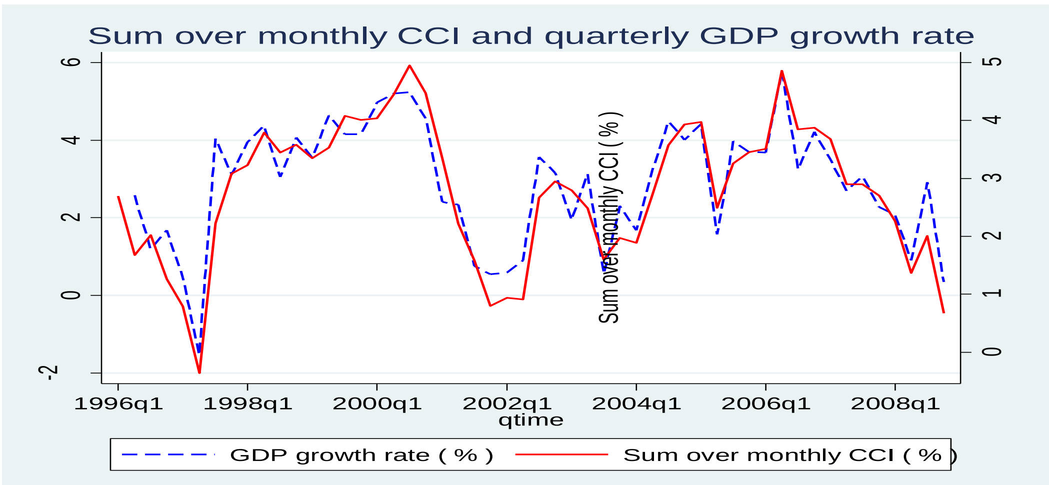

Figure 4.4 Comparison between sum over monthly CCI and quarterly GDP

Analyzing in detail some statistics related to the constructed index, it can be underlined that the actual index have positive high relationship with GDP (the correlation is equal to 0.80), moreover in most cases it correctly signals whether estimated monthly GDP growth rate is increasing or decreasing, (98 percent correct signal), which confirm that the coincident indicator move up or down directly with the change in the economy; or in another words the CCI moves in the same direction as the state of economy. Computation shows that the goodness of fit reported by adjusted R - square is around 84 percent indicate that the model has closer fit of the data in-sample or in the estimate period. Regarding to the forecast performance for the forecast period with T= 168 observation, the root mean square forecast error (RMSFE) is 0.88 percentage point, that is the average error in the forecast of the monthly GDP growth rate is about 0.9 percentage point and the mean absolute error (MAE) is 0.58 percentage point indicate that the model gives good performance; the bias of estimation is 0.01 as measured by Theil's U statistic, which is practically zero bias indicate that the model is free from systematic errors. Statistics from comparison with quarterly GDP shows that the transformed monthly CCI is closer to quarterly GDP with correlation equal to 0.93, and new indicator predict 99% of correct signal. The RMSFE and MAE is 0.64 and 0.52 percentage point respectively indicating good performance.

4.3 Quarterly Coincident Indicator

From the set of 93 economic time series, 62 principal components explain all 100 per cent of the panel data variability. Using the selection criteria explained in section 3.2 14 is the adequate number of principal components. Precisely 14 components explain more than 80 percent of the panel variability (82.05%), the first component explain 24.35 per cent; two PCs explain 39.19 percent; five PCs explain 59.89 percent, more than half of panel variability and the last 14th component explain only 1.63 per cent or increases in 1.63 percent of total variability of the panel data. One observation here is that, it was difficult to match all selection criteria; so that the number of PC fails to satisfy the first criteria mentioned in section 3.2 which requiring to consider first all components with eigenvalues greater than one, however the actual number of principal components satisfy the remaining three criteria's. For estimation of quarterly coincident indicator the quarterly GDP growth rate is used as a reference variable: $y_t = \ln GDP_t - \ln GDP_{t-4}$, see Figure 4.5.

Figure 4.5 Quarterly GDP growth rate

Using equation (3.3) the normalized series are combined as 14 weighted indexes generated by principal component analysis to obtain the 14 common factors (predictors); by the relation r = q(s+1), the possible number of lags for dependent variable s vary between 1 to 13. The regression of GDP growth rate on 14 common factors adding lags for dependent variable with s equal to 1, 2, 3, 4, 6 and 13; after trying different lags for s, the Ljung –Box Q Statistics test select s = 3, which is used for model estimation. The estimates for the quarterly GDP growth rate equation are reported in Table 4.5.

Table 4.5 OLS results for quarterly GDP growth rate equation

Dependent variable: y_t = quarterly GDP growth rate			
Variable	Estimate	Standard error	t - ratio
F_1	-0.006	0.032	-0.17
F_2	0.116	0.040	2.86
F_3	0.223	0.057	3.88
F_4	0.204	0.069	2.97
F_5	-0.005	0.107	-0.05
F_6	-0.121	0.086	-1.40
F_7	0.261	0.098	2.66
F_8	0.162	0.088	1.85
F_9	-0.179	0.127	-1.40
F_{10}	0.244	0.158	1.54
F_{11}	0.138	0.106	1.31
F_{12}	-0.012	0.161	-0.07
F_{13}	0.270	0.119	2.27
F_{14}	0.017	0.121	0.14
y_{t-1}	0.376	0.122	3.07
y_{t-2}	0.218	0.126	1.73
y_{t-3}	-0.065	0.121	-0.54
constant	1.217	0.421	2.89

The goodness of fit of the quarterly indicator that is adjusted R-square is 0.58 indicating close fit to the data; the estimation time period for the quarterly panel is 1993:Q1 to 2008:Q4 and the forecast period is 1996:Q1 to 2008:Q4, the constructed composite coincident indicator is shown in Figure 4.6.

