CENTRAL BANK OF ICELAND

WORKING PAPERS No. 33

Predicting recessions with leading indicators: An application on the Icelandic economy

by
Bruno Eklund

January 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

Predicting recessions with leading indicators: An application on the Icelandic economy

Bruno Eklund*
Central Bank of Iceland

Abstract

This paper focuses on the Stock and Watson methodology to forecast the future state of the business cycle in the Icelandic economy. By selecting variables available on a monthly basis that mimic the cyclical behaviour of the quarterly GDP, coincident and leading variables are identified. A factor model is then specified based on the assumption that a single common unobservable element drives the cyclical evolution of many of the Icelandic macroeconomic variables. The model is cast into a state space form providing a simple framework both for estimation and for predicting the future recession and expansion patterns. Based on the bootstrap resampling technique, a simple approach to estimate recession and expansion probabilities is developed. This method is completely nonparametric compared to the semi-parametric approach used by Stock and Watson.

Keywords: Leading indicators, Factor model, Kalman filter, Forecasting, Bootstrap, Recession probabilities.
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

Financial and monetary policy decisions have to be based on reliable knowledge of a large number of different economic variables and in particular on the business cycle. Central banks and markets pay special attention to the release of certain data, either because they can convey leading information on key variables, such as GDP or unemployment, or because they are inputs in estimates of such variables. The interest of developing and improving on methods of predicting the future economic conditions has therefore grown rapidly through the years. Methods concerning leading indicators have attracted considerable attention, ranging from the choice and evaluation of the indicators, to methods relating them to the general business cycle.

Forecasting of the business cycle, defined by Burns and Mitchell (1946) as representing comovements in a broad set of macroeconomic variables such as output, employment, and sales, has been one of the core problems in the literature. One possible method to study aggregate economic fluctuations is to choose a single important economic time series, for example industrial production or employment, as the object of interest for the subsequent analysis and forecasting. However, this approach is rather limited, since individual economic time series measure more or less well defined concepts, such as tons of steel produced or value of exported goods, not the state of the economy directly. Other methods have therefore been developed taking into account the behaviour of several economic time series at once, sometimes by constructing a single economic index from a number of variables.

In an important contribution to the practice in indicator analysis at that time, Stock and Watson (1989) formalised the early work of Burns and Mitchell using modern econometric techniques. They specified a factor model based on the assumption that comovements in many macroeconomic variables have a common element that can be captured by a single underlying unobservable variable. This variable, referred to as the state of the economy, is then by construction an economic index coincident with the general business cycle. Rather than basing the economic index on a weighted average of individual economic variables, as the more non-model based approach of the NBER business cycle dating committee, Stock and Watson used historical changes in this coincident economic index as well as other variables that lead the business cycle.

In general, if a coincident index truly reflects the state of the economy, a good forecast of this coincident index should naturally make a good leading index of the economic activity, see Stock and Watson (1991), and Stock and Watson (1993) for refinements of the methodology, discussion and further details. For an up to date survey on the construction, use and evaluation of leading indicators see Marcellino (2006).

This paper describes the steps of selecting coincident variables and leading indicators, and the specification of a factor model in state space form. The methodology is applied on the Icelandic economy, where potential monthly coincident variables and leading indicators are identified from a large set of macroeconomic and financial time series. The model is estimated by maximum likelihood through the Kalman filter, and used as a base for forecasting the state of the economy. As a final step, recession and expansion probabilities are estimated over the sample and forecast period considered.

Except for two stages, the outline of the empirical application will follow the Stock and Watson methodology. First, the state of the economy is estimated in a single estimation step where both coincident variables and leading indicators are added into the model. This should be compared with the two step estimation method used by Stock and Watson. Secondly, when estimating recession and expansion probabilities, an alternative estimation technique is developed based on bootstrap resampling. The advantage of this method over the semi-parametric approach proposed by Stock and Watson is that no extra parameters need to be estimated.

The Icelandic business cycle has previously been modelled using a Markov switching model, see Pétursson (2000) for details, but in contrast to this study only in a univariate setting and using yearly GDP.

The plan of the paper is as follows. Section 2 discusses methods of selecting potential coincident variables and leading indicators. The factor model is described in Section 3, while Section 4 presents the bootstrap method developed for estimating recession and expansion probabilities. Section 5 applies the methods and techniques outlined on the Icelandic economy. Final conclusions can be found in Section 6.

2 Selection of coincident and leading variables

If GDP would have been available on a monthly basis, it could provide a reliable summary of the current state of the economy. Since Icelandic GDP only is measured quarterly, it can not be used directly in the modelling process. Instead, candidate variables are selected, available on a monthly basis, that mimic the cyclical behaviour of GDP. These potential variables will in turn be classified as coincident, lagging or leading. Any of the coincident variables could be used as a proxy of GDP, but are usually used to form a single coincident economic index. 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 analyzing the correlation structure between the GDP and the candidate variable taken from a list of macroeconomic variables. Candidate variables can originate from any possible source available since we are interested in variables having a high correlation with the cyclical behaviour of GDP. A more formalized scoring system for selecting variables is due to the often quoted Moore and Shiskin (1967), where a number of criteria such as consistent timing, conformity to the business cycle, economic significance and more, are defined.

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 the GDP, while the behaviour of a coincident variable and the GDP, as the name suggests, coincides over time. Note that variables with a high negative correlation with GDP also can serve as good indicators of the evolution of the general business cycle, and should be added to the list of potential coincident and leading variables. Other possible selection criteria can be used to choose variables on the basis of spectral coherence or the lead time in turning points, but these methods are not considered in this paper.

Note that while the leading indicators may give early indications about changes in the direction of the business cycle, they do not provide any reliable information about the magnitude of that change. Also, applying leading indicators is many times difficult in practice because the lag relationship tends to be quite volatile. Ironically, the widespread use of a reasonably reliable leading indicator may, in fact, lead to less reliability in the indicator over time. This may happen if 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, a leading indicator series can help economists, business and government predict and prepare for significant changes in the economic environment.

