CENTRAL BANK OF ICELAND

WORKING PAPERS No. 36

Forecasting the Icelandic business cycle using vector autoregressive models

by
Bruno Eklund

September 2007

CENTRAL BANK OF ICELAND

Economics Department

Central Bank of Iceland Working Papers are published by the Economics Department of the Central Bank of Iceland. The views expressed in them are those of their authors and not necessarily the views of the Central Bank of Iceland.

Also available on the Central Bank of Iceland World Wide Web site (http://www.sedlabanki.is)

ISSN: 1028-9445

Forecasting the Icelandic business cycle using vector autoregressive models

Bruno Eklund*
Central Bank of Iceland

Abstract

This paper considers the modelling and forecasting of the Icelandic business cycle. The method of selecting monthly variables, coincident and leading, that mimic the cyclical behavior of the quarterly GDP is described. The general business cycle is then modelled by a vector autoregressive, VAR, model. The cyclical behavior of the business cycle is summarized by a composite coincident index, which is based on the root mean squared forecast error over a pseudo out of sample. By applying a bootstrap forecasting procedure, using the estimated VAR model, point and interval forecasts of the composite coincident index are estimated.

Keywords: Leading indicator, VAR model, Bootstrap forecasting, Composite coincident index.
JEL codes: C15, C32, E37.

*Bruno Eklund, Economics Department, Central Bank of Iceland, Kalkofnsvegur 1, IS-150 Reykjavik, Iceland, Bruno.Eklund@Sedlabanki.is

1 Introduction

Early information on how the state of the economy or the business cycle is evolving over time is of key interest to central banks, other business institutes and market agents. It is therefore no surprise that the literature on methods for extracting information from various economic variables, and forecasting the business cycle, has grown considerably in recent years. In the early work by Burns and Mitchell (1946) the business cycle was defined as representing comovements in a broad set of macroeconomic variables. Their non-model based approach, applied by the NBER business cycle dating committee, was later formalised using modern econometric techniques in an important study by Stock and Watson (1989). After this, further development has been rapid and, in particular, methods concerning leading indicators have attracted considerable attention. Numerous techniques exist today for the selection and evaluation of coincident and leading variables, and for describing the relationship between these variables and the general business cycle using an increasing number of statistical models. Marcellino (2006) contains an extensive survey on methods for selecting coincident and leading variables, and for the modelling of the business cycle.

This paper will describe a simple method for selecting coincident variables and leading indicators based on the correlation structure between the variables and the yearly growth rate of GDP. Applied on the Icelandic economy, the selection method identifies a number of potential coincident variables and leading indicators from a large set of macroeconomic and financial time series that are available on a monthly basis. A vector autoregressive, VAR, model is then specified, which provides one of the simplest model based framework for understanding the relationship between coincident and leading variables, and for the construction of composite coincident indexes. The VAR model is estimated using generalized least squares, and a composite coincident index, CCI_t, is constructed based on the inverse of the root mean squared forecast error over a pseudo out of sample. Point forecasts and confidence intervals of the CCI_t are estimated using a bootstrap forecasting method. A common problem of VAR models is often the large number of parameters to estimate, preventing the analysis of large data sets. Other methods exist that can overcome this problem, such as Bayesian techniques and factor models. Nevertheless, the simplicity of the modelling and forecasting procedure of the VAR model makes it a good candidate describing the relationship between the variables and for forecasting the business cycle.

In recent years, the Icelandic business cycle has been studied on a few occasions. In Pétursson (2000), the business cycle dynamics were modelled using yearly GDP and a univariate Markov switching model. Eklund

(2007) based the modelling procedure on the classic Stock and Watson (1989) methodology using monthly data and specified a model in state space form, and provided probabilities for recession and expansion over a twelve months forecast horizon. The main properties of the Icelandic business cycle have also been analysed in Daníelsson et al. (2006) using a macroeconomic model for quarterly data.

The remainder of the paper is organized as follows. Section 2 discusses the method of selecting potential coincident variables and leading indicators. The VAR model is described in Section 3, while Section 4 presents the bootstrap forecasting method. Section 5 applies the methods and techniques outlined in previous sections on the Icelandic economy. Final conclusions can be found in Section 6.

2 Selection of coincident and leading variables

Icelandic GDP could provide a reliable summary of the current state of the economy if it were available on a monthly basis and does not tend to be revised heavily. Since Icelandic GDP is only measured on a quarterly basis, it can only be used indirectly in the modelling process. Candidate variables, mimicking the cyclical behaviour of GDP and available on a monthly basis, are instead selected and classified as coincident, lagging or leading. Any of the coincident variables can be used as a proxy for GDP, but usually a single coincident economic index is formed to represent the business cycle dynamics. Several methods both in model and non-model based frameworks exist today for such a construction.

The selection of variables can typically be performed by analysing the correlation structure between GDP and the candidate variable taken from a list of macroeconomic variables. Originating from any possible source available, candidate variables having a high correlation with the cyclical behaviour of GDP are of interest. Other numerous criteria for selecting variables can be considered, such as consistent timing, conformity to the business cycle, economic significance and more, as defined in the more formalized scoring system due to the often quoted Moore and Shiskin (1967). Variables can also be chosen on the basis of spectral coherence or the lead time in turning points, but these methods are not considered in this paper.

If the estimated correlation between a variable x_t and GDP is high at some lag k, then x_t is said to be a potential leading indicator if $k > 0$, a potential coincident variable if $k = 0$, or lagging if $k < 0$ respectively. Since lagging variables do not contain any early information of the present or the future economic state, they can be excluded from the analysis. The cyclical behaviour of a leading indicator thus mimics but precedes that of GDP, while the behaviour of a coincident variable and GDP, as the name suggests, coincides over time. Note that variables with a high negative correlation with GDP can also serve as good indicators of the evolution of the business cycle, and should be added to the list of potential coincident and leading variables.

By construction leading indicators will give early indications about changes in the direction of GDP and business cycle. However, they will not provide any reliable information about the magnitude of that change. Furthermore, the lag relationships can be quite volatile which is why it can many times be difficult in practice to apply leading indicators. Also, the widespread use of a reasonably reliable leading indicator may in fact, ironically, lead to less reliability in the indicator over time. This can result if the agents in the economy act on the forecast and alter either the economic outcome or the lead time between the indicator and the economy. Despite its drawbacks, leading indicator series can help economists, business and government predict and prepare for significant changes in the economic climate.

3 The vector autoregressive model

The vector autoregressive (VAR) model is one of the simplest forms of multivariate models. Its popularity for analysing the dynamics of economic systems is due to the influential work by Sims (1980). It is particularly convenient for estimation and forecasting, and provides the simplest model based framework for relating leading indicators to coincident variables, and for the construction of regression based composite indexes.