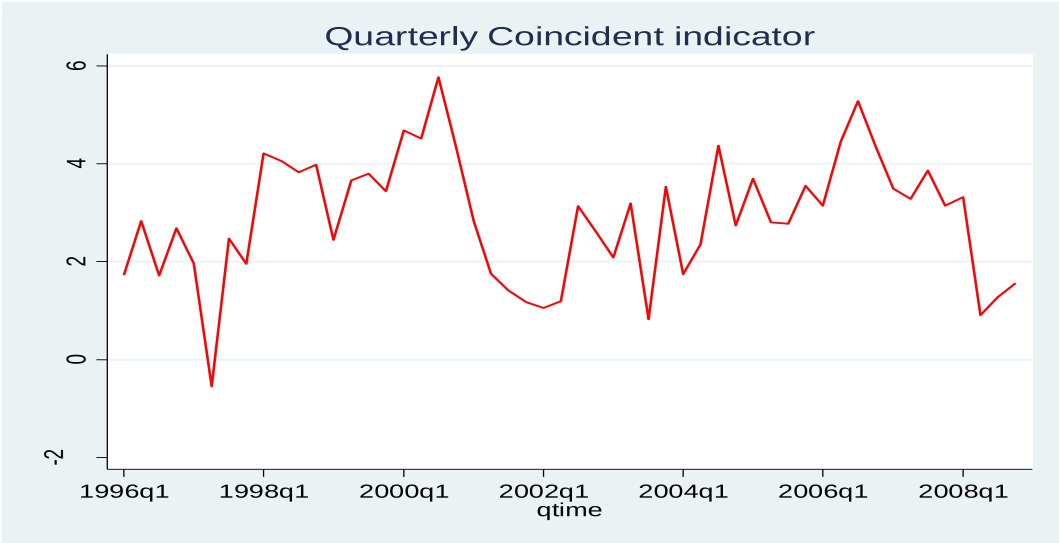

Figure 4.6 Quarterly Coincident indicator

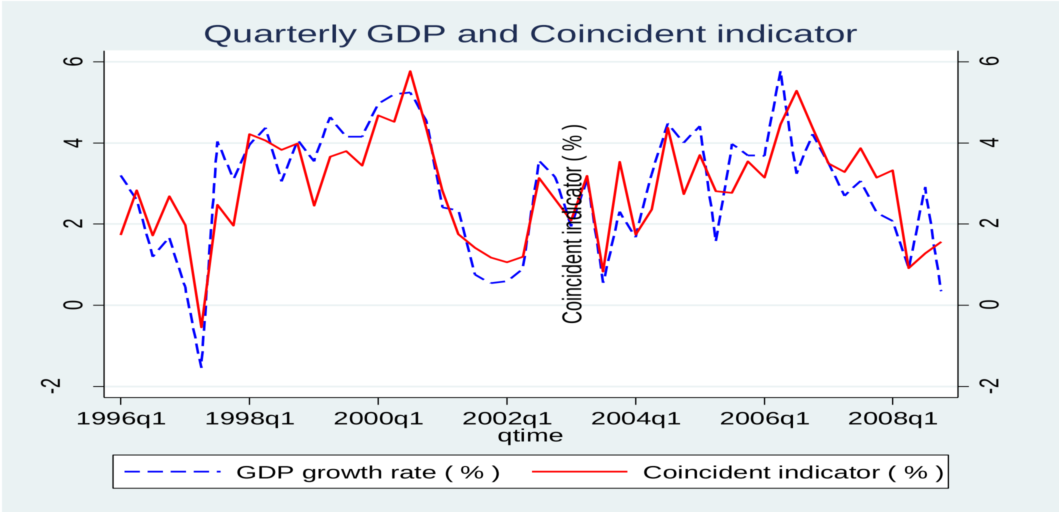

Figure 4.7 Quarterly GDP and Coincident indicator

Figures 4.7 illustrate a close relationship between the constructed quarterly coincident indicator (solid line) and GDP (dashed line); the corresponding value of correlation is equal to 0.83. The pattern of the composite coincident indicator follows the quarterly GDP, note that these graphic practically up and down turns of GDP growth rate series are captured by the composite coincident indicator series and the discrepancies are relatively small, indicating that the new indicator reproduce the same information obtained by GDP, with advantage for CCI because the new indicator takes account all variables, thus describe co - movement of the state of economy.

Analyzing the results in detail, it can be observed that, the ability of the new indicator to approximate the GDP for the forecast period is good. The root mean square forecast error, (RMSFE) is equal to 0.86 percentage point. The corresponding mean absolute error is equal to 0.70 percentage point. The Theil's U statistics "bias" is $5.0{*}10^{-7}$ indicating the absence of systematic errors in the forecast. One another measure of interest when forecasting is the ability of the new indicator to predict the correct signal of the changes; this indicator is measured by the percentage of correct signal, for this specific case the forecast value predict the GDP series correctly about 99.9%, which is good and confirm that the coincident indicator move up or down approximately with the change in the economy.