3 The model

Developed by Geweke (1977) and Sargent and Sims (1977), the dynamic factor models became well known with the publication of Stock and Watson (1989) attempting to formalise a probabilistic base for the coincident and leading indicators considered by Burns and Mitchell (1946). Following the general idea outlined by Stock and Watson, assume that the cyclical evolution of the economy can be described by a single underlying unobservable factor

c_t, the state of the economy. This variable is thus the only source for any cyclical movements in the economy, and drives the evolution of the macroeconomic and the financial variables that exhibit business cyclical behavior. The coincident variables and the leading indicators are thus dependent on c_t, where c_t, in turn, is allowed to depend on lagged values of both itself and the leading indicators. The following general model can be defined:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu}_y + \boldsymbol{\Lambda}_{yc}(L)\, c_t + \mathbf{u}_t && (1) \\ \mathbf{x}_t &= \boldsymbol{\mu}_x + \boldsymbol{\Lambda}_{xc}(L)\, c_t + \boldsymbol{\Lambda}_{xx}(L)\, \mathbf{x}_t + \boldsymbol{\varepsilon}_{xt} && (2) \\ c_t &= \mu_c + \lambda_{cc}(L)\, c_t + \boldsymbol{\Lambda}_{cx}(L)\, \mathbf{x}_t + \varepsilon_{ct} && (3) \\ \mathbf{u}_t &= \boldsymbol{\Lambda}_{uu}(L)\, \mathbf{u}_t + \boldsymbol{\varepsilon}_{yt}, && (4) \end{aligned}$$

where $\mathbf{y}_t$ is a vector of coincident variables, $\mathbf{x}_t$ is a vector of leading indicators, and c_t is the state of the economy, $t = 1, \ldots, T$. Furthermore, $\boldsymbol{\mu}_y$, $\boldsymbol{\mu}_x$, and μ_c are vector and scalar means respectively, $\boldsymbol{\Lambda}_{yc}$, $\boldsymbol{\Lambda}_{xc}$, and $\boldsymbol{\Lambda}_{cx}$ are lag polynomial vectors, $\boldsymbol{\Lambda}_{xx}$ and $\boldsymbol{\Lambda}_{uu}$ are lag polynomial matrices, where $\boldsymbol{\Lambda}_{uu}$ is assumed to be diagonal, and λ_{cc} is a lag polynomial. Shocks to the equation system enter the model through the error vectors $\boldsymbol{\varepsilon}_{yt}$, $\boldsymbol{\varepsilon}_{xt}$, and scalar error ε_{ct}. Finally, equation (4) is added to take into account any possible autocorrelation in $\mathbf{u}_t$, the errors of the coincident variables. Since $\boldsymbol{\Lambda}_{uu}$ is diagonal these errors are uncorrelated, implying that a shock to one of the coincident variables does not affect any of the others. Note that the equation for c_t is a scalar valued equation, while the other three are vector valued.

Instead of specifying the whole model as above, Stock and Watson (1989) apply a simplified two step estimation procedure. First estimating the state of the economy, c_t, using equations (1), (3) and (4) under the assumption that $\boldsymbol{\Lambda}_{cx}(L) = 0$. No leading indicators are thus assumed to contain any helpful information for estimating the current state of the economy. In the second step the leading indicators are added into the analysis, and the parameters in equations (1) and (2) are estimated conditionally on the parameter values and the state of the economy from step one. Such a estimation procedure is robust to misspecification of the restricted model in the second step, but can be inefficient when either the whole model (1)-(4) is correctly specified or, at least, when lags of the leading indicators contain valuable information for estimating the current state of the economy, see Marcellino (2006) for a discussion. Since expecting the leading indicators not to have any such valuable information would in practise be unrealistic and overly restrictive, only the whole model as given in equation (1)-(4) above will be considered in the analysis that follows. Also, the relationships between the leading indicators and the coincident variables are more clear than using the two step estimation method.

Reformulating the model in state space form, allows to express it as:

$$\begin{aligned} \mathbf{z}_t &= \mathbf{G}\mathbf{h}_t + \mathbf{A}\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t \\ \mathbf{h}_t &= \mathbf{F}\mathbf{h}_{t-1} + \mathbf{B}\boldsymbol{\xi}_t + \mathbf{w}_t, \end{aligned} \tag{5}$$

where the vector $\mathbf{z}_t$ consists of coincident variables $\mathbf{y}_t$ and leading indicators $\mathbf{x}_t$, $\boldsymbol{\xi}_t$ is a vector of lagged leading indicators, $\boldsymbol{\varepsilon}_t$ and $\mathbf{w}_t$ are error vectors, and $\mathbf{h}_t$ is the state vector consisting of the state of the economy c_t and $\mathbf{u}_t$. Finally, $\mathbf{G}$, $\mathbf{A}$, $\mathbf{F}$, and $\mathbf{B}$ are the corresponding parameter matrices of the model. The model can be estimated, using the Kalman filter, under the assumption that the error vectors, $\boldsymbol{\varepsilon}_t$ and $\mathbf{w}_t$, are independent, serially uncorrelated and multivariate normally distributed, see for example Harvey (1989) or Hamilton (1994) for details on properties, estimation and forecasting using the Kalman filter.

Forecasting s steps ahead based on the state space representation (5) can easily be attained by calculating s single step forecasts. Each variable in the model has an estimated data generating process, that is, all included variables are represented on the left hand side in (5). Any value of the number of forecast steps s can therefore be considered, using the previous point forecasts in the following forecast step if necessary.

4 Estimating probabilities using Bootstrap resampling

Estimating recession and expansion probabilities is an intuitively appealing way to summarize the features and tendencies in the economy that are given by the estimated variable c_t, the state of the economy. This section describes a simple technique to estimate these probabilities based on bootstrap resampling. The developed method does in fact generate realizations of the empirical distribution of the state of the economy at each time point, and the location of the observed state in this distribution forms a basis for estimating the probabilities.

Conditionally on the estimated parameters of the model, $\widehat{\mathbf{G}}$, $\widehat{\mathbf{A}}$, $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{B}}$, it is a simple task to resample the estimated state of the economy $\widehat{c}_t$. To generate a bootstrap realization of $\widehat{c}_t$, $T + s$ errors are drawn and inserted into the estimated model:

$$\begin{aligned} \mathbf{z}_t^b &= \widehat{\mathbf{G}}\mathbf{h}_t^b + \widehat{\mathbf{A}}\boldsymbol{\xi}_t + \widehat{\boldsymbol{\varepsilon}}_t^b \\ \mathbf{h}_t^b &= \widehat{\mathbf{F}}\mathbf{h}_{t-1}^b + \widehat{\mathbf{B}}\boldsymbol{\xi}_t + \widehat{\mathbf{w}}_t^b, \end{aligned} \tag{6}$$

where $\mathbf{z}_t^b$ and $\mathbf{h}_t^b$ are the resulting resampled bootstrap values of $\mathbf{z}_t$ and $\mathbf{h}_t$, and $\widehat{\boldsymbol{\varepsilon}}_t^b$ and $\widehat{\mathbf{w}}_t^b$ are the drawn errors. Note that we also need to condition on the forecasts of the leading indicators in $\mathbf{x}_t$. The resampled values of $\widehat{c}_t$, denoted by c_t^b, can then be collected from the resampled state vector $\mathbf{h}_t^b$. By generating a large number, R, of realization of $\widehat{c}_t$ the empirical distribution $\widehat{D}_t$ of the state of the economy can be estimated at each time point.