Let $\mathbf{y}_t$ be a set of m coincident variables, and $\mathbf{x}_t$ a set of n leading indicators. Collect the variables $\mathbf{y}_t$ and $\mathbf{x}_t$ in the $(m+n)$-dimensional column vector $\mathbf{z}_t = (\mathbf{y}_t', \mathbf{x}_t')'$. The vector autoregressive model with p lags, VAR(p), can then be defined as follows;

$$\mathbf{z}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{z}_{t-1} + \ldots + \mathbf{\Phi}_p \mathbf{z}_{t-p} + \boldsymbol{\varepsilon}_t, t = 1, \ldots, T, \tag{1}$$

where $\mathbf{c}$ is a column vector of constants, $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p$ are parameter matrices, and $\boldsymbol{\varepsilon}_t$ is a $(m+n)$-dimensional i.i.d. error vector process with zero mean and covariance matrix $\mathbf{\Sigma}$. Assuming normally distributed errors, the model parameters can be estimated by maximum likelihood, or equivalently by ordinary least squares equation by equation, see for example Hamilton (1994) for details. To be able to estimate the model under more general parameter restrictions, and to perform tests on single parameters, the model is re-expressed in the equivalent alternative form

$$
\begin{aligned}
y_{1,t} &= \mathbf{w}_{1,t}\boldsymbol{\beta}_1 + \varepsilon_{1,t} \\
&\vdots \\
y_{m,t} &= \mathbf{w}_{m,t}\boldsymbol{\beta}_m + \varepsilon_{m,t} \\
x_{1,t} &= \mathbf{w}_{m+1,t}\boldsymbol{\beta}_{m+1} + \varepsilon_{m+1,t} \\
&\vdots \\
x_{n,t} &= \mathbf{w}_{m+n,t}\boldsymbol{\beta}_{m+n} + \varepsilon_{m+n,t},
\end{aligned}
\tag{2}
$$

where $\mathbf{w}_{i,t}$ is a row vector consisting of all variables that appear in equation i, that is, a possible constant and all lags of $\mathbf{y}_t$ and $\mathbf{x}_t$, and $\boldsymbol{\beta}_i$ is the corresponding parameter vector. Let k_i be the number of parameters to be estimated in equation i, and denote by $k = k_1 + \ldots + k_{m+n}$ the total number of parameters in the model. Collect these in the $(k \times 1)$ column vector

$$
\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_{m+n}'\right)'. \tag{3}
$$

The equation system (2) can then be expressed as

$$
\mathbf{z}_t = \mathbf{H}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \tag{4}
$$

where $\mathbf{H}_t$ is the $((m+n) \times k)$ matrix

$$
\mathbf{H}_t = \begin{bmatrix} \mathbf{w}_{1,t} & \mathbf{0}_2 & \ldots & \mathbf{0}_{m+n} \\ \mathbf{0}_1 & \mathbf{w}_{2,t} & \ldots & \mathbf{0}_{m+n} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0}_1 & \mathbf{0}_2 & \ldots & \mathbf{w}_{m+n,t} \end{bmatrix}, \tag{5}
$$

where $\mathbf{0}_i$ are zero vectors of order $(1 \times k_i)$, $i = 1, \ldots, m+n$, added to make $\mathbf{H}_t$ conformable with the $(k \times 1)$ vector $\boldsymbol{\beta}$.

Estimation of (4) is performed by an iterative process. First an initial estimate $\widehat{\boldsymbol{\Sigma}}(0)$ of the covariance matrix is obtained given a start vector of parameter values $\widehat{\boldsymbol{\beta}}(0)$. Then, by using $\widehat{\boldsymbol{\Sigma}}(0)$, an estimate $\widehat{\boldsymbol{\beta}}(1)$ of the parameter vector $\boldsymbol{\beta}$ can be obtained by generalized least squares, which in the next step produces a new estimate $\widehat{\boldsymbol{\Sigma}}(1)$ of $\boldsymbol{\Sigma}$. Iterating in this manner will produce the maximum likelihood estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}$, though these estimates after just one iteration have the same asymptotic distribution as the final maximum likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, see Magnus (1978) for details.

4 Forecasting

It follows immediately that the optimal one-step ahead forecast of model (4) is

$$\widehat{\mathbf{z}}_{T+1} = \mathbf{H}_{T+1}\widehat{\boldsymbol{\beta}}. \tag{6}$$

Note that $\mathbf{H}_t$ in (4) consists of possible constants and only lagged values of $\mathbf{y}_t$ and $\mathbf{x}_t$, so all elements of $\mathbf{H}_{T+1}$ in (6) are known.

The estimation of higher step forecasts is, however, not as straight forward. Several techniques can be applied. Two commonly used methods are the naive forecast and the bootstrap forecasting method. In the naive case, the presence of forecast error is ignored, producing, for example, the two-step point forecast

$$\widehat{\mathbf{z}}_{T+2}^{naive} = \widehat{\mathbf{H}}_{T+2}\widehat{\boldsymbol{\beta}}, \tag{7}$$

where the unknown lags of $\mathbf{y}_{T+2}$ and $\mathbf{x}_{T+2}$ in $\widehat{\mathbf{H}}_{T+2}$ are replaced by their respectively point forecast from $\widehat{\mathbf{z}}_{T+1} = \left(\widehat{\mathbf{y}}'_{T+1}, \widehat{\mathbf{x}}'_{T+1}\right)'$. The matrix $\widehat{\mathbf{H}}_{T+h}$ is thus updated for each forecast step h with the point estimates of $\mathbf{y}_t$ and $\mathbf{x}_t$ from the earlier forecast steps. This method is easy to use, but since it does not take into account any errors of the forecasts, it will usually be badly biased.

The bootstrap forecasting method, on the other hand, incorporates forecast errors when estimating higher step forecasts, see for example Granger and Teräsvirta (1993) and Teräsvirta (2006). The two-step point forecast is in this case

$$\widehat{\mathbf{z}}_{T+2}^{b} = \frac{1}{R-1}\sum_{i=1}^{R-1}\widehat{\mathbf{z}}_{i,T+2}^{b} = \frac{1}{R-1}\sum_{j=1}^{R-1}\widehat{\mathbf{H}}_{j,T+2}\widehat{\boldsymbol{\beta}}, \tag{8}$$

b denoting the bootstrap forecast, and where the unknown lags of $\mathbf{y}_{T+2}$ and $\mathbf{x}_{T+2}$ in the matrix $\widehat{\mathbf{H}}_{j,T+2}$ are replaced as before by their respectively point forecast, but now also adding a drawn value of the estimated error vector of $\boldsymbol{\varepsilon}_t$ in the model (4). More precisely, when adding the first step forecast of, for example, $\widehat{y}_{1,T+1}$ to $\widehat{\mathbf{H}}_{j,T+2}$, a drawn value $e_{1,j}$ from the residual vector $\mathbf{e}_t$ in (4) is added to $\widehat{y}_{1,T+1}$, thus $\widehat{y}_{1,T+1} + e_{1,j}$ will be included in $\widehat{\mathbf{H}}_{j,T+2}$, and so on for the other variables. For the two-step forecast $R-1$ residual vectors are drawn, each resulting in a point forecast $\widehat{\mathbf{z}}_{i,T+2}^{b}$, $i = 1, \ldots, R-1$. As a two-step point forecast it is natural to choose the mean of the $R-1$ values of $\widehat{\mathbf{z}}_{i,T+2}^{b}$ as in (8), even though other forms can be considered, such as the median. The method can clearly be used for multi-step forecasts although it quickly becomes complex. For example, the recommended three-step point forecast is

