Statistics Netherlands

Discussion Paper

Analysing short term developments in key economic indicators: predicting the sign of period on period changes

2014 | 20

Floris van Ruth

1 augustus 2014

Index

Summary

The need for ever faster information on current economic conditions and the focus on short term changes tends to clash with a desire for realiability and clarity. This is especially so for monthly economic indicators, where the month-on-month changes tend to be rather volatile, making analysis difficult. One way to assist in this could be to use predictions of next month's developments as tool to assess the current month's realisations. More precisely, if it is predicted that the next month change will probably be a decrease, then the current month's increase should get less attention, as it is likely to be only transitory. This study applies a range of predictive techniques to forecasting the sign of the next month's change in five key economic indicators: consumer confidence, household consumption, producer confidence, manufacturing production and exports, all for the Netherlands. Techniques tested range from standard regression models to advanced machine learning techniques. In all cases, predictive models outperformed naive benchmarks. The best out of sample forecasting accuracies ranged between 65,5% and 80%. Unfortunately, it was found that the amount of available data was too small to get a robust estimate from the machine learning techniques. The overal best performing and most robust technique was found to be a combination of a GLM model with dynamic factor estimates.

Keywords

forecasting, model evaluation, business cycle, machine learning, short term economic analysis

1. Introduction

Markets, media and policymakers need ever more current and speedy information on economic conditions. This desire to see changes in economic conditions as soon as they happen has been amplified after the financial crisis of 2008, when economic analysts felt surprised and many people were anxious to see whether policy implemented to combat the crisis was effective. One consequence of this was increased attention for the monthly short term economic indicators such as consumption, industrial production and consumer confidence. Another was a clear shift in focus to the short-term, i.e. period on period, development in these statistics. Where most analysis used to be based on either year on year growth rates or deviations from trend, a clear preference for month on month growth rates emerged.

This development leads to this study, for even though month on month changes in monthly economic indicators may give current information, they tend to be very noisy as well. There is a lot of short-term variability in these series, which is at least partly random. This means that an increase in one month might well be followed by a decrease in the next month, without this actually indicating a change in underlying economic conditions. It is hard to distill reliable information from these short term changes. A potential solution, which is tested in this study, is when analysing developments in month t, to make a prediction for the development in month t+1. This additional information can help in deciding whether the current month's development is likely to persist, or is transitory.

This study builds on and leans heavily on Cornec and Mikol [2011], who use a range of regression-based and machine learning models to predict whether the quarterly growth rate of GDP will increase or decrease. They are not focusing on the growth rate per se, but whether the growth rate in quarter t is larger or smaller than the growth rate in quarter t-1. This gives valuable information on the current development in the dynamics of the economy. They find that linear discriminant analysis yields superior results to general linear model variants, and to machine learning techniques such as support vector machines and decision trees, and that all of these dominate naive strategies. Another interesting result from that study is the fact that business survey indicators contain highly relevant information for nowcasting these economic dynamics.

Most studies into forecasting economic variables tend to target the actual level or growth rate, focusing less on the dynamics. Current state of the art tends to be Markov switching models, or dynamic factor models. Markov switching models focus on identifying different regimes for growth or decline phases, characterised by respectively positive and negative growth rates [Hamilton(1989), Chauvet, M. and Piger, J.M. (2003). The basic idea of the Markov switching models could be of use here, but non-linear models tend to be somewhat fragile, and relatively poor at forecasting. Dynamic factor-based models [Forn et al. (2005), Stock and Watson (2002)] tend to be quite good at forecasting short-term economic developments, and are also quite attractive of construction. The dynamic factors are extracted from the set of individual explanatory variables, making the models relatively insensitive to the availability of and the noise in individual indicators. Therefore, this approach is potentially useful here, when modified to predicting the sign of period on period changes.