There are several ways the error vectors $\widehat{\boldsymbol{\varepsilon}}_t^b$ and $\widehat{\mathbf{w}}_t^b$ can be generated. If one is confident that the estimated errors are normal, or close to normal, the series $\widehat{c}_t$ can be resampled by drawing normally distributed errors. On the other hand, if the residuals are not believed to have a normal distribution, a bootstrap method can be applied. In the case of negligible autocorrelated residuals, the error vectors $\widehat{\boldsymbol{\varepsilon}}_t^b$ and $\widehat{\mathbf{w}}_t^b$ are drawn randomly one by one with replacement and with equal probability, from the residual vectors $\widehat{\boldsymbol{\varepsilon}}_t$ and $\widehat{\mathbf{w}}_t$ obtained in the model estimation. This is what usually is referred to as a simple bootstrap. If the residuals exhibit strong autocorrelation, the bootstrap method can be modified to capture the correlation structure by drawing the residuals by a moving blocks bootstrap. This method, compared to the simple bootstrap, draws blocks of a given number of consecutive residuals. The autocorrelation structure is then maintained within each block. This will take into account the presence of the autocorrelation, and result in a more robust estimate of the distribution of the state of the economy. Effects of any deviation from the normal distribution assumption will in addition be minimized when resampling from the estimated errors, see Fitzenberger (1997) for details. Note that these features of the bootstrap are not an excuse to ignore proper modelling practice. To validate the estimation procedure and the testing of model parameters, the estimated error must have properties as close to the model assumptions as possible, that is, independent, serially uncorrelated and multivariate normally distributed.

As mentioned above, the R resampled values of $\widehat{c}_t$ can be used to estimate the distribution $\widehat{D}_t$ of $\widehat{c}_t$ at time t. The location of $\widehat{c}_t$ in this empirical distribution $\widehat{D}_t$ can give us valuable information of the tendencies in the economy towards an expansion or a recession. Consider for explanatory purposes Figure 1. Assume that the estimated state of the economy, $\widehat{c}_t$, has the smooth evolution over time as depicted in the figure. The overall empirical distribution is denoted by $\widehat{D}$ which shows the density of $\widehat{c}_t$ over its probable outcomes. Intuitively, it is easy to accept the idea that when $\widehat{c}_t$ takes on a large or a small value, such that $\widehat{c}_t$ is located in any of the two tails of the distribution $\widehat{D}$, the economy is said to be in an expansion or in a recession respectively. The tendency of the economy towards, for example, a recession at time t_1 can be obtained by analyzing the location of $\widehat{c}_{t_1}$, the estimated value of $\widehat{c}_t$ at time t_1, in the empirical distribution $\widehat{D}_1$. In the figure $\widehat{c}_{t_1}$ is located in the lower half of $\widehat{D}_1$ indicating that the tendency towards a recession is higher than towards an expansion at this particular time point.

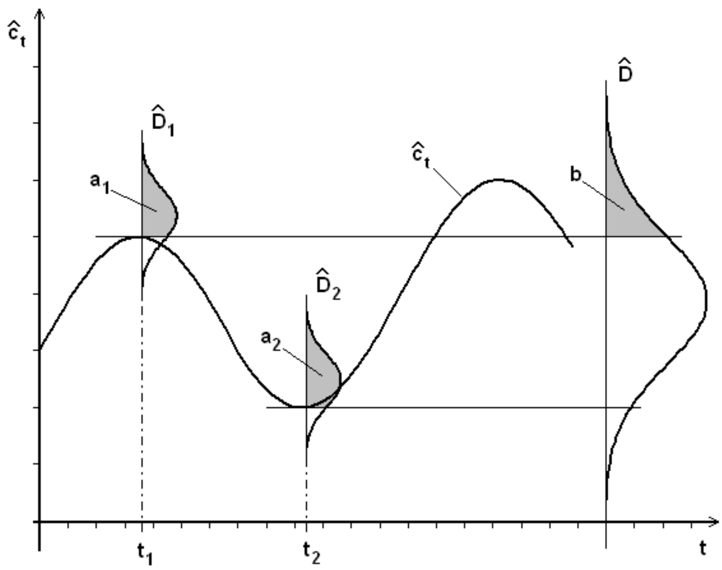

Figure 1. The estimated state of the economy $\widehat{c}_t$, its empirical distribution $\widehat{D}$, and the empirical distributions $\widehat{D}_1$ and $\widehat{D}_2$ of the resampled realization of $\widehat{c}_t$ at time points t_1 and t_2.

The shaded area a_1 is a measure of this tendency, and it is obvious that the tendency towards a recession will be larger if $\widehat{c}_{t_1}$ would be located further down in the lower tail of $\widehat{D}_1$. The opposite is true for the tendency towards an expansion. Assume now that the two areas a_1 and a_2 are equal. This does not mean that the probability of a recession is equal at the two time points t_1 and t_2. Since the state of the economy is at a peak of the business cycle at t_1, the tendency of a recession is low compared to the tendency at t_2 when $\widehat{c}_t$ is at a trough. The probability of a recession is thus smaller at t_1 than at t_2. This is easily seen in the estimated distribution $\widehat{D}$ of the overall state of the economy, where the shaded area b obviously is smaller than the corresponding area at time t_2. Thus, estimating the probability of a recession, or expansion, at time t, both local and global tendency must be considered. Conditional on global and local tendency, b and a_1 at time t_1, the probability of a recession can be estimated applying Bayes theorem. The probability of a recession, $P_{r,i}$, at time point i can be defined as

$$P_{r,i} = \frac{P\left(c_t^b \geq \widehat{c}_i\right) P\left(\widehat{c}_t \geq \widehat{c}_i\right)}{P\left(c_t^b \geq \widehat{c}_i\right) P\left(\widehat{c}_t \geq \widehat{c}_i\right) + P\left(c_t^b \leq \widehat{c}_i\right) P\left(\widehat{c}_t \leq \widehat{c}_i\right)}, \tag{7}$$

where $\widehat{c}_i$ is the value of $\widehat{c}_t$ at time point i, and c_t^b is the resampled series of $\widehat{c}_t$, $t = 1, \ldots, i, \ldots, T+s$. Probabilities of an expansion, $P_{e,i}$, follow analogously. For the case in Figure 1, the probability of a recession at t_1 would then be

$$P_{r,t_1} = \frac{a_1 b}{a_1 b + (1 - a_1)(1 - b)}. \tag{8}$$

The main advantage of this approach over the method applied by Stock and Watson (1993) is that this method is completely nonparametric. The semi-parametric approach proposed by Stock and Watson involves assumptions concerning two specific patterns of how a recession and an expansion look like. Furthermore, to be able to classify time points as a recession or expansion, and to estimate the probabilities, they needed to estimate three parameters describing two stochastic limits.