$$\widehat{\mathbf{z}}^b_{T+3} = \frac{1}{(R-1)(R-2)} \sum_{i=1}^{(R-1)(R-2)} \widehat{\mathbf{z}}^b_{i,T+3} = \frac{1}{(R-1)(R-2)} \sum_{j=1}^{R-1} \sum_{k=1}^{R-2} \widehat{\mathbf{H}}_{j,k,T+3}\widehat{\boldsymbol{\beta}}, \tag{9}$$

where the matrices $\widehat{\mathbf{H}}_{j,k,T+2}$, $j = 1, \ldots, R-1$, $k = 1, \ldots, R-2$, are constructed using each of the $R-1$ matrices $\widehat{\mathbf{H}}_{j,T+2}$ in (8) and updating them, as before, with the values of $\mathbf{y}_t$ and $\mathbf{x}_t$ from the previous forecast steps adding $R-2$ drawn residuals. Proceeding in this fashion forecast errors can be accounted for from each forecast step.

The bootstrap method is fairly easy to apply, but obviously very computationally intensive and time consuming for higher forecast horizons. When the model is correctly specified it has good properties, yielding approximately unbiased forecasts of $\mathbf{z}_{T+h}$, but can otherwise be biased. The big advantage over other methods is that it offers not only the point forecast $\widehat{\mathbf{z}}^b_{T+h}$, but also a way to obtain an estimate of the h-step ahead forecast distribution. For example, the $R-1$ values of $\widehat{\mathbf{z}}^b_{j,T+2}$ in (8) can be viewed as a drawn sample from the distribution, which in turn can be used for constructing confidence intervals, hypothesis testing, estimation of quantiles and obtaining other features of the forecast distribution that can be of interest.

4.1 Composite coincident index

In the literature there is some agreement that the best overall forecasting method is a combined, or composite, forecast as a weighted average of a variety of forecasts, each generated by a different technique, see for example Timmermann (2006) for a discussion and a survey of existing techniques. This average should prove superior to any single forecasting model because the errors in the separate forecasts will tend to cancel one another out. It has often been found that simple combinations, that is, combinations that do not require estimating many parameters, such as arithmetic averages or weights based on the inverse root mean squared forecast error (RMSFE), do better than more sophisticated methods. As shown by Gupta and Wilton (1987) the performance of equal weighted combinations depends strongly on the relative size of the variance of the forecast errors associated with the different forecasting methods. When these are similar, equal weights perform well, while when larger differences are observed, differential weighting of forecasts is generally required.

A number of different methods can therefore be considered when constructing a composite coincident index, CCI_t. This study will base the weights on the inverse RMSFE, defining the composite coincident indicator as

$$CCI_t = \boldsymbol{\omega}\mathbf{y}_t. \tag{10}$$

The elements of the vector of weights $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)$ sum to one, where each individual weight ω_i is given by

$$\omega_i = \frac{\frac{1}{\mathrm{RMSFE}_i}}{\frac{1}{\mathrm{RMSFE}_1} + \ldots + \frac{1}{\mathrm{RMSFE}_m}}, \qquad i = 1, \ldots, m. \tag{11}$$

The index will thus reflect the behaviour of the business cycle using information from all m coincident variables included in the model.

5 Empirical application

5.1 The data set

In this section the variable selection procedure of Section 2 is applied to the Icelandic economy. Monthly data is available for in total 104 macroeconomic and financial variables from a number of different sectors and markets. Since many of these variables only have been collected and reported for about ten years, the sample size is limited to the period between January 1999 and February 2006, thus giving 86 observations in total. Since the Icelandic GDP is only measured quarterly, the variables are aggregated into quarterly data in the initial process of selecting coincident variables and leading indicators.

Furthermore, to model and forecast changes in the Icelandic economy, GDP is transformed into yearly growth rate in the analysis that follows. A number of possible data transformations could be considered for the candidate variables at this point. The large technical literature concerns various methods to remove long term movements and high frequency fluctuations. However, except for using seasonally adjusted series when needed and expressing all variables in real values if necessary, the data series are only analysed in levels or in yearly growth rate. For index series and yield spread, the transformation to yearly growth rate can not be applied directly. Therefore, index series are expressed in yearly percentage growth rate, and the yield spread in yearly change.

As the data is available on a monthly basis it would be possible to transform the data into monthly, instead of yearly, growth rate. Yearly growth rate has the advantage that it decreases the effect of any possible seasonality. Using data in monthly growth rate and not accounting for possible seasonality in the modelling procedure would make the estimate of the CCI_t badly biased. The VAR model (4) does allow for modelling seasonality. The sample size of the data set available does however restrict the number of possible variables to be included in the model. Adding seasonal components would thus be more appropriate for larger sample sizes and is therefore left for later studies. Analysing the correlation structure of each candidate variable in levels, and in growth rate, with the real GDP annual growth rate, a number of potential leading indicators and coincident variables can be selected for the Icelandic economy. The chosen candidate variables and type of transformation are given in Table 1.

Table 1. Selected potential coincident variables and leading indicators, including acronyms and type of transformation.

Name	Description	Transformation
Coincident variables		
CEM	Cement sales	growth rate
ICG	Import of consumer goods, semi-durable	growth rate
IFB	Import of food and beverages	growth rate
NWP	Number of new work permits	growth rate
RER	Real exchange rate, Index	% growth rate
VRA	Number of vacancies in greater Reykjavik area	growth rate
Leading indicators		
CRED	Credit cards: Total number of transactions	growth rate
EMP	Export of marine products	levels
NVR	New Vehicles Registration, SA	growth rate
OIL	Oil price (UK Brent 38), USD/barrel	growth rate
RIH	Real index of housing	% growth rate
TRAB	Trade balance	levels
TREB	Treasury Bonds indices 1 year	growth rate
WS	Wages and Salaries, Index	% growth rate
YBN	Yield spread, Tr Bonds 20 year - Tr Notes 5 year	change
YNB	Yield spread, Tr Notes 5 year - Tr Bills 3 month	levels

5.2 Model specification and estimation

When specifying the model the available sample size limits the number of potential variables to use. It is therefore not possible to include all variables in Table 1 into the model. On the other hand this should not be necessary for obtaining reliable estimates of the model parameters and for the construction and forecasting of the composite coincident index CCI_t. Basing the modelling of the business cycle only on a subset of the potential variables poses, however, a question of which variables to choose. There is an obvious trade off between using as much information in the data as possible, and the problem of the dimension of the model and number of parameters to estimate.