One way this study differs from Cornec and Mikol[2011] is that it focuses on prediction the direction of change in the five most important monthly economic indicators: manufacturing production, exports, household consumption, producer confidence and consumer confidence. This focus on monthly indicators is due to the need for continual information on current economic conditions. Also, interest tends to be in whether the level of an economic indicator is above or below that of the previous month, therefore the focus on the sign of the month on month change in the level and not in the growth rate. Another difference is that this study tests a broader range of models both standard econometric and machine learning. Variants of standard regression based models tested here are OLS and GLM models with automatic variable selection, OLS and GLM models using principal components based dynamic factors, and GLM ensemble models. Machine learning models evaluated here are decision trees, random forests, support vector machines, LDA, boosting and bagging. What these approaches have in common is either automatic variable selection or no need for variable selection, and the fact that they can be run automatically. This means that these techniques are relatively robust, they still perform when variables are missing or relations between indicators change, without the need of manual intervention. This is important if one wants to develop a system to support the analysis of monthly economic indicators, as time tends to be in short supply in monthly production and analysis cycles, whilst at the same time missing or changing indicators tend to be relatively common.

2. Methods used

The aim of the models tested here is to predict whether the level of the target variable y_t will be larger or smaller at t+1, i.e. whether the period on period change in the level is positive or negative.If necessary for the model tested, a dummy variable was created taking the value 0 if the month on month change in the level Δy_{t+1} was negative, and 1 if it was positive.
Model evaluation was done by performing predictive analysis on a test set not used in the model construction. It consisted of the last 10% of the total data set available, or about 30 months. Scoring was done by calculating the % of month on month changes correctly classified.

Automatic model selection
In automatic model selection, an algorithm is used to determine the ideal set of explanatory variables from the complete base set. Here, the model selection functionality of the R MASS package was used. It offers the option to perform variable selection based on a combination of forward and backward selection, where variables are entered one by one or entered all together and then deleted one by one. This based on a fit criterion such as AIC, where the variable with the highest explanatory power is added first/with the lowest explanatory power is deleted first. The models used were OLS and GLM-logit.

Dynamic Factor Models
Dynamic factor models are having a large impact on the practice of economic forecasting [Forni et al. (2005) , Stock and Watson (2002)]. The basic idea is that developments in the data are driven by only a few underlying factors [Reijer (2005]. Extracting these will result in both a dramatic dimension reduction and noise reduction. It will also make the model more robust to the availability of individual indicators. The basic model structure is:

$$X_t = \boldsymbol{\Lambda F_t} + e_t$$

$$y_{t+1} = \beta F_t + \varepsilon_t$$

Where F_t are the latent dynamic factors, $\boldsymbol{\Lambda}$ is a matrix of factor loadings, X_t is a matrix containing the explanatory variables, and y is the target indicator.
Here, principal components analysis is used to extract the common factors from the explanatory data set X_t. It is then possible to build a OLS or GLM-logit model using for example the first principal component (the strongest dynamic factor) or the full common component which is based on summing the total number of dynamic factors (in the case of the Dutch economy the estimated number of dynamic factors is three).

Ensemble models
Ensemble modeling is a form of forecast averaging, whereby predictions of a large number of related models are averaged to come to a consensus forecast. The underlying idea is that a model which performs well/badly in one period, does not necessarily perform well/badly in another, which makes ex ante model selection problematic. However, combining the outcomes of a large number of reasonable models should result in a reasonable prediction where the extremes average out. In this study, GLM-logit forecasts were made from all possible models of combinations of 1, 2 or 3 of the explanatory variables. The explanatory variable set was augmented by including the 2, 3 and 12 month lag of each explanatory variable as well. The ensemble prediction was computed by taking the majority vote, i.e. whether the fraction predicting 0 was larger or smaller than the one predicting a 1.

Decision trees
Classification trees were generated using the RPART algorithm [Breiman (1984)]. This is based on stepwise partitioning the data into separate samples, selecting in each step the optimal splitting variable and threshold value. The evaluation is done using the relative homogeneity of the resulting partitions as measure of the goodness of fit. The aim is for ever more homogeneous subsamples, here characterised by either an decrease (0) or increase (1) of the target variable. A second, pruning stage is included where splits which add relatively little to the explanatory power are removed. This is evaluated by cross-validation.