5 Empirical application

5.1 The data set

Applying the variable selection procedure of Section 2 to the Icelandic economy, this study uses monthly data available for in total 104 macroeconomic and financial variables from a number of different sectors and markets. The sample size is limited to the period between January 1999 and September 2005, thus giving 81 observations in total. All variables are expressed in real values if necessary. 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 can 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, the data series are only analysed in levels or in yearly growth rate. For index series the transformation to the yearly growth rate can not be applied directly, why they instead are expressed in yearly growth rate in percent.

Since the data is available on a monthly basis it would be possible to transform the data into monthly, instead of yearly, growth rate. However, using yearly growth rate decreases the effect of any possible seasonality. Not accounting for seasonality in the model could affect the estimate of the variable c_t. The model as specified in equations (1)-(4) does allow for modelling seasonality. However, the sample size of the data set available restricts the number of possible variables to be included in the model. Adding seasonal components to the model would thus be more appropriate for larger sample sizes and is therefore left for later studies.

A number of potential leading indicators and coincident variables can be selected for the Icelandic economy by analyzing the correlation structure of each candidate variable in levels, or in growth rate, with the annual real GDP growth rate. The chosen candidate variables are listed and described 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	growth rate
YNB	Yield spread, Tr Notes 5 year - Tr Bills 3 month	levels

This selection method of coincident and leading variables differs from the two step procedure applied by Stock and Watson. The suggested method in this study, to a priori select both coincident and leading variables, is a consequence of how the model is specified and estimated. As mentioned in Section 3, Stock and Watson used a two-step model estimation procedure.

Their variable selection approach started by first choosing coincident variables. Then, what they called, a coincident economic index was estimated. The leading indicators were then selected in a second step on the basis of this estimated coincident index series, not on the GDP. Using this two-step approach made it possible to select coincident variables and leading indicators at different stages in the analysis. This implied, however, that all possible valuable information in the leading indicators was neglected in the estimation of the current state.

5.2 Model specification and estimation

Due to the limited sample size, it is not possible to include all the potential variables in Table 1 when specifying the model. That should, on the other hand, not be necessary for obtaining reliable estimates of model parameters and of the state of the economy. Including only 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.

Focusing first on 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 in the model. Selecting between the two, the variable with the strongest correlations with real GDP in growth rate is the most natural choice.

Four coincident variables are thus included in the model; cement sales, imports of food and beverages, vacancies in the Reykjavik area, and real exchange rate. These four variables originate from different sectors of the Icelandic economy, it is therefore reasonable to believe that they will form a good basis for describing the general business cycle. The two variables, imports of consumer goods, and new work permits, will be left out of the study due to the lower correlation with GDP and to their close relation to the variables import of food and beverages, and vacancies in the Reykjavik area respectively.

Selecting leading indicators to include in the model poses a larger problem. An obvious reason to include leading indicators is to take into account early information of the behaviour of the business cycle. Using a single leading indicator can not be recommended since economic theory and experience suggest that recessions can have a number of different sources and characteristics. On the other hand, the more indicators included in the model, the larger and more complex the model becomes. Therefore, considering the 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 has been set to three in the analysis that follows.

Given the list in Table 1 of ten potential leading indicators, there are 120 possible combinations containing three indicators. Only a small fraction of these combinations have been considered in this study. By choosing different sets containing three leading indicators from various sectors in the economy, only five of the combinations have been thoroughly analysed. The selection between the different sets 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 four selected coincident variables. Secondly, the model forecast performance is evaluated over a hold out sample. This is done by estimating the model over the sample period saving the last 12 observations for a comparison with the point forecasts. The forecast performance is measured by the standard mean squared error (MSE), where the average mean squared errors summarize the performance.

Credit cards, export of marine products and trade balance are chosen as the leading indicators to be included in the model. In Table A1 in the Appendix the forecast of the coincident variables, the leading indicators, their MSE and average MSE are reported for the hold out sample of 12 observations. The other four sets of analysed leading indicators, not reported here, contain various combinations of the three indicators from above and the variables; wages and salaries, new vehicles registration, oil price, and yield spread between 20 year treasure bonds and 5 year treasure notes.

Analysis of the correlation structure between the coincident variables and the leading indicators, and between the individual leading indicators, is used as a basis for specifying the lag structure of the model (5). Estimation is then performed by maximum likelihood 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

As in any exercise of estimating an econometric model, it is necessary to get as good estimates as possible of the model parameters. Be it forecasting or policy evaluation, any implementation of the model following the estimation will depend on the parameters. Therefore, finding out whether or not the model appears to satisfy the assumptions under which it was estimated is an integral part of any modelling exercise. A number of diagnostic tests and methods is applied in this study to confirm that the model assumptions are not violated, and that the estimated model has desirable properties such as a mean reverting behavior. The appendix contains tables with the most important results of the evaluation of the estimated state space model.

Two important model assumptions are that the error vectors $\boldsymbol{\varepsilon}_t$ and $\mathbf{w}_t$ in (5) are multivariate normally distributed and serially uncorrelated. The estimation of the model is based on maximum likelihood, so any large deviation from normality would result in inefficient parameter estimates. Table A3 presents results of the univariate Jarque and Bera (1980) tests of normality of each of the residual vectors. The residual vector ε_t^{RER} has a p-value of just over 1%, thus showing moderate nonnormality. However, taking into account the p-values of the other residual vectors, who all are high, the overall normality assumption of the error vectors is not in any major extent affected by this departure from normality.

When analysing the presence of autocorrelation in the residuals, each residual vector is regressed against a constant and q lags of itself. Each reported p-value in Table A4 is the result of a joint test of the hypothesis that all parameters except the constant are zero. A few individual p-values of the 48 tests show presence of mild autocorrelation. Given a test level α, the overall conclusion does, however, indicate very modest or negligible autocorrelation in the error vectors $\boldsymbol{\varepsilon}_t$ and $\mathbf{w}_t$. The result thus implies that we would not reject the assumption that the error vectors are serially uncorrelated.

It is sometimes argued that misspecification of the model can make the error variances time-varying. A multivariate Lagrange multiplier test is therefore applied to test if the model error variances are heteroskedastic. Table A5 presents results of the LM test, see Eklund and Teräsvirta (2007) for specific details of the test. As the alternative to constant variances, three time-varying variance specifications are considered. First, an alternative where the variances are functions of the explanatory variables, referred to as the White case. Next, a case of ARCH(q) time-varying variances, $q = 1, 3, 5$, and finally a smooth transition of the variances over time. 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.