First consider the choice of coincident variables, Table 1 contains six possible variables. Among these six variables there are two import variables, and two variables concerning employment. It would be realistic to assume, for each of these pairs of variables, that the information on the business cycle contained in one of the variables would be similar, or the same, as in the other one. It should therefore be sufficient to include only one variable from each pair into the model. Selecting between the two, the variable with the strongest correlations with the real GDP in growth rate is the most natural choice. Of the six candidates the three variables cement sales (CEM), imports of food and beverages (IFB) and vacancies in the Reykjavik area (VRA) are chosen. Originating from different sectors in the Icelandic economy, it is reasonable to believe that they will form a good basis for describing the general business cycle. The other three variables, imports of consumer goods (ICG), new work permits (NWP), and real exchange rate (RER) will be left out of the study due to their lower correlation with GDP and to their close relationship to the variables import of food and beverages, and vacancies in the Reykjavik area respectively.

The selection of leading indicators to include in the model poses a larger problem. The obvious reason to include leading indicators is to take into account early information of the behaviour of the business cycle. In general, using a single leading indicator can not be recommended since economic theory and experience suggest that recessions and expansions can have a number of different sources and characteristics. More indicators included in the model implies, however, a larger and more complex model to specify and estimate. Considering this trade off between the number of parameters to be estimated and early information of the business cycle contained in the leading indicators, the number of indicators is set to four in the analysis that follows. When selecting a set of four leading indicators from the ten potential leading indicators in Table 1, there are as many as 210 possible combinations to choose from. To evaluate each of these combinations is very time consuming, at present it has not been possible to automate the whole modelling process, which is why only a small fraction of the combinations have been considered in this study. Choosing different sets of four leading indicators from various sectors in the economy, only six combinations have been thoroughly analysed. The selection between the different sets of indicators is based on two criteria. First, they are all evaluated according to how strongly the individual leading indicators are correlated both with the real GDP growth rate and with the three selected coincident variables. Secondly, the model forecast performance is evaluated over a pseudo hold out sample of 9 observations. This is done by estimating the model over the sample period saving the last 9 observations for a comparison with the estimated point forecasts. The forecast performance is measured by the standard mean squared error (MSE), where the root mean squared forecast errors ($RMSFE$) summarize the performance. Note that the naive forecasting method is used at this stage for simplicity reasons. Based on these two criteria: credit cards ($CRED$), new vehicles registration (NVR), real index of housing (RIH), and the yield spread between treasure bonds 20 years and treasury notes 5 years (YBN), are chosen as the leading indicators to be included in the model. Table A1 in the Appendix shows the forecasts of the coincident variables, the leading indicators, their MSE and RMSFE for the pseudo hold out sample of 9 observations. The other five sets of analysed leading indicators, not reported here, contain various combinations of the four indicators from above and the variables; trade balance ($TRAB$), wages and salaries (WS), oil price (OIL), and yield spread between 5 year treasure notes and 3 month treasure bills (YNB).

An initial lag structure of the model (4) is determined by analysing the correlation structure between the coincident variables and the leading indicators, and between the individual leading indicators. Estimation is then performed by GLS using the standard backward elimination of variables, one at a time, with the most nonsignificant parameter value. The final specification and parameter estimates are reported in Table A2 in the Appendix.

5.3 Evaluation

To determine the quality of the estimated model it is necessary to perform adequate evaluation tests, and to analyse the model properties. Whether it should be used for forecasting or policy evaluation, any implementation of a model following the estimation step will depend on the estimated parameters. Finding out whether or not the model appears to satisfy the assumptions under which it was estimated is therefore an important part of any modelling exercise. Several diagnostic tests and simulation methods are applied in this section to confirm that the model assumptions are not violated, and that the estimated model has desirable properties such as a mean reverting behaviour. Tables and figures with the most important results of the evaluation of the estimated VAR model can be found in the appendices.

Table A3 shows the results of the univariate Jarque and Bera (1980) tests of normality of each of the residual vectors. All individual p-values are high, suggesting that the assumption of a normally distributed model error vector is satisfied.

To test the possible presence of autocorrelation in the residuals, each residual vector is regressed against a constant and q lags of itself. The joint hypothesis that all parameters except the constant are zero is then tested for $q = 1, \ldots, 6$. The p-values are reported in Table A4, which then contains 42 separate tests. Given a test level of 5%, one of the tests show very mild significance. The individual p-value is 0.042, for the residual ε_t^{VRA} with 3 lags. All other p-values for the other lags are high, indicating that this mild autocorrelation can be neglected since it is not strong enough to affect the joint hypothesis that the whole error vector is serially uncorrelated. It is therefore safe to conclude that the assumption of a serially uncorrelated error vector is not rejected.

It is sometimes argued that misspecification of a model can make the error variances time-varying. Testing if the model error variances are heteroskedastic is therefore performed using a multivariate Lagrange multiplier test, see Eklund and Teräsvirta (2007) for specific details of the test. Table A5 presents the results of the LM test, where two time-varying variance specifications are considered as alternatives to constant variances. The first case is the one of conditional heteroskedasticity, that is, a case of ARCH(q) time-varying variances, $q = 1, 3, 5$. The second case, called the smooth transition case, is an alternative where the variances are allowed to change monotonically over time between two extreme regimes. The LM p-value is the value implied by the asymptotic properties of the test, while the bootstrap p-value is estimated using resampling which corrects for any possible size distortion present. The reported p-values show that no heteroskedasticity is present in the residual vectors of the estimated model, implying that the hypothesis of constant variances appears to be correct.

To investigate the mean reverting properties of the estimated model, a realization of 10 000 observations of $\mathbf{z}_t$ in (4) is generated from the model conditional on the parameter estimates. Any possible serious nonstationary features of the model can then effectively be revealed. Table A6 presents resulting values of the minimum, mean, maximum, and standard deviation of the simulated realization, showing that all the individual time series in $\mathbf{z}_t$ evolve over time within lower and upper bounds and, furthermore, with low standard deviations. There is thus no indication that the realization of $\mathbf{z}_t$ would be nonstationary, why these results are a strong indication that the model is stationary and mean reverting.

The same conclusion, that the model appears to be stationary, can be stated when analysing the impulse-response function of the VAR model. Figures B1 and B2 show the impulse-response function of the coincident variables, leading indicators and the composite coincident index CCI_t when a unit shock is introduced into the model via the error ε_t^{IFB} of the equation of the variable IFB at time point $t = 0$. As the shock is introduced into the system, depending on the lag structure of each of the equations of the model, each variable reacts in a consequence of its relationship with the other variables. As shown in the figures, it is clear that the shock fades out and becomes negligible after about three years, that is after 36 observations. Shocks into the model through the other error terms of the model, not reported in the paper, have similar characteristics showing a diminishing effect over time on the variables, even though a shock to the error of the equation of RIH tends to take more time to fade out, about another year. There is therefore no evidence indicating that shocks to any of the equations will cause a permanent effect on the variables as would be the case if any of the series would contain a unit root or have another form of nonstationary. It is therefore safe to conclude that the estimated VAR model is stationary and mean reverting.