4.4 Comparison between coincident indicator and alternative indicators

In sections 4.2 and 4.3, when reporting the results for the composite coincident indicator (CCI) it was assumed that the number of common factors and lagged for dependent variable (GDP growth rate) is fixed and OLS was used. In this section the robustness of the indicator is investigated when changes are made in the number of common factors or by changes the model. I calculate alternative indicators using OLS with changes in the number of common factors taking for example 2, 5, 10, 14 and Diffusion index (DI), which is a calculated using only serially uncorrelated common factors without additional lags on the dependent variable. Its also use the univariate ARMA model, precisely an AR(p) process and its constructed the Diffusion index – Autoregressive (DI_AR). The basic idea is to use the equation $\hat{y}_t = \hat{\alpha} + \hat{\beta} \hat{F}_t + a_t$ where a_t which now is assumed to follow an AR(p) process. The main goal in this section is to compare the forecast performance between different indicators.

From monthly series of CCI and alternative indicators with T = 168 forecast points each; the root mean squared forecast error, RMSFE; the mean absolute error, MAE; Theil U statistic, percentage of correct signal predicted by indicator and the correlation between each indicator with GDP growth rate was calculated. Table 4.6 shows that, in terms of RMSFE and MAE, the OLS performs better when decreasing the number of common factors, however in general all average errors in the forecasts is about 0.83 percentage point, that is, the difference with changes the number of common component is not very large; comparison between indicators with the same number of common factors CCI (column 2) and DI (column 7), can see that the CCI outperform DI that indicate that, the effect of lags on dependent variable is positively significant, because the RMSFE and MAE for DI is 1.54 and 1.28 percentage point, quite large compared with CCI measures.

Table 4.6 Forecast performance indicator for monthly indexes

	CCI	Alternative indicators					
		r = 2(1)	r = 5(1)	r =10(1)	r =14(1)	DI(2)	DI_AR(3)
RMSFE	0.88	0.78	0.80	0.83	0.85	1.54	0.75
MAE	0.58	0.43	0.45	0.51	0.50	1.28	0.46
U theil	0.01	0.01	0.01	0.00	0.01	0.12	0.01
Correct signal	0.98	0.99	0.99	0.98	0.98	0.97	0.98
Correlation with GDP	0.80	0.84	0.83	0.82	0.81	0.22	0.85

Notes: (1) – OLS regression for r = 2, 5, 10, 14 common factors and lags on dependent variable;
(2) – OLS regression for r = 18 common factors without lags on dependent variable;
(3) – AR(2) process for r =18 common factors without lags on dependent variable.

Comparing OLS and AR models, CCI (column 2) and DI_AR (column 8) calculated using AR(2) it is see that AR model outperform OLS model in terms of RMSFE and MAE, the ability of the new indicator to predict the correct signal of the changes in the direction of the economy, from table above the indicators report closer similarities, with prediction all about 98 percent. The correlation between each indicator and GDP growth rate (dependent variable) is in general high more than 0.80 except correlation between GDP and DI indicate large discrepancies between corresponding series.

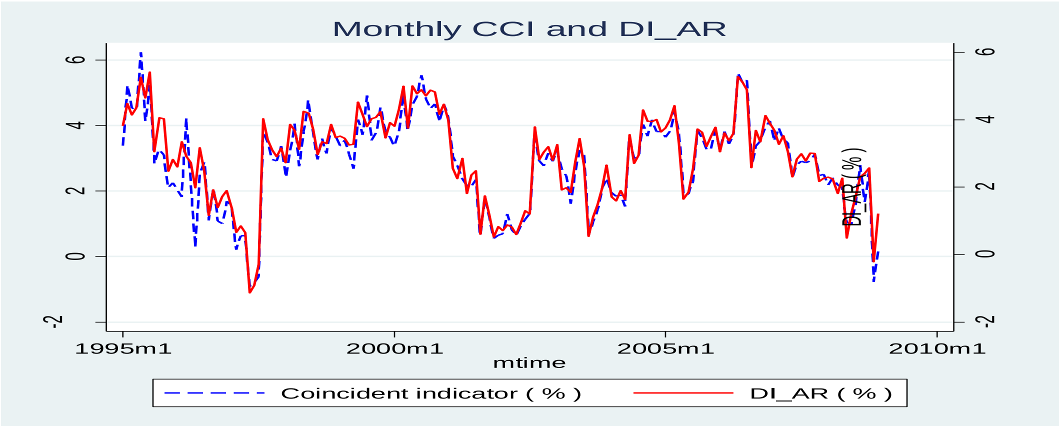

Figure 4.8 Monthly CCI and Diffusion index- Autoregressive

To compare the patterns for two new indicators, Figure 4.8 shows CCI (dashed line) and DI_AR (solid line), the indexes are very similar in terms of oscillation (co-movements up and down of the economic activity), so it can be conclude that the basic information contained in the GDP growth rate is capture by both indicators.

For quarterly panel data, the alternative indicators are indexes with 2, 5, 10 common factors and DI using OLS and DI_AR using AR model. For the forecast period 1996:Q1 to 2008:Q4 five alternative indicators were obtained. The goodness of fit corresponding to the five indicators are 0.28, 0.44, 0.55 for regression with 2, 5 and 10 common factors; 0.40 for DI and 0.61 for DI_AR. In Table 4.7 is reported the important measures of forecast performance, from this table we can see that the RMSFE and MAE improves better when increasing the number of common factors which is 1.23 percentage point (column 3) for forecast from regression with 2 common factors to 0.86 (column 2) percentage point for forecast using 14 common factors.