Random forest
Random forests are an ensemble extension of decision trees.[Breiman (2001)] The thesis is that it is better to average the outcomes of a large number of decision trees that to look for an optimal one. A multitude of trees is grown by selecting for each tree a bootstrap sample from the training data set. For each node in this tree, a random set of m indicators is selected from the n available explanatory variables. The split is then decided on the optimal indicator from this subset. This is repeated until the tree is fully grown, after which the tree growing procedure is repeated until the forest is completed. Estimation is performed by majority vote.

Boosting
Boosting results from testing whether using many weak models or "learners" can yield strong results [Schapire. (1990)]. Here, the ADABOOST algorithm [Freund and Schapire (1997)] was used. A large number of simple models or weak learners is used to model the data. The algorithm works by performing a number of sequential boosting steps. In a step, two sets of weights are computed, one for each data point and one for each learner. The weight of the data point is based on the accuracy with which it is predicted. An accurately modeled data point gets a lower weight. This means that in the next step, more effort is directed to the "difficult" points. Each learner gets a weight based on its error rate, and the estimate is a weighted average of the predictions of the individual learners. This means that strong learners get progressively more influence.

Bagging
Bootstrap aggregating or bagging is another ensemble based machine learning method [Breiman (1996)]. Here, the bagging function from the R package ada was used. The idea is to base the modeling on n random samples (with replacement) of size m' from the original data set of size m. A classification model is then trained on each sample, and the ensemble prediction is computed by taking the average over the n classification models.

Support Vector Machines
Support vector machines work by transforming the data to a higher dimensional set, and then performing separation in the resulting hyperplane [Cortes and Vapnik (1995)]. Thus, this technique is able to perform separation and classification on complex problems where linear separation is useless. The classification boundary of a SVM is given by:

$$f(x) := h(x)\beta + \beta_0$$

Where h(x) is a non-linear function of the observations x which maps them to a higher-dimensional space. Classification is then performed by assigning an observation to the nearest support vector in the higher dimensional space, which results in highly non-linear classification boundaries in the original space. The actual calculation involves an inner product space H(x,x') of the function h(x). A range of kernel functions H is available. In this study three options available in the R package e1017 were used:

$$linear: H(x, x') = x * x'$$

$$sigmoid: H(x, x') = tanh(\gamma x' x + coef0)$$

$$radial: H(x, x') = e^{-\gamma \|x - x'\|^2}$$

The type of kernel function determines the shape of the classification region around each support vector.

3. Data

The dataset consisted of five “target” indicators, key short-term macro economic indicators, and a broad set of explanatory indicators which potentially contain information on the development of the target indicators. All analyses were performed on each of the five target indicators in turn, in which case the four others were added to the set of explanatory variables. The dataset consisted of monthly observations over the period 1990:01-2013:06. The five target indicators were:

- Consumer Confidence
- Producer confidence
- Production in the manufacturing industry (index)
- Household consumption (index)
- Exports

The explanatory dataset consisted of component indicators from the consumer survey and the manufacturing industry business survey, production in sub-branches of the manufacturing industry (2-digit), imports, stock prices, wages, and consumer and producer confidence in Germany. The total number of indicators in the set was 30, for exact descriptions and details see appendix A. In this study, indicators from the consumer and business survey are collectively termed “confidence indicators”, and the other indicators will be described as “real indicators”. All indicators were corrected for seasonal effects using the Census X12 methodology, and in index form when appropriate. The indicators were selected to be representative of different aspects of the economy, though the options were restricted by the required availability of a monthly time series dating back to 1990.

The data were transformed in several ways, depending on the analysis performed. The aim of this study is predicting the sign of month on month changes in the five key indicators. Thus, for the two confidence indicators simply the month on month absolute change was computed, whilst for the three real indicators the month on month growth rates were computed. If the method tested required this, these changes were transformed to a dummy form, taking the value 0 if the change was smaller or equal than zero, and 1 otherwise.