5.4 Forecasting the composite coincident index CCI

The evaluation of the VAR model in Section 5.3 shows a good fit with the general model assumptions. The estimated composite coincident index CCI_t should therefore be a good representation of the cyclical behaviour of the Icelandic economy. Since it is constructed as a weighted average of the coincident variables it is only a reflection of the economic situation, not an estimate of the size of the economy or the magnitude of growth. As such, it is only a representation of the cyclical dynamics of GDP and can thus not be used to predict the actual GDP growth.

Figure 1 shows the estimated composite coincident index and the real GDP yearly growth rate over the sample period considered as a solid line and bars respectively. In the figure CCI_t appears to be a good fit of the cyclical behaviour in GDP. The estimated correlation between the two variables is high, 0.75, which also confirms this finding. As the CCI_t closely follows the evolution of GDP over time, it is interesting to note the period with the downward trend of CCI_t starting around September 2001. Following the slow decrease of GDP growth rate, CCI_t clearly reaches a trough in March 2002 several months before the one in GDP is reached. The turn upward of the Icelandic GDP growth rate between December 2002 and March 2003 is very rapid, in contrast with the more smooth increase of CCI_t which already started in March 2002. This asymmetric behaviour of GDP suggests that the VAR model might be unable to properly capture all the dynamics of the business cycle, and possibly some nonlinear model would be preferred in the modelling procedure. This was also proposed by Pétursson (2000), even though the study used yearly data over an earlier sample period. Asymmetric features were also reported in Eklund (2007), showing that a recession in the economy is easier to predict than an expansion. Nevertheless, besides not capturing the turning points of the cycle properly, the CCI_t has a high correlation with GDP and should therefore be considered a good measure of the general business cycle.

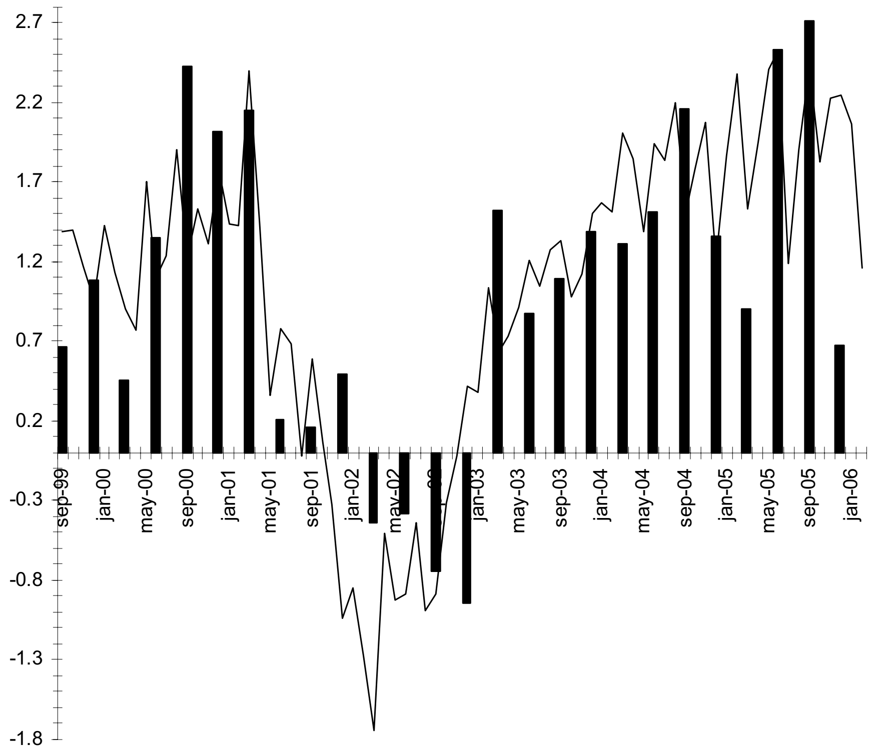

Figure 1. Estimated CCI_t (line), and real GDP in yearly growth rate (bars). Both variables standardized having a unit variance.

Figure 2 shows the point forecasts of CCI_t as a white line plotted over the nine months forecasting horizon. The last few observations from Figure 1 are also included, where February 2006 is the last observation in the sample. The bootstrap confidence intervals are also included in the figure and are estimated as described in Section 4. At each forecast step, each area in the different shades of grey contains ten per cent of the observations of the empirical forecasting distribution, while the black area in the centre contains twenty per cent. Together the different areas cover the whole forecasting range of the CCI_t and can be viewed as a contour plot of the distribution. As such, they can be used directly for hypothesis testing. For example, the range of the fifth forecast step, July 2006, covers values from about 1.20 till 2.05, clearly indicating that CCI_t is positive. This implies a strong growth in GDP and further tendencies of an expansion of the Icelandic economy at this point in time. Note that the point forecasts, the white line, are located close to the centre of the distribution, the black area, which mid-point is the median of the forecast distribution. The distribution is thus close to symmetric at the forecast steps, showing no asymmetric behaviour of the forecasts.