Table 4.7 Forecast performance indicator for quarterly indexes

	CCI	Alternative indicators				
		r = 2$^{(1)}$	r = 5$^{(1)}$	r =10$^{(1)}$	DI$^{(2)}$	DI_AR$^{(3)}$
RMSFE	0.86	1.23	1.10	0.93	1.07	0.91
MAE	0.70	0.97	0.86	0.76	0.90	0.75
U theil	0.000	0.001	0.001	0.001	0.001	0.002
Correct signal	0.99	0.98	0.98	0.99	0.98	0.99
Correlation with GDP	0.83	0.59	0.70	0.80	0.72	0.81

Notes: (1) – OLS regression for r = 2, 5, 10 common factors and lags on dependent variable;
(2) – OLS regression for r = 14 common factors without lags on dependent variable;
(3) – AR(2) process for r = 14 common factors without lags on dependent variable.

Comparing the CCI and DI we can see that the CCI report better forecast performance for all measures, the difference is due to the regression for DI not taking account of the effect of the idiosyncratic component. Forecast for CCI and DI_AR shows that the OLS outperforms the AR model (column 2 and column 7). The good thing here is that all indicators in table above predict correctly the change of direction of economic activity; the percentage of correct prediction is about 0.98 percent and the correlation between each indicator and quarterly GDP is high for all pairs. Finally Figure 4.9 shows the CCI and DI_AR with similar patterns to underline that one of each pair indicators can be uses to describe the co-movement of the state of the economy.

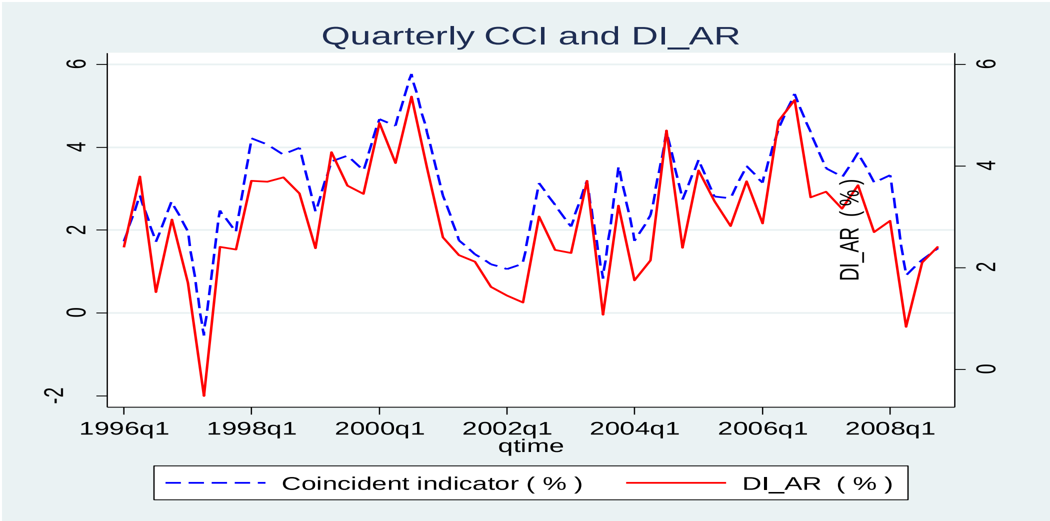

Figure 4.9 Quarterly CCI and Diffusion index –Autoregressive

Summarizing this result discussion it is possible to note that, few time series are available in a timely fashion and most time series variables are published with delays. Moreover this problem is crucial for quarterly data where the delay for some series is over two months. For this analysis we use lags to the series according to corresponding delay time in order to take account of the different release dates. To be stationary the data used in this study in general was transformed by first differences because they are dominated by stochastic trend that was confirmed with Augmented Dickey and Fuller unit root test and for dependent variable (GDP growth rate) the transformation was log-differences.

By using factor model techniques two CCI and respectively alternative indicators was estimated, where for each group of indicators the Figures 4.3 and 4.7 show a close fit in pattern and the relationship between each CCI and corresponding GDP growth rate with correlation 0.80 and 0.83 respectively for the monthly and quarterly indicators. The measures of forecast performance for the two indicators, in general shows that the RMSFE is approximately 0.87 percentage point, the MAE is around 0.64 percentage point and Theil U statistic is about 0.01; the ability of the new indicators to capture the signal of economic activity is high about 98%. These results indicate that the new indicators can describe the economic activity well; details for measures of forecast performance can be seen in Table 4.6 and Table 4.7.

In order to compare the composite coincident indicators sex and five alternative indexes was estimated for monthly and quarterly panel data. From the comparison of forecast performance (Table 4.6 and 4.7), it can be seen that, for monthly indicators the forecast performance does not change much with changes in the number of common factors and DI_AR outperform OLS, while for quarterly indicators the change in number of factors affects RMSFE and MAE and CCI outperform DI_AR. In general CCI and DI_AR in both situations perform similarly. While both new indicators can be used to describe the state of economy, is necessary to underline that the monthly composite coincident indicator has an advantage, because it can provide more timely information than the quarterly CCI, provided that it is constructed from series with less delay.

V CONCLUSIONS

At the end of the 1980s, Stock and Watson developed a new system of composite indexes of coincident economic indicator, for the United States of America, using modern econometric techniques. This paper proposes monthly and quarterly coincident indicator for the Swedish economy built with the same methodology used in factor model structure, Stock and Watson.

The main conclusions of the paper are the following:

(1) The new composite coincident indicator, which includes different categories of economic activities, tracks both the monthly and quarterly GDP growth rate very well ;

(2) The correlation between GDP and the new composite coincident indicators is positive and high around 0.80 and 0.83; moreover the indicators predict in 98 percent and 99 percent correct signal of the changes in the economic activity;

(3) The measures of forecast performance, shows RMSFE approximately 0.87 percentage point and 0.64 percentage points for MAE indicate that the indicators are reasonable and can be used as indicators for the state of the economy.