When analysing the mean reverting properties of the model, 10000 observations of the vectors $\mathbf{z}_t$ and $\mathbf{h}_t$ are generated from the estimated model. Such a simulated realization of the model variables can be an effective method to reveal any nonstationary properties of an estimated model. Table A6 reports values of the minimum, mean, maximum, and the standard deviation of the simulated series. These results are a strong indication that the model is stationary since all individual time series have realization that evolve within lower and upper bounds and, furthermore, with low standard deviations. This feature of stationarity, or mean reversion, is clearer when the impulse response is considered. Figures A1 and A2 show the response over the months to come of a unit shock to the error of the state of the economy ε_t^c. Figure A1 depicts the effect on the leading indicators and c_t, while Figure A2 shows the effect on the coincident variables. As the shock is introduced into the system at time point $t = 0$, depending on the lag structure of the model, each variable reacts in a consequence of its relationship with c_t and the other variables. As shown in the figures, it is clear that the shock fades out already after about two years, that is after 24 observations. Therefore, there is no evidence of a permanent effect of the shock as would be the case if any of the series would be nonstationary or contain a unit root. 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. It is therefore safe to conclude that the model is stationary and mean reverting.

5.4 Estimated recession and expansion probabilities

The estimated state of the economy $\widehat{c}_t$ represents the economy's common element that drives the evolution of variables with a business cyclical behavior. It will, however, only reflect the pattern of the economic situation, not the size of the economy or the magnitude of growth. As a consequence, it can only show whether the economy is in a recession or in an expansion, or more precisely serve as a base for estimating the probabilities of a recession or a expansion in the economy. It can thus not be used to predict the GDP growth.

Figure 2 depicts the estimated state of the economy $\widehat{c}_t$ and the real GDP yearly growth rate over the sample period considered including 12 month forecasts of c_t. The estimated correlation between $\widehat{c}_t$, when quarterly aggregated, and the yearly growth rate of GDP is as high as 0.79, indicating that key features of the business cycle has been accounted for in the modelling process. What is interesting to note is the start of the downward trend of $\widehat{c}_t$ in April 2001, just preceding the one of the GDP growth rate, and the turn upward, around June 2002, before the trough in the cycle of GDP is reached. It also appears that $\widehat{c}_t$ has a peak around May 2005, which could indicate that the Icelandic economy is slowly turning towards a recession. Estimating the probabilities can be helpful in determining how strong this downturn in the economy is.

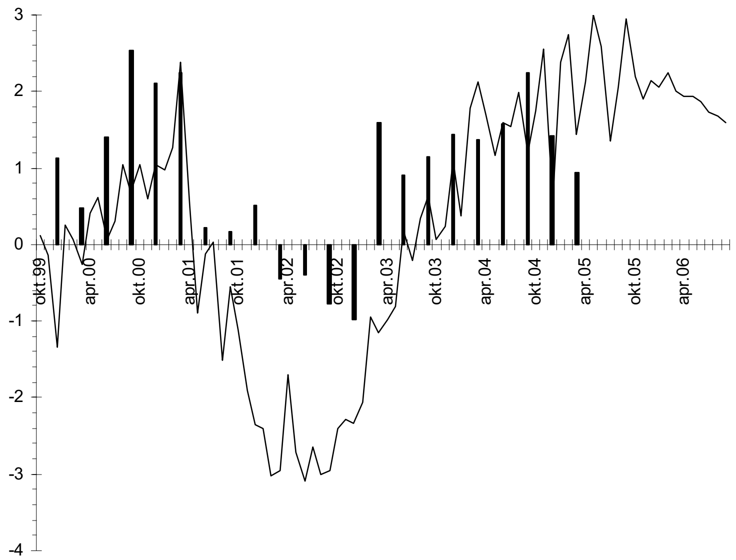

Figure 2. Estimated state of the economy $\widehat{c}_t$, (line), and real GDP in yearly growth rate (bars).

When estimating the recession and the expansion probabilities the resampling technique from Section 4 is implemented. Conditionally on the estimated parameters, and on the forecasts of coincident and leading variables, $R = 200$ realization of $\widehat{c}_t$ are generated from the model. Since no autocorrelation is detected in the residual vectors there is no need for applying a moving blocks bootstrap. Results from the evaluation of the estimated model also indicate, since the normality hypothesis is not rejected, that errors can be drawn either directly from a multivariate normal distribution or alternatively from the residual vectors using a simple bootstrap. When comparing the two techniques, resampling both by drawing errors from the residual vectors and from the multivariate normal distribution, only negligible differences can be detected in the resulting recession and expansion probabilities. In this case with no autocorrelation and normal errors, it can thus be concluded, as expected, that the two resampling techniques are equivalent.

Figure 3 shows the estimated state of the economy for the sample period and the 12 month forecasts after the estimated period. The figure also includes 30 of the 200 bootstrap realization plotted at each fifth time point. This figure can directly be compared to Figure 1 in Section 4, where the probability estimation method is described. Each dot represents an observed value of a resampled series c_t^b for some $b = 1, \ldots, 200$. At a given time point t, the 200 values of c_t^b describe the empirical distribution $\widehat{D}_t$ of $\widehat{c}_t$, which will serve as a base for estimating the probabilities. It is then a simple task to estimate the frequencies of observed values of c_t^b above and below the estimated economic state $\widehat{c}_t$ at all time points $t = 1, \ldots, T + s$, and then to estimate recession and expansion probabilities.

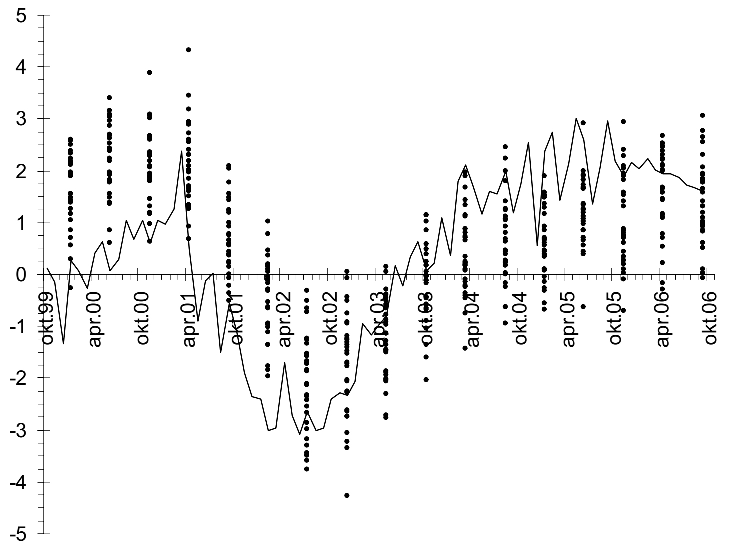

Figure 3. Estimated state of the economy $\widehat{c}_t$, (line), and 30 resampled values of c_t^b, (dots), at each fifth time point.