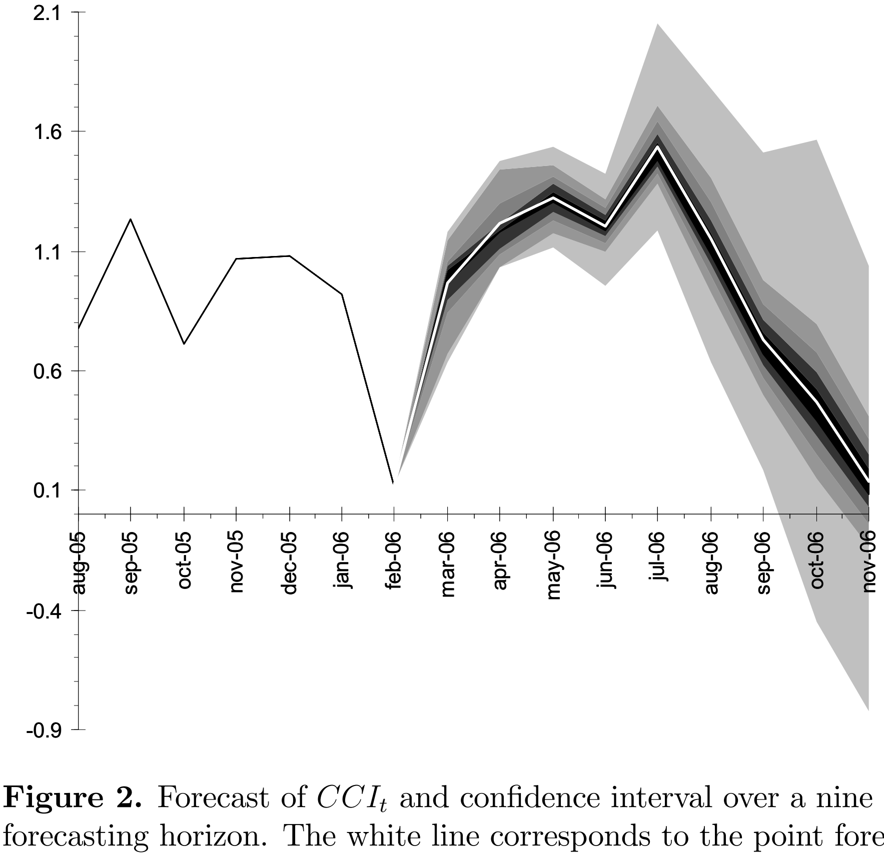

Figure 2. Forecast of CCI_t and confidence interval over a nine months forecasting horizon. The white line corresponds to the point forecasts, each area in grey shades covers 10% of the forecast distributions, and the black area in the centre covers 20%.

It is clear from Figure 2 that not until in the last two forecast steps, the forecast ranges of the CCI_t all are located above zero, indicating strong growth of Icelandic GDP over this time period. The first time the forecast range covers zero is in October 2006, where between five and ten per cent of the observations are negative. Given a hypothesis test on, for example, the 5% level, testing if CCI_t is equal to zero would then not be rejected. This result would suggest mild growth of GDP this month, or possibly no growth at all. In November 2006 this is more accentuated where between 20 and 30 per cent of the observations are negative, indicating an even milder growth of GDP or an ever larger possibility of no or negative growth. These results indicate that the Icelandic economy will continue to have a strong growth of GDP over the first three quarters of 2006, while the economy shows some tendencies to move towards a recession some time in the last months of 2006.

6 Conclusions

This paper has described a simple method for selecting coincident variables and leading indicators based on the analysis of the correlation structure between candidate variables and the yearly growth rate of GDP. This method has been applied to the Icelandic economy, where several coincident variables and leading indicators have been identified. The number of possible combinations of coincident and leading variables that can be included in the modelling procedure is very large. Only a handful of variables have therefore been analysed thoroughly, leaving the majority of the possible combinations for later studies.

To model the relationship between coincident and leading variables, and to construct a composite coincident index, CCI_t, representing the Icelandic business cycle dynamics, a vector autoregressive, VAR, model has been specified and estimated. Evaluated over a pseudo hold out sample, weights for constructing the CCI_t have been estimated based on the inverse of the RMSFE. The evaluation of the model shows no violation of model assumptions, showing that the cyclical behaviour of the Icelandic economy is satisfactory estimated by the CCI_t, also indicated by the high correlation between the index and the yearly GDP growth rate.

In a final exercise, point forecasts and confidence intervals of the CCI_t are estimated using a bootstrap forecasting method over a nine months forecasting horizon. The method is very time consuming and computationally intensive but produces not only a point forecast, but also a realization of the forecast distribution. This realization can serve as a basis for constructing forecast confidence intervals, and for hypothesis testing. Analyzing the results from the forecasting procedure, there is a strong indication that the Icelandic economy will continue to have a strong growth of the GDP over the first three quarters of 2006. Later, in October and November 2006, the economy shows some tendencies of moving towards a recession.

References

Burns, A. F., and W. C. Mitchell (1946): "Measuring business cycles," *NBER Studies in Business Cycles no. 2 (New York).*

Daníelsson, Á., L. Elíasson, M. Gudmundsson, B. Hauksson, R. Jónsdóttir, T. Ólafsson, and T. Pétursson (2006): "QMM: A quarterly macroeconomic model of the Icelandic economy," Central Bank of Iceland, Working Papers No. 32.

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

Eklund, B., and T. Teräsvirta (2007): "Testing the constancy of the error covariance matrix in vector models," *Journal of Econometrics*, Forthcoming.

Granger, C. W. J., and T. Teräsvirta (1993): *Modelling nonlinear economic relationships.* Oxford, Oxford University Press.

Gupta, S., and P. C. Wilton (1987): "Combination of Forecasts: An Extension," *Management Science*, 33, 356–372.

Hamilton, J. D. (1994): *Time Series Analysis.* Princeton, Princeton University Press.

Jarque, C. M., and A. K. Bera (1980): "Efficient tests for normality, homoskedasticity and serial independence of regression residuals," *Economics Letters*, 6, 255–259.

Magnus, J. R. (1978): "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix," *Journal of Econometrics*, 7, 281–312.

Marcellino, M. (2006): "Leading indicators: What have we learned?," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann. Elsevier.

Moore, G. H., and J. Shiskin (1967): "Indicators of business expansions and contractions," NBER Occasional Paper, No. 103.

Pétursson, T. G. (2000): "Business Cycle Forecasting and Regime Switching," Central Bank of Iceland, Working Papers No. 7.

Sims, C. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Stock, J. H., and M. W. Watson (1989): "New indexes of coincident and leading economic indicators," in *NBER Macroeconomics Annual*, ed. by O. Blanchard, and S. Fischer, pp. 351–394. MIT Press.

Teräsvirta, T. (2006): "Forecasting Economic Variables with Nonlinear Models," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann. Elsevier.

Timmermann, A. (2006): "Forecast Combinations," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann. Elsevier.

Appendix A. Tables

Table A1. Forecasts, mean squared errors (MSE), and root mean squared forecast error (RMSFE) of coincident variables and chosen leading indicators over a hold out sample.

	Coincident variables			Leading indicators				
h	CEM_{T+h}	IFB_{T+h}	VRA_{T+h}	$CRED_{T+h}$	NVR_{T+h}	RIH_{T+h}	YBN_{T+h}	CCI_{T+h}
1	1.811	1.483	0.697	0.472	2.299	2.084	−0.919	1.324
MSE	0.348	0.123	0.683	1.504	0.987	0.071	0.017	0.158
2	1.481	1.134	1.077	1.174	1.693	1.395	−1.261	1.254
MSE	0.630	1.777	1.025	0.744	0.002	0.501	0.149	0.333
3	1.139	0.497	0.488	−0.036	2.001	1.904	−1.100	0.754
MSE	0.059	2.030	0.023	0.395	0.156	0.147	0.126	0.114
4	1.873	1.418	0.793	0.188	2.154	1.761	−1.505	1.370
MSE	0.865	0.758	0.272	0.148	0.208	0.075	0.439	0.217
5	1.752	2.329	0.923	1.347	2.104	1.655	−1.182	1.573
MSE	0.284	8.455	0.112	0.700	0.280	0.250	0.258	0.483
6	2.784	2.415	0.713	0.603	2.165	1.688	−1.398	1.933
MSE	3.552	0.477	0.263	0.098	0.315	0.214	0.323	0.644
7	2.195	2.598	0.681	0.291	1.642	1.576	−1.229	1.723
MSE	0.391	5.652	0.259	0.400	0.036	0.286	0.761	0.381
8	2.959	2.282	0.941	0.536	1.741	1.579	−1.360	2.058
MSE	0.171	0.239	3.350	0.015	1.248	0.120	1.006	0.502
9	3.042	2.108	0.665	1.129	1.738	1.423	−1.427	1.951
MSE	1.403	5.971	3.195	0.024	1.241	0.011	1.146	0.968
RMSFE	0.979	1.737	1.064	0.669	0.705	0.432	0.685	0.650