(4) Because monthly CCI is computed before the real GDP, consequently before quarterly CCI the monthly CCI has an advantage and could provide timely and useful information about the state of the economy.

REFERENCES

Ahmad, N. (**2008**):"Malaysia Economic Indicators – leading, coincident and lagging indicators". *Department of statistics Malaysia*

Altissimo, F., A. Bassanetti, R. Cristadoro, M. Forni, M. Lippi, L. Reichlin and G.Veronese (**2001**):" A real time coincident indicator of the euro area business cycle". *Centre for Economic Policy Research Discussion Paper*, number 436, December 2001

Altissimo, F., R. Cristadoro, M. Forni, M. Lippi, and G. Veronese (**2007**):"New Eurocoin: Tracking Economic Growth in real time". *Economic Research Department*, Discussion Paper 631, Bank of Italy

Boivin, J. and S. Ng (**2005**): "Understanding and Comparing Factor – Based Forecasts," *International Journal of Central Banking. 3*, 117-151.

Boivin, J. and S. Ng (**2006**): "Are more data always better for factor analysis?".*Journal of Econometrics*

Burns, A.F. and W.C. Mitchell (**1946**): "Measuring Business Cycle. Now York. NBER

Chamberlain, G., and M. Rothschild (**1983**): "Arbitrage, factor structure and mean - variance analysis in large asset markets", *Econometrica*, 51, 1305 = 1324.

Crone, T.M. (**2000**): "A new look at Economic indexes for the States in the third district" *Business review*, November/December 2000

Crone, T.M. (**2006**): " What a new set of indexes tells us about state and national business cycles". *Business review*, January 2006

D'Agostino, A, D. Giannone (**2006**): " Comparing Alternative Predictors Based on Large – Panel Factor Models". *European Central Bank*, Working Paper Series No. 680 October 2006

Eickmeier, S. and C. Ziegler (**2006**): "How good are dynamic factor models at forecasting output and inflation?. *EuroSystem Deutsche Bundesbank*. Discussion Paper Series 1: Economic studies, No 42.

Forni, M., M. Hallin, M Lippi and L. Reichlin (**2000**): "The generalized dynamic factor model: identification and estimation" *Review of Economics and Statistics*, 82(4), pp 540 – 554.

Forni, M., M. Hallin, M. Lippi and L. Reichlin (**2001**): "Coincident and leading indicators for the euro area". *The economic journal* 111, pp 62 – 85

Forni, M., M. Hallin, M. Lippi and L. Reichlin (**2001**):"Do financial variables help forecasting inflation and real activity in the Euro area?. *Fundation Banque de France*. December 2001.

Forni, M., M. Hallin, M Lippi and L. Reichlin (**2005**): “The generalized dynamic factor model: one- side estimation and forecasting” *Journal of the America Statistical Association*, 100, pp.830 - 840.

Fukuda, S. and O. Takashi (**2001**): “A new composite index of coincident economic indicators in Japan: How can we improve forecast performances?”. *International Journal of forecasting 17*, pp.483 - 498

Gaudreault, C, R. Lamy and Y. Liu (**2003**):” New coincident, leading and recession index for the Canadian Economy. An application of the Stock and Watson methodology*”. Department of Finance*. Working Paper, December 2003.

Gillitzer, C., J. Kearns and A. Richards (**2005**): ” The Australian Business Cycle: A Coincident Indicator Approach”, *Economic Group*, Reserve Bank of Australia, Working paper pp.262 -307

Goodwin et al (**2007**): *"*Economic growth." In: *Encyclopedia of Earth*. Eds. Cutler J. Cleveland. [First published in the Encyclopedia of Earth April 23, 2007].: Retrieved January 19, Site: http://www.eoearth.org/article/Economic_growth,

Hall, S.G, N.G. Zonzilos (**2003**): ”An indicator measuring underlying economic activity in Greece”. *Bank of Greece*. Working Peper No.4, August 2003

Hassler, J., P. Lundvik, T. Persson and P. Soderlind (**1992**): ”The Swedish Business Cycle: Stylized facts over 130 years”, *Institute for International Economic Studies* – Stockholm University.

Inklaar, R., J. Jacobs and W. Ramp (**2003**).”Business cycle indexes: does a leap of data help?” University of Groningen, CCSO, Working Paper No 2003/12.

Johnson, R.A and D.W. Wichern (**2007**):” Applied multivariate Statistical Analysis”. *Pearson International Edition,* Sixth edition. Printice Hall 773 pages.

Kaiser, H.F. (1960).”The application of electronic computers to factor analysis*”. Educational and Psychological measurement*, 20, 141 – 151.

Kubandi, A. (**2004**): “The generalized dynamic factor models: forecasting inflation rate in France with business surveys”. *Ninth annual conference on econometric modeling for Africa*. Rand Afrikaans University (South Africa). 30 June to 2 July 2004

Patnaik, I and R. Sharma (**2002**): “Business cycles in the Indian economy”. Margin, 35(1):71–79, October–December 2002

Pindyck, R.S and D.L. Rubinfeld (**1998**): “Economic models and economic forecasts”, *Irwin*, McGraw – Hill.

Quah, D. and T. Sargent (**1993**): ”A dynamic index model for large cross sections “, Ch.7 in J.H. Stock and M.W. Watson (eds), *Business Cycle, indicators and forecasting*. University of Chicago Press for the NBER. 285 – 306.

Schumacher, C. (**2006**): “Forecasting German GDP using alternative factor models based on large dataset,” *Journal of Forecasting*, forthcoming.