Figure 4 shows the resulting estimated recession probabilities for the sample and forecast periods together with the estimated state of the economy $\widehat{c}_t$, and Figure 5 depicts the probabilities and the yearly growth rate of GDP over the same time period. Figure 6 depicts the GDP growth rate and the corresponding expansion probabilities which are, naturally, the mirror image of the recession probabilities.

As shown in these three figures, the first couple of months in the sample period indicate a high probability of a recession, even though the GDP yearly growth rate is positive. This can be explained by the low, and relatively volatile, values of the estimated state of the economy $\widehat{c}_t$, which even is negative for some of the months during the first year of the sample. Later, around October 2000, as the economic situation stabilises, the recession probabilities decrease with a smallest value obtained in March 2001. This is followed by an abrupt increase of the probability in April 2001, which starts an almost two year long recession in Iceland. The estimated probabilities correspond very well with the actual facts of the Icelandic economy over this period, which showed a sharp contraction in several sectors of the economy. This can also be detected by analysing the GDP growth rate, showing low, and even negative, growth over this two year period. As the economy recovers from the recession the probabilities again decrease slowly. The estimated state of the economy does also correspond well with the measure of the output gap reported by the Central Bank of Iceland.

After February 2004, and for the remaining sample and forecast period, the recession probabilities are, with a few individual exceptions, very low. This indicates that the Icelandic economy is in a stage of expansion, and will remain there for some time. The probabilities of a recession do increase slightly towards the end of the forecast period, but the expansion tendencies would probably still be a dominant factor at this point in time.

A very interesting feature in the figures is the asymmetric characteristics of the probabilities. As information of a recession appears in the economy, information that the leading indicators can capture, there is a very rapid increase of the recession probabilities. On the other hand, when there are information of expansion tendencies in the economy, the adjustment of the probabilities are much slower, indicating that it is easier to predict a recession than an expansion in the economy. This is a feature that also other studies have noted, see for example the discussion in Teräsvirta (2006).

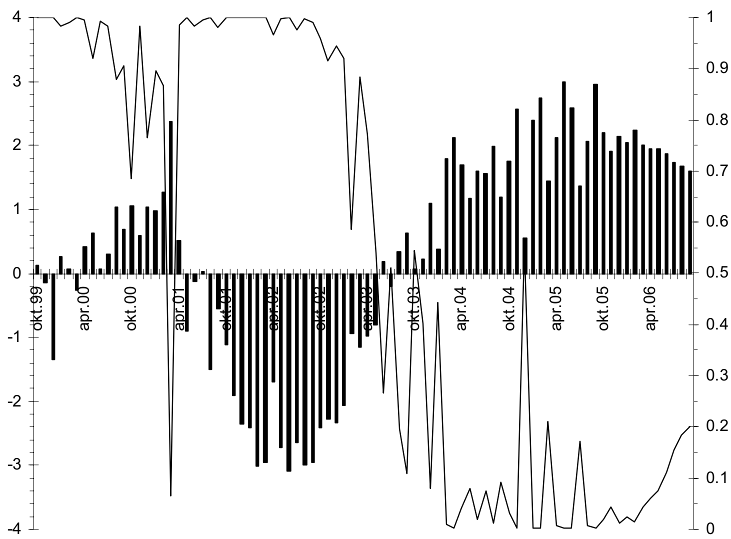

Figure 4. Estimated state of the economy $\widehat{c}_t$, (bars), and estimated probabilities of a recession, (line).

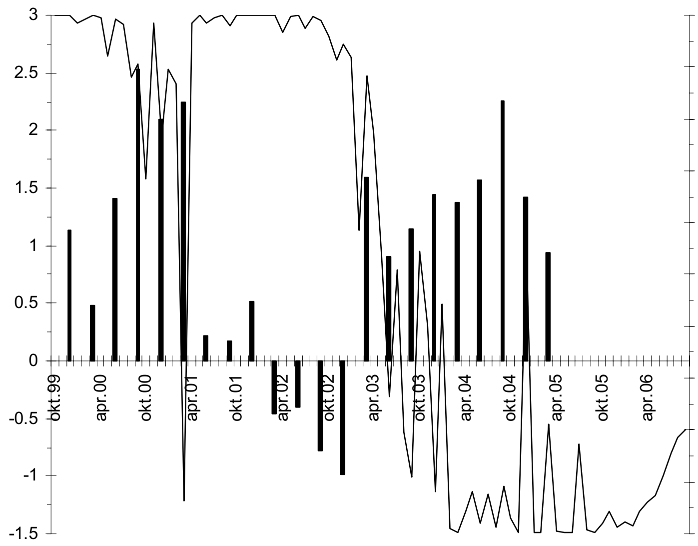

Figure 5. Real GDP in yearly growth rate, (bars), and estimated probabilities of a recession, (line).

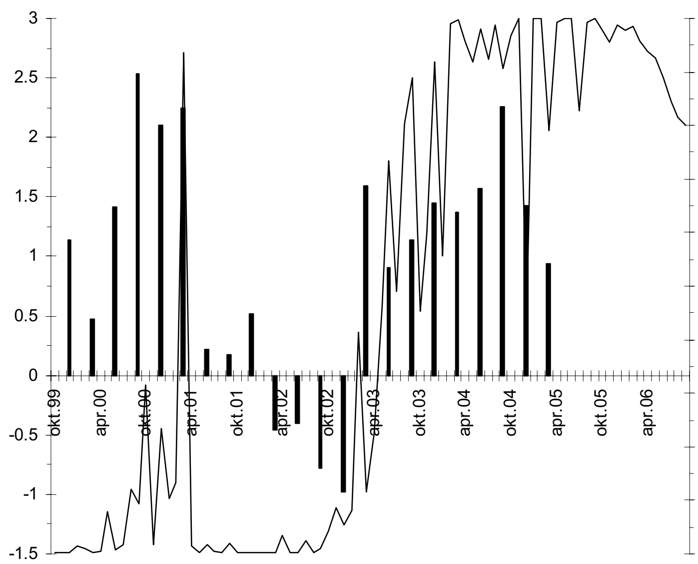

Figure 6. Real GDP in yearly growth rate (bars) and estimated probabilities of an expansion (line).

6 Conclusions

The purpose of this paper is to investigate the Icelandic business cycle, and to estimate the probability that the economy will be in a recession or in an expansion over the analysed sample and forecast period. The analysis is based on the Stock and Watson methodology, where the business cyclical behaviour present in the Icelandic macroeconomic and financial time series is assumed to depend on a single unobservable factor, the state of the economy.

Several coincident variables and leading indicators are identified for the economy. A large number of different sets of variables and indicators can therefore be combined and used in the modelling procedure. This study only investigates the performance of five different sets thoroughly. Further analysis of the contribution of other choices of coincident variables and leading indicators is left for future studies.

A factor model is specified and estimated through the Kalman filter. The cyclical behaviour of the Icelandic economy is satisfactory estimated by the state of the economy, indicated by a high correlation with the yearly GDP growth rate.