Table A2. Model parameters estimated by generalized least squares, standard deviations of each parameters below in parenthesis.

$$
\begin{aligned}
\mathrm{CEM}_t = {} & \underset{(0.002)}{0.160} + \underset{(0.002)}{0.048}\mathrm{IFB}_{t-3} - \underset{(0.005)}{0.334}\mathrm{VRA}_{t-5} + \underset{(0.007)}{0.362}\mathrm{CEM}_{t-1} \\
& + \underset{(0.012)}{0.142}\mathrm{NVR}_{t-4} + \underset{(0.011)}{0.248}\mathrm{NVR}_{t-5} + \underset{(0.026)}{0.248}\mathrm{RIH}_{t-1} - \underset{(0.062)}{0.706}\mathrm{RIH}_{t-3} \\
& + \underset{(0.048)}{0.450}\mathrm{RIH}_{t-4} + \underset{(0.016)}{0.480}\mathrm{RIH}_{t-8} - \underset{(0.005)}{0.191}\mathrm{YBN}_{t-4} + \varepsilon_t^{\mathrm{CEM}}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{IFB}_t = {} & \underset{(0.005)}{0.111} - \underset{(0.009)}{0.122}\mathrm{IFB}_{t-1} + \underset{(0.009)}{0.332}\mathrm{NVR}_{t-4} - \underset{(0.013)}{0.391}\mathrm{CRED}_{t-6} \\
& + \underset{(0.013)}{0.150}\mathrm{CRED}_{t-7} + \underset{(0.018)}{0.407}\mathrm{RIH}_{t-7} - \underset{(0.019)}{0.344}\mathrm{YBN}_{t-3} + \varepsilon_t^{\mathrm{IFB}}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{VRA}_t = {} & -\underset{(0.003)}{0.208}\mathrm{IFB}_{t-5} + \underset{(0.007)}{0.416}\mathrm{VRA}_{t-1} + \underset{(0.005)}{0.148}\mathrm{CEM}_{t-4} + \underset{(0.004)}{0.075}\mathrm{NVR}_{t-6} \\
& + \underset{(0.005)}{0.264}\mathrm{CRED}_{t-5} + \underset{(0.004)}{0.201}\mathrm{CRED}_{t-8} - \underset{(0.010)}{0.174}\mathrm{YBN}_{t-1} + \underset{(0.008)}{0.115}\mathrm{YBN}_{t-3} + \varepsilon_t^{\mathrm{VRA}}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{CRED}_t = {} & -\underset{(0.004)}{0.232}\mathrm{IFB}_{t-5} - \underset{(0.007)}{0.218}\mathrm{VRA}_{t-6} - \underset{(0.007)}{0.235}\mathrm{CEM}_{t-1} + \underset{(0.009)}{0.474}\mathrm{NVR}_{t-5} \\
& - \underset{(0.009)}{0.156}\mathrm{CRED}_{t-1} + \underset{(0.006)}{0.303}\mathrm{CRED}_{t-2} + \underset{(0.007)}{0.284}\mathrm{CRED}_{t-3} + \underset{(0.024)}{0.176}\mathrm{YBN}_{t-1} \\
& - \underset{(0.033)}{0.252}\mathrm{YBN}_{t-2} - \underset{(0.023)}{0.354}\mathrm{YBN}_{t-3} + \varepsilon_t^{\mathrm{CRED}}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{NVR}_t = {} & \underset{(0.002)}{0.049}\mathrm{IFB}_{t-2} + \underset{(0.003)}{0.206}\mathrm{VRA}_{t-3} - \underset{(0.004)}{0.114}\mathrm{CEM}_{t-1} + \underset{(0.004)}{0.596}\mathrm{NVR}_{t-1} \\
& + \underset{(0.006)}{0.498}\mathrm{NVR}_{t-4} + \underset{(0.003)}{0.209}\mathrm{CRED}_{t-1} - \underset{(0.002)}{0.137}\mathrm{CRED}_{t-4} - \underset{(0.003)}{0.122}\mathrm{CRED}_{t-5} \\
& - \underset{(0.005)}{0.180}\mathrm{RIH}_{t-2} - \underset{(0.005)}{0.129}\mathrm{RIH}_{t-8} + \varepsilon_t^{\mathrm{NVT}}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{RIH}_t = {} & \underset{(0.001)}{0.032} + \underset{(0.001)}{0.105}\mathrm{VRA}_{t-1} - \underset{(0.001)}{0.128}\mathrm{CEM}_{t-1} + \underset{(0.002)}{0.223}\mathrm{NVR}_{t-3} \\
& - \underset{(0.002)}{0.060}\mathrm{NVR}_{t-5} + \underset{(0.001)}{0.087}\mathrm{CRED}_{t-1} + \underset{(0.011)}{0.834}\mathrm{RIH}_{t-1} + \underset{(0.021)}{0.221}\mathrm{RIH}_{t-2} \\
& - \underset{(0.009)}{0.249}\mathrm{RIH}_{t-3} + \underset{(0.003)}{0.152}\mathrm{YBN}_{t-2} - \underset{(0.003)}{0.154}\mathrm{YBN}_{t-3} + \varepsilon_t^{\mathrm{RIH}}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{YBN}_t = {} & -\underset{(0.001)}{0.031}\mathrm{IFB}_{t-2} - \underset{(0.003)}{0.097}\mathrm{VRA}_{t-1} - \underset{(0.002)}{0.202}\mathrm{VRA}_{t-5} + \underset{(0.002)}{0.102}\mathrm{CEM}_{t-1} \\
& - \underset{(0.002)}{0.234}\mathrm{NVR}_{t-1} - \underset{(0.002)}{0.144}\mathrm{CRED}_{t-2} - \underset{(0.002)}{0.189}\mathrm{CRED}_{t-3} + \underset{(0.019)}{0.426}\mathrm{RIH}_{t-1} \\
& - \underset{(0.019)}{0.201}\mathrm{RIH}_{t-2} + \underset{(0.004)}{0.498}\mathrm{YBN}_{t-1} + \underset{(0.004)}{0.278}\mathrm{YBN}_{t-4} - \underset{(0.003)}{0.191}\mathrm{YBN}_{t-7} + \varepsilon_t^{\mathrm{YBN}}
\end{aligned}
$$

Table A3. Univariate Jarque-Bera tests of normality of residual vectors.