The explanatory variables were transformed in different ways as well, depending on the type of target indicator and the type of model used, see table 3.1. These transformations were dictated by a need to both maximise the correlation between target indicator and explanatory variables, and to incorporate efficiently information on current and past relevant developments.

Table 3.1; variable types/transformations used in different models

Method	Type input variables	Real	Confidence
Machine learning	Real	Period on period change (+lag 1) 3 month change 6 month change Year on year change	Period on period change (+lag 1) 3 month change 6 month change Year on year change
	confidence	Level Period on period change 3 month change 6 month change Year on year change	Level Period on period change 3 month change 6 month change Year on year change
Model	Real	Period on period change (+ lag 1, lag 2, lag 3 and lag 12)	Period on period change in the year on year change (+ lag 1, lag 2, lag 3 and lag 12)
	confidence	Level (+ lag 1, lag 2, lag 3 and lag 12)	Period on period change (+ lag 1, lag 2, lag 3 and lag 12)

4. Results

The aim of this study is to find the best method to predict the sign of *next* periods period on period changes, which is also a test of the predictability of short-term dynamics. For each of the five key short-term economic statistics predictions were made of the sign of the next periods change, using both model-based approaches and machine learning techniques. The value of t+1 was predicted using all data up to and including t. Thus, for each target indicator a variable was constructed with the value of the change at t+1.

Depending on the method tested, these were entered as a 1(increase)/0(decrease) variable, or just the value of the change itself. Several of the machine learning techniques yield as output a direct classification into 1(+) or 0(-). If the output was a probability (for example GLM-logit models), a value of smaller than or equal to 0.5 was taken to indicate a decrease(-), a value greater than 0.5 an increase(+). For methods which yield a direct prediction of the next periods change, the actual value of the prediction was ignored, only the sign (+/-) was taken into account.

After the necessary transformations, data available for estimation ranged from 1992:01 to 2013:06. Of this sample, the last two and a half years were set aside for the out of sample prediction test. The score on this prediction test is the measure on which all methods are evaluated and compared. The size of 30 observations is rather small in machine learning terms, but quite respectable for an economic forecasting evaluation, and comprises 10% of the total sample. For each method, the predicted change signs were simply compared to the realised values, and the percentage correct was calculated.

To assess how valuable these predictions could be in practice, a benchmark is needed. It is good practice to include a naïve benchmark, i.e. what quality of predictions could be achieved without going to the trouble of building a forecasting model. Here, two types of naïve prediction are included. The first is taking the current periods sign (the period on period change in t) as a prediction for next periods sign. This is a very rudimentary AR(1) model, which might be representative of much actual expectation formation. As table 4.1 shows, in the case of these short-term economic indicators it is a very poor prediction strategy. The variability inherent in these series means that less than half of these predictions are correct. An increase is more likely to be followed by a decrease and vice versa. Formulated in another way, persistence is very low in these series if one focuses on period on period changes. The second naïve benchmark is using the unconditional expectation. This is simply the majority vote; if most realisations are positive (negative), always predict a positive (negative) change. Table 4.1 shows the success rate of this strategy both for the whole sample, and when applying the whole sample unconditional expectation to the test sample. This strategy is more successful, in general resulting in more than 50% correct predictions. The methods tested here need to perform better than these naïve benchmarks to be of value. Table 4.1 presents for each target indicator the performance of both the best machine learning method and the best model base method. For the full results, see appendix B.