Stock, J.H and M. W. Watson (**1989**):”New Indexes of Coincident and Leading Economic indicators”, *NBER Macroeconomic Annual*, 1989 pp 351 – 94.

Stock, J.H., and M.W. Watson (**1999**): ”Forecasting inflation”, *Journal of Monetary Economic*, 44 (2), pp 293 – 33.

Stock, J.M and M.W. Watson (**2002a**): “Forecasting using principal components from a large number of predictors”, *Journal of the American Statistical Association* 97 (460), pp 1167 – 1179.

Stock, J.M and M.W. Watson (**2002b**): “Macroeconomic forecasting using diffusion indexes”. *Journal of Business and Economic Statistics* 20(2), pp147 – 162.

Stock, J.M and M.W. Watson (**2004a**): “Forecasting with many predictors”, *Unpublished Manuscript*, Princeton University.

The Conference Board U.S. Business Cycle Indicators (**2008**):” U.S. leading economic indicators and related composite indexes for September 2008”.

Tsay, R. S. (**2005**): “Analysis of financial time series” *Second edition Wiley Interscience*.

Watson, M.W. (**2001**): “Macroeconomic forecasting using many predictors”, M. Dewatripont, LP Hansen and SJ Turnovsky (eds). *Advances in economics and econometrics: theory and applications*, Eighth Word Congress, Volume III, Cambridge University Press, Cambridge, pp 87-114.

Appendix A – List of macroeconomic time series and they transformations

Table A1: Monthly economic variables by category (panel 1990:01 to 2008:12) and type of transformation used: x = time series; l = natural logarithm; d = first difference; z = normalization

Categories and economic time series	Transformation
Public finances (= 5 series)	
01. Monthly central government debit - total debit	zdlx
02. Monthly central government debit - government bonds	zdlx
03. Monthly central government debit - premium bonds	zdlx
04. Monthly central government debit - everyman's savings	zdlx
05. Monthly central government debit - treasury bills	zdlx
Business activities (= 9 series)	
06. Industrial production index - working day	zdx
07. Industrial production index - non working day	zdx
08. Industrial production index - trend	zdlx
09. Deliveries in industry - non working days in total market, cp	zdx
10. Deliveries in industry - non working days in domestic market, cp	zdx
11. Deliveries in industry - non working days in export market, cp	zdx
12. Deliveries in industry - working days in total market, cp	zdx
13. Deliveries in industry - working days in domestic market, cp	zdx
14. Deliveries in industry - working days in export market , cp	zdx
Prices and consumption (= 20 series)	
15. Inflation rate according to CPI	zdx
16. Inflation rate according to CPIX	zdx
17. Inflation rate according to CPIF	zdx
18. Consumer price index - food and non-alcoholic beverages	zdlx
19. Consumer price index - clothing and footwear	zdlx
20. Consumer price index - housing, water, electricity, gas & fuels	zdlx
21. Consumer price index - health	zdlx
22. Consumer price index - transport	zdlx
23. Consumer price index - communication	zdlx
24. Consumer price index - CPI over Living Cost Index	zdlx
25. Consumer price index - total shadow Index numbers	zdlx
26. Consumer price index - total fixed Index numbers	zdlx
27. Consumer price index - mortgage interest cost	zdlx
28. Consumer price index - goods	zdlx
29. Consumer price index - services	zdx
30. Consumer price index - housing	zdx
31. Consumer price index - weight mortgage interest cost	zdx
32. FPI for agricultural buildings - material	zdlx
33. FPI for agricultural buildings - labour	zdlx
34. FPI for agricultural buildings - plant and equipment	zdlx
Trade in goods and services (= 6 series)	
35. Foreign trade - total imports, SEK million	zdlx
36. Foreign trade - total exports, SEK million	zdlx

37. Foreign trade - trade of goods, SEK million	zdlx
38. Retail trade - total retail	zdlx
39. Retail trade - mostly food	zdlx
40. Retail trade - mostly durables	zdlx
Energy (= 6 series)	
41. Electricity consumption - total electricity	zlx
42. Electricity consumption - export	zlx
43. Electricity consumption - consumption in the county	zlx
44. Electricity consumption - quarrying and manufacturing	zdlx
45. Electricity consumption - housing, services, etc	zlx
46. Electricity consumption - housing temperature corrected	zlx
Labour market (= 3 series)	
47. Labour force survey - employment	zdlx
48. Labour cost index – total	zdx
49. Labour cost index – manufacturing industry	zdx
Transport and communications (= 6 series)	
50. Vehicles - bil registter of cars	zdlx
51. Vehicles - bil registter of tracks	zdlx
52. Vehicles - bil registter of buses	zdlx
53. Vehicles - bil registter of tractors	zdlx
54. Vehicles - bil registter of trailers	zdlx
55. Vehicles - bil registter of snowmobiles	zdlx
Financial markets (= 13 series)	
56. Montlhy exchange rate - SEK/Canada	zdx
57. Montlhy exchange rate - SEK/ USA	zdx
58. Montlhy exchange rate - SEK/Japan	zdx
59. Montlhy exchange rate - SEK/Switzerland	zdlx
60. Montlhy exchange rate - SEK/Denmark	zdx
61. Montlhy exchange rate - SEK/Norway	zdlx
62. Montlhy exchange rate - SEK/Euro	zdlx
63. Montlhy exchange rate - SEK/Import weighted	zdx
64. Montlhy exchange rate - SEK/Export weighted	zdlx
65. Montlhy exchange rate - SEK/Competitor weighted	zdlx
66. Montlhy exchange rate - SEK/KIX	zdlx
67. Interest rates - treasury bills 3 months	zdx
68. Interest rates - government bonds 10 years	zdx

Table A2: Quarterly economic variables by category (panel 1993:Q to 2008:Q4) and type of transformation used: x = time series; l = natural logarithm; d = first difference; d2 = second difference; z = normalization.