For estimating recession and expansion probabilities, a nonparametric method is developed based on bootstrap resampling. This technique requires no assumptions concerning recession and expansion patterns, or the need of estimating extra parameters, as the semi-parametric approach of Stock and Watson. The estimated probabilities show that there is a high probability for an expansion in Iceland over the forecasting period, and that this probability will remain high for some time. The results also show an interesting asymmetric feature, that predicting a recession appears to be easier than predicting an expansion. This suggests additional research topics in the area of predicting business cycles using leading indicators.

References

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

Eklund, B. and Teräsvirta, T. (2007), 'Testing the constancy of the error covariance matrix in vector models', *Journal of Econometrics* . Forthcoming.

Fitzenberger, B. (1997), 'The moving blocks bootstrap and robust inference for linear least squares and quantile regressions', *Journal of Econometrics* **82**, 235–287.

Geweke, J. (1977), The dynamic factor analysis of economic time series, *in* D. J. Aigner and A. S. Goldberger, eds, 'Latent Variables in Socio-Economic Models', North Holland Publishing (Amsterdam).

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

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

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

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

Moore, G. H. and Shiskin, J. (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.

Sargent, T. J. and Sims, C. A. (1977), Business cycle modeling without pretending to have too much a priori economic theory, *in* 'New Methods in Business Cycle Research', Federal Reserve Bank of Minneapolis.

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

Stock, J. H. and Watson, M. W. (1991), A probability model of the coincident economic indicators, *in* K. Lahiri and G. H. Moore, eds, 'Leading Economic Indicators: New Approaches and Forecasting Records', New York: Cambridge University Press, chapter 4, pp. 63–89.

Stock, J. H. and Watson, M. W. (1993), A procedure for predicting recessions with leading indicators: Econometric issues and recent experience, *in* J. H. Stock and M. W. Watson, eds, 'New Research on Business Cycles, Indicators and Forecasting', University of Chicago Press, pp. 95–156.

Teräsvirta, T. (2006), Forecasting economic variables with nonlinear models, *in* G. Elliott, C. W. J. Granger and A. Timmermann, eds, 'Handbook of Economic Forecasting', Elsevier.

Appendix A. Tables

Table A1. Forecasts and mean squared errors of coincident variables and chosen leading indicators over a hold out sample.

h	IFB_{T+h}	VRA_{T+h}	CEM_{T+h}	RER_{T+h}	EMP_{T+h}	TRAB_{T+h}	CRED_{T+h}
1	0.877	0.800	0.870	−0.448	−0.658	−0.752	−0.107
MSE	0.017	0.088	0.028	0.007	0.003	0.024	0.200
2	0.458	0.696	0.805	−0.365	0.019	−0.372	0.171
MSE	1.147	0.033	0.168	0.065	0.528	0.022	0.000
3	0.506	0.587	0.692	−0.306	−0.077	−0.266	−0.150
MSE	0.808	0.220	0.329	0.457	0.001	0.234	0.132
4	0.467	0.529	0.670	−0.279	−0.087	−0.349	−0.060
MSE	0.002	0.195	0.483	0.332	3.237	0.450	1.665
5	0.398	0.430	0.554	−0.254	−0.099	−0.387	−0.184
MSE	3.716	0.453	0.911	0.415	0.035	0.427	0.605
6	0.330	0.354	0.461	−0.234	−0.006	−0.229	−0.197
MSE	0.189	0.405	0.128	1.029	0.046	1.206	0.085
7	0.307	0.310	0.421	−0.216	−0.055	−0.245	−0.258
MSE	0.005	0.691	1.149	0.917	0.182	0.507	0.272
8	0.291	0.277	0.395	−0.200	0.085	−0.130	−0.240
MSE	1.791	0.484	2.296	0.595	0.025	3.667	0.001
9	0.239	0.226	0.325	−0.179	0.085	−0.123	−0.275
MSE	0.811	1.900	1.019	0.875	0.126	4.027	3.751
10	0.207	0.191	0.279	−0.162	0.102	−0.082	−0.289
MSE	0.099	0.001	0.214	1.085	3.408	6.538	0.376
11	0.158	0.148	0.213	−0.145	0.127	−0.046	−0.294
MSE	1.977	0.078	0.578	1.122	2.244	9.282	0.803
12	0.124	0.118	0.168	−0.130	0.133	−0.009	−0.310
MSE	5.439	1.639	0.660	1.748	2.542	10.643	0.793
Average MSE	1.333	0.516	0.664	0.721	1.031	3.086	0.724

Table A2. Model parameters estimated by maximum likelihood, each parameters standard deviations below in parenthesis.

Observation equation

$$
\begin{array}{rcl}
IFB_t & = & \underset{(0.1508)}{0.4025}\, c_t + u_t^{IFB} \\
VRA_t & = & \underset{(0.1171)}{0.2855}\, c_t + u_t^{VRA} \\
CEM_t & = & \underset{(0.1399)}{0.3949}\, c_t + u_t^{CEM} \\
RER_t & = & \underset{(0.0465)}{0.0123}\, c_t + u_t^{RER} \\
EMP_t & = & \underset{(0.1855)}{0.3815}\, c_{t-1} - \underset{(0.1729)}{0.3475}\, c_{t-2} + \underset{(0.1038)}{0.1992}\, EMP_{t-1} + \underset{(0.1223)}{0.1907}\, TRAB_{t-2} \\
 & & \underset{(0.1216)}{-0.2868}\, CRED_{t-1} - \underset{(0.1370)}{0.1922}\, CRED_{t-5} + \varepsilon_t^{EMP} \\
TRAB_t & = & \underset{(0.1239)}{-0.2736}\, c_{t-2} + \underset{(0.1137)}{0.1911}\, EMP_{t-1} + \underset{(0.1216)}{0.2039}\, TRAB_{t-1} \\
 & & \underset{(0.1301)}{-0.1570}\, CRED_{t-5} + \underset{(0.1382)}{0.1060}\, CRED_{t-6} + \varepsilon_t^{TRAB} \\
CRED_t & = & \underset{(0.1057)}{0.0825}\, c_{t-1} - \underset{(0.1055)}{0.1080}\, c_{t-2} - \underset{(0.0877)}{0.0561}\, EMP_{t-1} \\
 & & \underset{(0.0996)}{+0.2425}\, CRED_{t-1} + \underset{(0.1041)}{0.5516}\, CRED_{t-2} + \varepsilon_t^{CRED}
\end{array}
$$