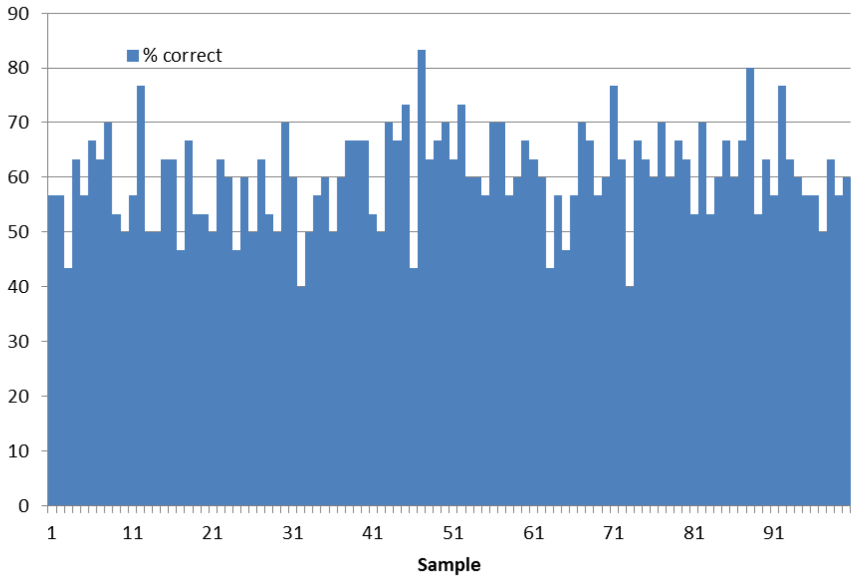
Table 4.1; Results of predicting the sign of next month's change in the five key economic indicators. Accuracy rate (% of signs correct) in 30-month out of sample test set, naïve strategies compared to best performing machine learning technique and best regression type model.

% correct	Naive, t-1	Naive, unconditional	Naive, unconditional test sample	Best Machine Learning	Best Model
				random forest	Factor, LM CC
Producer confidence	43,1	52,5	61,3	73,3	73,3
				SVM sigmoid	Factor, GLM CC
Consumer Confidence	50,1	57,4	64,6	65,5	66,7
				SVM sigmoid	Selection GLM
Consumption	36	56,4	54,8	73,3	80
				Bagging	Selection LM
Manufac-turing production	39,2	56,7	48,4	70	66,7
				Boosting	Selection, GLM
Exports	40,6	59,2	51,6	76,7	66,7

SVM sigmoid=Support vector machine with sigmoid transfer function, Factor LM CC=OLS with common component from dynamic factor analysis, Factor GLM CC=logit model with common component from dynamic factor analysis, Selection GLM/LM=variable selection logit/OLS model.

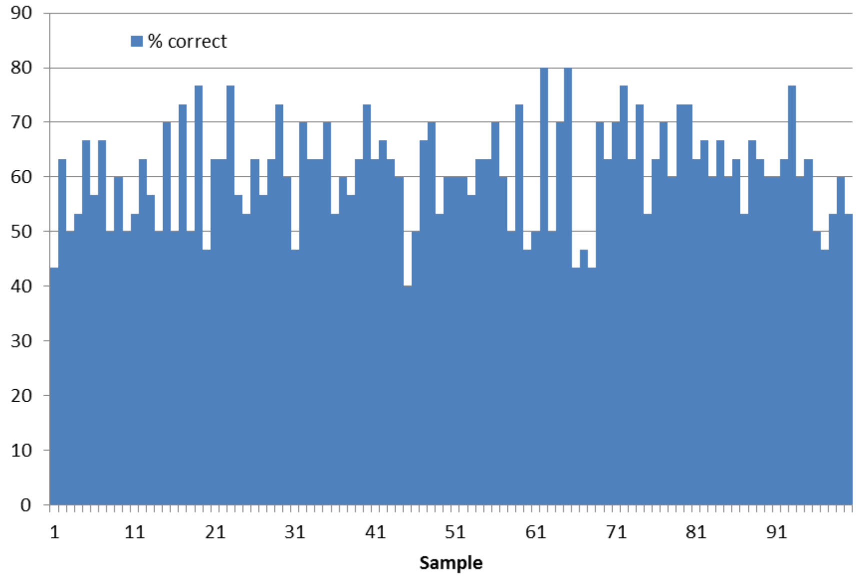
In all cases, the prediction methods perform better than the naïve benchmarks, though in the case of consumer confidence only marginally so. In most other cases, an increase in predictive accuracy of around 20%-points seems to be attainable, a worthwhile improvement. There seems to be no clear winner as to which method performs best. In two cases, machine learning methods have the highest accuracy, in two cases model based methods, and one yields a tie. There is also no dominant method in either of the two classes. However, there is a complicating factor. As appendix B shows, the performance of the machine learning techniques varied quite sharply, with the best and worst performing method for a certain indicator sometimes being as much as 30%-points apart. The intuition was, that this was due to the relatively small sample size of ~250 observations, whilst usually for machine learning exercises samples of several tens of thousands of observations are considered small. To test this idea, the prediction exercise was repeated a hundred times for the machine learning techniques, and for each iteration, the data were randomly divided into a training/test set. Thus, for each iteration the machine learning model was built on a different data set. If the dataset was large/homogenous enough, this should not influence parameter estimation much, and all test sets should be comparable enough to each other and the training data to result in consistent predictive performance. As graphs 4.1 and 4.2 show, this was not the case.