	ε_t^{CEM}	ε_t^{IFB}	ε_t^{VRA}	ε_t^{CRED}	ε_t^{NVR}	ε_t^{RIH}	ε_t^{YBN}
Skewness:	0.120	0.245	−0.039	0.231	0.012	−0.312	−0.005
Kurtosis:	2.630	3.220	3.232	3.301	3.500	2.829	2.275
JB value:	0.633	0.940	0.194	0.990	0.816	1.364	1.709
p-value:	0.729	0.625	0.907	0.610	0.665	0.505	0.425

Table A4. Univariate test of no autocorrelation against errors described by an AR(q) process. Each residual vector is regressed against a constant and itself with q lags. The presented p-values result from testing the hypothesis that all parameters of the lags equal zero.

q	ε_t^{CEM}	ε_t^{IFB}	ε_t^{VRA}	ε_t^{CRED}	ε_t^{NVR}	ε_t^{RIH}	ε_t^{YBN}
1	0.563	0.256	0.059	0.438	0.718	0.144	0.099
2	0.367	0.518	0.134	0.182	0.192	0.337	0.168
3	0.140	0.128	0.042	0.271	0.200	0.362	0.213
4	0.159	0.176	0.086	0.399	0.308	0.065	0.312
5	0.189	0.256	0.143	0.543	0.251	0.097	0.409
6	0.480	0.194	0.233	0.566	0.272	0.154	0.549

Table A5. Testing constancy of the error covariance matrix, see Eklund and Teräsvirta (2007), against time varying variances specified by:

	ARCH(q)			Smooth transition
	$q=1$	$q=3$	$q=5$	
LM test p-value:	0.288	0.563	0.626	0.656
Bootstrap p-value:	0.180	0.450	0.520	0.660

Table A6. Simulated realization: 10000 observations.

	CEM_t	IFB_t	VRA_t	CRED_t	NVR_t	RIH_t	YBN_t
Min:	−5.244	−3.839	−3.419	−4.636	−3.743	−3.940	−3.673
Mean:	0.141	0.002	−0.175	−0.223	−0.024	−0.137	0.284
Max:	5.060	4.588	3.543	3.690	3.825	3.460	4.463
Stdv:	1.389	1.049	1.032	1.031	1.058	1.021	1.236

Appendix B. Figures

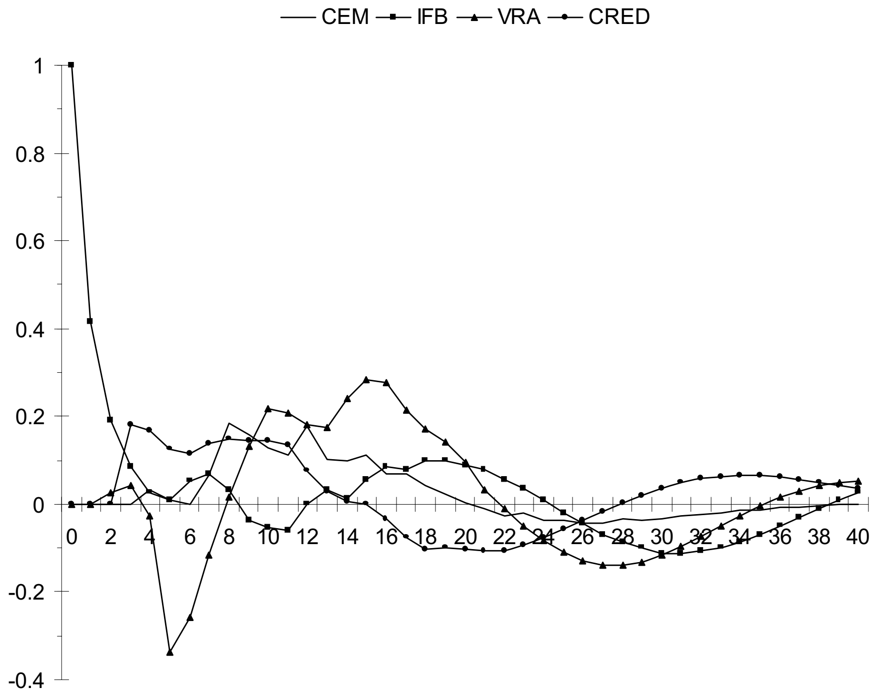

Figure B1. Impulse response over 35 months on the yearly growth rate of the state of the economy and on the leading indicators of a unit shock to ε_t^{CCI}, the error of CCI_t.

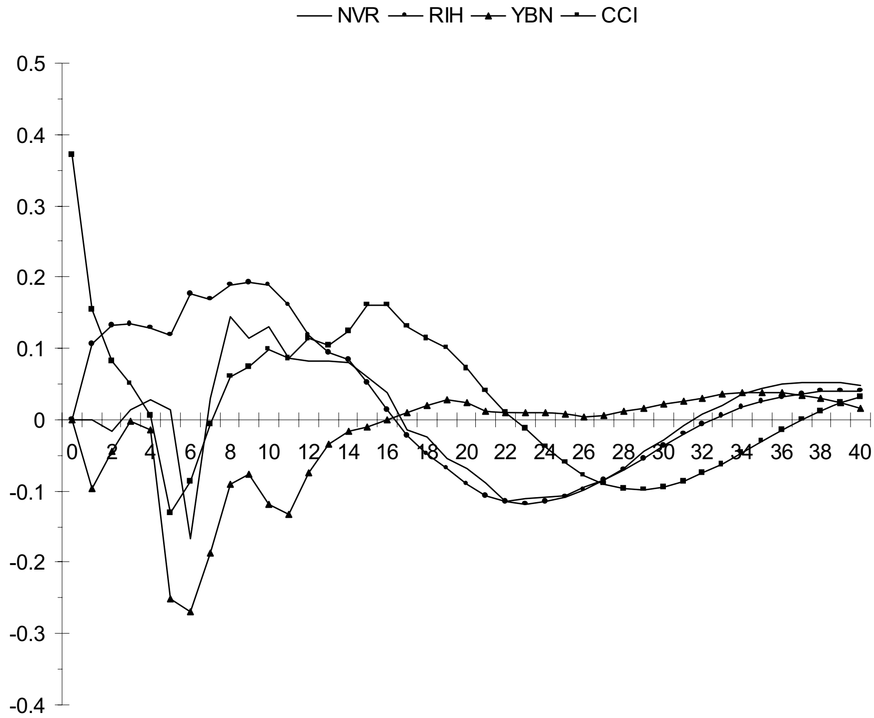

Figure B2. Impulse response over 35 months on the coincident variables of a unit shock to ε_t^{CCI}, the error of CCI_t.