Categories and economic time series	transformation
Business activities (=7 series)	
01. NACE Mines and quarries and manufacturing industry, s.a.	zdlx
02. IPI - Working day adjusted	zdlx
03. IPI - Non working day adjusted	zdlx
04. IPI - Working day adjusted and seasonally	zdlx
05. IPI - Working day in manufacturing industry	zdlx
06. Investment in buildings and construction works	zdlx
07. Investment in machinery and equipment	zdlx
Housing, construction and building (=8 series)	
08. Conversion - starts after conversion in multi-dwelling buildings	zlx
09. Conversion - completed after conversion in multi-dwelling buildings	zdlx
10. New construction - starts multi-dwelling buildings	zdlx
11. New construction - starts one or two dwelling buildings	zdlx
12. New construction - completed multi-dwelling buildings	zdlx
13. New construction - completed one - or two-dwelling buildings	zdlx
14. Real estate price index for buildings for seasonal and secondary use	zdlx
15. Real estate price index for one- and two-dwelling buildings for permanent use	zdlx
Labour market (= 13 series)	
16. Number of employees - total private sector	zdx
17. Number of employees - mining, quarrying and manufacturing	zdx
18. Number of employees - mines and quarries	zdx
19. Number of employees - manufacturing industry	zdx
20. Number of employees – food, product, beverage & tobacco industry	zdx
21. Number of employees - industry for coke refined petroleum products and nuclear fuel and industry for chemicals	zdx
22. Number of employees - industry for rubber and plastic	zdlx
23. Number of employees - industry for transport equipment	zd2lx
24. Number of employees - retail trade and repair shops for personal and household goods	zdlx
25. Number of employees - real estate and business services	zdlx
26. Number of employees - financial institutions and insurance companies	zdlx
27. Number of employees - institutes for research and development	zdlx
28. Number of employees - other manufacturing industry	zdlx
National accounts (= 34 series)	
29. GDP - GDP at market prices, cp	zdlx
30. GDP - Imports of goods and services, cp	zdlx
31. GDP - Imports of goods, cp	zdlx
32. GDP - Imports of services , cp	zdlx
33. GDP - Household consumption expenditure, include NPISH, cp	zdlx

34. GDP - Central government. final consumption expenditure, cp	zdlx
35. GDP - Market producers and NPISH, fixed capital formation, cp	zdlx
36. GDP - Local government final consumption expenditure, cp	zdlx
37. GDP - Gross fixed capital formation, cp	zd2lx
38. GDP - Export of services, cp	zdlx
39. GDP - Exports of goods and services, cp	zdlx
40. GDP - Export of goods, cp	zdlx
41. GDP - Exports of services, cp	zdlx
42. Household consumption - goods and non-alcoholic beverages, cp	zdlx
43. Household consumption - alcoholic beverages and tobacco, cp	zdlx
44. Household consumption - clothing and footwear, cp	zdlx
45. Household consumption - housing, water, electric., gas & fuels, cp	zdlx
46. Household consumption – health, cp	zdlx
47. Household consumption – transport, cp	zdlx
48. Household consumption – communication, cp	zdlx
49. Household consumption - resident expenditure abroad, cp	zdlx
50. Disposable income – net	zdlx
51. Disposable income - wages and salaries, resources	zdlx
52. Disposable income - employees social contributions, resources	zdlx
53. Disposable income - individual consumption expenditure	zdlx
54. Labour input - average number of persons employed	zdlx
55. Labour input - hours worked, ten thousands	zdlx
56. Hours worked - total market producers for own final use, sa	zdlx
57. Hours worked - producers of goods, sa	zlx
58. Hours worked - agriculture, hunting, forestry and fishing, sa	zdlx
59. Hours worked – manufacturing, sa	zdlx
60. Hours worked - construction industry, sa	zd2lx
61. Hours worked - wholesale and retail trade, sa	zd2lx
62. Hours worked - financial institutions and insurance companies, sa	zd2lx
Prices and consumption (= 9 series)	
63. Input FPI - multi-dwelling buildings, total including vat	zdlx
64. Input FPI - multi-dwelling buildings, total excluding vat	zdlx
65. Input FPI - multi-dwelling buildings, Labor	zdlx
66. Input FPI - built one- or two-dwelling buildings, total including vat	zdlx
67. Input FPI - built one- or two-dwelling buildings, total excluding vat	zdlx
68. Input FPI - built one- or two-dwelling buildings, labor	zdlx
69. New residential FPI - total factor price index	zdlx
70. New residential FPI - material	zdlx
71. New residential - Labour	zdlx
Trade in goods and services (= 23 series)	
72. Volume imported - food beverages and tobacco	zdlx
73. Volume imported - total	zdlx
74. Volume imported - crude materials, except fuels	zd2lx
75. Volume imported - mineral fuels, rubric. and related materials	zlx
76. Volume imported - chemicals and related products, n.e.s.	zd2lx
77. Volume imported - processed goods	zd2lx