State equation

$$
\begin{array}{rcl}
c_t & = & \underset{(0.0938)}{0.0559} + \underset{(0.1400)}{0.4088}\, c_{t-1} + \underset{(0.1382)}{0.4140}\, c_{t-3} - \underset{(0.1485)}{0.2990}\, EMP_{t-2} \\
 & & \underset{(0.1659)}{+0.0648}\, TRAB_{t-2} + \underset{(0.1561)}{0.2513}\, CRED_{t-8} + \varepsilon_t^c \\
u_t^{IFB} & = & \underset{(0.1347)}{-0.2486}\, u_{t-1}^{IFB} + \varepsilon_t^{IFB} \\
u_t^{VRA} & = & \underset{(0.0817)}{0.7782}\, u_{t-1}^{VRA} + \varepsilon_t^{VRA} \\
u_t^{CEM} & = & \underset{(0.0860)}{0.7972}\, u_{t-1}^{CEM} + \varepsilon_t^{CEM} \\
u_t^{RER} & = & \underset{(0.1142)}{1.6541}\, u_{t-1}^{RER} - \underset{(0.1965)}{0.9673}\, u_{t-2}^{RER} + \underset{(0.1142)}{0.2786}\, u_{t-3}^{RER} + \varepsilon_t^{RER}
\end{array}
$$

Estimated residual variances

$$
\begin{array}{lll}
\sigma^2_{\mathrm{IFB}} = \underset{(0.0894)}{0.4440}, & \sigma^2_{\mathrm{VRA}} = \underset{(0.0394)}{0.2068}, & \sigma^2_{\mathrm{CEM}} = \underset{(0.0315)}{0.1133} \\
\sigma^2_{\mathrm{RER}} = \underset{(0.0075)}{0.0444}, & \sigma^2_{\mathrm{EMP}} = \underset{(0.0936)}{0.4991}, & \sigma^2_{\mathrm{TRAB}} = \underset{(0.0955)}{0.5487} \\
\sigma^2_{\mathrm{CRED}} = \underset{(0.0631)}{0.3772}, & \sigma^2_c = \underset{(0.3862)}{0.5168} &
\end{array}
$$

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

	ε_t^{IFB}	ε_t^{VRA}	ε_t^{CEM}	ε_t^{RER}	ε_t^{EMP}	ε_t^{TRAB}	ε_t^{CRED}	ε_t^{c}
Skewness:	−0.099	−0.112	−0.105	−0.233	0.022	0.002	0.038	−0.129
Kurtosis:	2.837	3.156	2.678	4.671	2.975	2.912	3.990	2.412
JB value:	0.197	0.224	0.443	9.030	0.008	0.023	2.960	1.236
p-value:	0.906	0.894	0.801	0.011	0.996	0.988	0.228	0.539

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^{IFB}	ε_t^{VRA}	ε_t^{CEM}	ε_t^{RER}	ε_t^{EMP}	ε_t^{TRAB}	ε_t^{CRED}	ε_t^{c}
1	0.432	0.690	0.224	0.270	0.634	0.940	0.281	0.550
2	0.667	0.802	0.465	0.213	0.019	0.105	0.264	0.655
3	0.839	0.846	0.568	0.072	0.012	0.202	0.332	0.024
4	0.044	0.896	0.703	0.074	0.036	0.358	0.507	0.028
5	0.068	0.810	0.570	0.121	0.058	0.156	0.042	0.058
6	0.151	0.809	0.770	0.039	0.123	0.235	0.007	0.171

Table A5. Testing constancy of the error covariance matrix against time varying variances specified by:

	White	ARCH(q)			Smooth transition
		$q = 1$	$q = 3$	$q = 5$	
LM test p-value:	0.115	0.170	0.129	0.116	0.537
Bootstrap p-value:	0.280	0.190	0.140	0.120	0.540

Table A6. Simulated realization: 10000 observations.

	IFB_t	VRA_t	CEM_t	RER_t	EMP_t	$TRAB_t$	$CRED_t$	c_t
Min:	−3.805	−2.942	−2.712	−4.289	−3.327	−3.629	−2.961	−5.012
Mean:	0.085	0.066	0.081	−0.118	−0.009	−0.095	−0.033	0.235
Max:	3.350	2.938	2.947	3.534	3.530	3.999	3.136	5.132
Stdv:	0.906	0.831	0.819	1.069	0.909	0.961	0.842	1.467

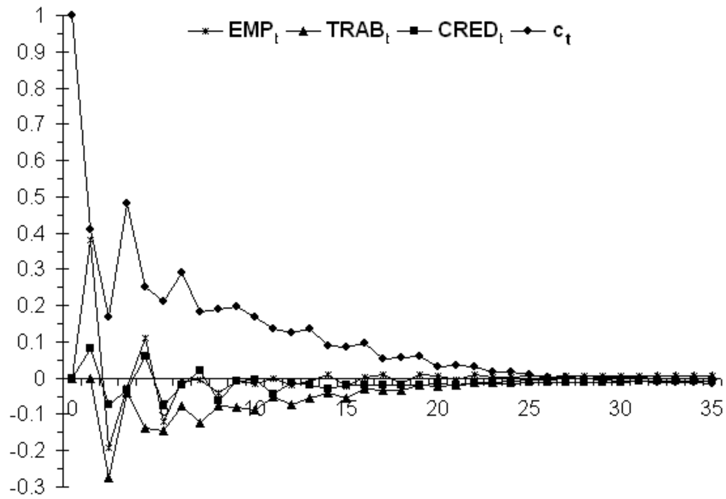

Figure A1. Impulse response over 35 months on the yearly growth rate of the state of the economy and on the leading indicators of a unit shock to ε_t^c, the error of c_t.

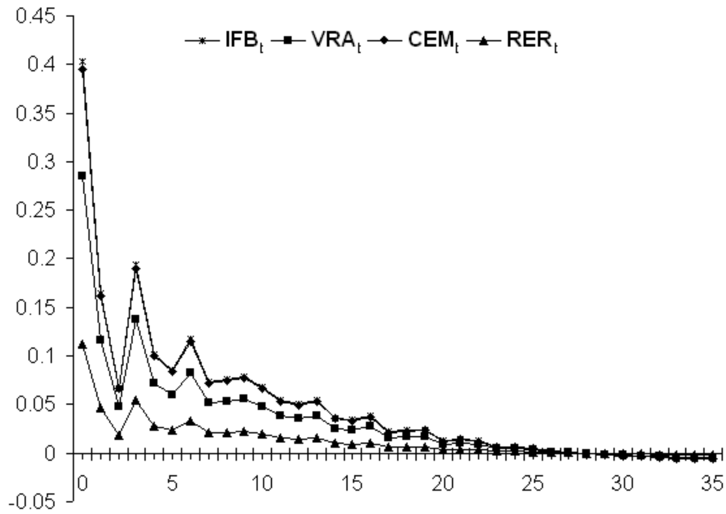

Figure A2. Impulse response over 35 months on the coincident variables of a unit shock to ε_t^c, the error of c_t.