78. Volume imported - manufactured goods classify by material	zd2lx
79. Volume imported - machinery and transport equipment	zd2lx
80. Volume imported - other manufactured articles	zdlx
81. Volume exported - food beverages and tobacco	zd2lx
82. Volume exported - total	zd2lx
83. Volume exported - crude materials, except fuels	zdlx
84. Volume exported - mineral fuels, rubric. and related materials	zd2lx
85. Volume exported - chemicals and related products, n.e.s	zd2lx
86. Volume exported - processed goods	zd2lx
87. Volume exported - manufactured goods classify by material	zd2lx
88. Volume exported - machinery and transport equipment	zd2lx
89. Volume exported - other manufactured articles	zdlx
90. Turnover index, services and activities, constants prices	zdlx
91. Turnover index, working day in service and activities, c.p.	zdlx
92. Turnover index, working day in wholesale trade and commission	zdlx
93. Turnover index, working day in wholesale on a fee or contract basis	zdlx

Appendix B. Table of r PCs and STATA code

Table B1: Eigenvalues for the first 18 PC, 67 explain total variability of monthly panel data.

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1 |	7.69369	0.29796	0.1131	0.1131
Comp2 |	7.39573	1.18688	0.1088	0.2219
Comp3 |	6.20885	1.08182	0.0913	0.3132
Comp4 |	5.12703	0.38816	0.0754	0.3886
Comp5 |	4.73887	1.24965	0.0697	0.4583
Comp6 |	3.48922	0.57717	0.0513	0.5096
Comp7 |	2.91204	0.40653	0.0428	0.5524
Comp8 |	2.50551	0.30949	0.0368	0.5893
Comp9 |	2.19602	0.21025	0.0323	0.6216
Comp10 |	1.98576	0.16925	0.0292	0.6508
Comp11 |	1.81650	0.24161	0.0267	0.6775
Comp12 |	1.57488	0.16997	0.0232	0.7006
Comp13 |	1.40491	0.05290	0.0207	0.7213
Comp14 |	1.35201	0.12610	0.0199	0.7412
Comp15 |	1.22590	0.08204	0.0180	0.7592
Comp16 |	1.14386	0.08519	0.0168	0.7760
Comp17 |	1.05866	0.04810	0.0156	0.7916
Comp18 |	1.01056	0.08375	0.0149	0.8065

Table B2: Eigenvalues for the first 14 PC, 62 explain total variability of quarterly panel data.

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1 |	22.64350	8.83736	0.2435	0.2435
Comp2 |	13.80610	5.55766	0.1485	0.3919
Comp3 |	8.24844	1.66265	0.0887	0.4806
Comp4 |	6.58579	2.17034	0.0708	0.5514
Comp5 |	4.41546	1.11237	0.0475	0.5989
Comp6 |	3.30308	0.23525	0.0355	0.6344
Comp7 |	3.06783	0.23090	0.0330	0.6674
Comp8 |	2.83692	0.40309	0.0305	0.6979
Comp9 |	2.43383	0.33404	0.0262	0.7241
Comp10 |	2.09979	0.10787	0.0226	0.7467
Comp11 |	1.99191	0.20440	0.0214	0.7681
Comp12 |	1.78751	0.21635	0.0192	0.7873
Comp13 |	1.57116	0.05856	0.0169	0.8042
Comp14 |	1.51260	0.12610	0.0163	0.8205

STATA code

#1 Calculus of r common factors (k is maximum number of variables in the panel)

```
1) . pca var1, var2, ..., vark                 (# output PC’s, then select r)
2) . matrix w1 = (value11\vaue21\value31\ ...\valuek1) (# define column of
   . Matrix w2 = (value12\value22\value32\...\valuek2)    weights )
     ...
   . Matrix wr = (value1r\value2r\value3r\...\valuekr)

3) . mkmat var1
     ...
   . mkmat vark
```

```
4) . matrix X = var1, var2, ..., vark  (# Define matrix X= (x1, …, xk))
   . mat list X
5) . matrix factor1 = X*w1
   . matrix factor2 = X*w2
     ...
   . matrix factorr = X*wr   (# equation 3.3 in the text)
6) . mat list factor_i

Predict f1 f2 ... fr  (* produce score components)
```

#2 Equations for monthly indicators

```
. arima Y, ar(1/4)
. reg y f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 f11 f12 f13 f14 f15 f16 f17 f18 l(1/3).y
. reg y f1 f2 l(1/3).y
. reg y f1 f2 f3 f4 f5 l(1/3).y
. reg y f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 l(1/3).y
. reg y f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 f11 f12 f13 f14 f15 f16 f17 f18
. arima y f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 f11 f12 f13 f14 f15 f16 f17 f18, ar(1/2)
```

#3 Equations for quarterly indicators

```
. reg y f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 f11 f12 f13 f14 l(1/3).y
. reg y f1 f2 l(1/3).y
. reg y f1 f2 f3 f4 f5 l(1/3).y
. reg y f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 l(1/3).y
. reg y f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 f11 f12 f13 f14
. arima y f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 f11 f12 f13 f14, ar(1/2)
```

#4 Forecast procedure

```
set more off
capture: drop f_*
* First forecast origin
local first = tm(1994:12)
* Last forecast origin
local last = tm(2009:4)
* Max horizon to evaluate forecasts at
local fhor = 4
* Number of forecast origins
local numfor = `last' - `first' + 1 - `fhor'

forvalues i = 1/`numfor' {
   local firstfor = `first'+`i'
   * Model specification
   Equation if mtime < `firstfor'
   * Forecast for dependent variable
   predict f_gdp`i'
   * Replace forecast horizons we are not interested in with missing values
so we don't plot them
   replace f_gdp`i' = . if mtime < `firstfor' | mtime > `firstfor'+`fhor'-1

}
```