Graph 4.1; % of period on period changes in consumer confidence correctly classified by random forest for different samplings of training/test set.



Graphs 4.1 and 4.2 are two examples of typical results of the sampling exercise for the machine learning techniques. Respectively for the random forest method applied to consumer confidence and support vector machines on producer confidence. In both cases, high variability in the performance over different samplings of the data is visible. Predictive accuracy varies between 45% and 85%, whilst the only difference between the samples is the division of the data over training/test set. Method and variables included are identical.

Graph 4.2; % of period on period changes in producer confidence correctly classified by SVM (sigmoid transfer function) for different samplings of training/test set.



The conclusion here is that in the case of a small data set, the performance of a machine learning technique over on one specific training set is not a good measure of its true relative predictive powers in this context. To test what type of performance on average could be expected of the different techniques applied to the problem considered in this study, the average success rate over the 100 samplings was computed. Table 4.2 shows the techniques which performed best on this measure, again compared to the benchmarks and model based outcomes.

This leads to different conclusions. The average performance of the machine learning techniques is (of course) worse than the best performing ones presented in table 4.1.
Now, model based methods outperform the machine learning techniques. Also, support vector machines with sigmoid transfer functions now dominate the machine learning class.
This does not mean that the results presented in table 4.2 are false, or that those techniques are not capable of performing as well as shown there. It is just the case that one cannot be sure that those methods will perform as well as that in future. The averages probably give a more realistic indication of the predictive quality that can be expected for future use.

Table 4.2; Results of predicting the sign of next month's change in the five key economic indicators. Accuracy rate (% of signs correct) in 30-month out of sample test set, naïve strategies compared to best average performance machine learning technique and best regression type model.

% correct	Naive, t-1	Naive, unconditional	Naive, unconditional test sample	Best Machine Learning	Best Model
				SVM sigmoid	Factor, LM CC
Producer confidence	43,1	52,5	61,3	63,2	73,3
				SVM sigmoid	Factor, GLM
Consumer Confidence	50,1	57,4	64,6	65,5	66,7
				LDA	Selection GLM
Consumption	36	56,4	54,8	65,5	80
				Bagging	Selection LM
Manufac-turing prodcution	39,2	56,7	48,4	65,5	66,7
				SVM sigmoid	Selection, GLM
Exports	40,6	59,2	51,6	65,5	66,7

So for the problem studied here, model based techniques seem to be the best choice. Though one cannot be sure that these will keep performing as well as in these tests, as is the nature of regression models. The dynamic factor models did not perform particularly well as classical forecasting models, with high rmse's. And the models based on automatic variable selection generally contained between the 40 and 80 explanatory variables, rather a lot. However, this means that both the variable selection models and the dynamic factor models rely on a broad range of variables, making them more robust, and diminishing the influence of individual variables.

Another aspect investigated here is whether performance is improved by performing (a crude form of) variable selection, or whether with the methods considered here it is preferable to use the full data set. For each of the five target indicators, a rough and ready indicator set was constructed consisting of variables which based on knowledge of previous research and a few simple tests should have a stronger connection with the target variable. All methods were tested on both the full data set, and the selection. Table 4.3 shows for each target variable/prediction method whether using the full dataset resulted in superior performance, or the selection. Blanks indicate a tie.

Table 4.3; Per predictive technique and key economic indicator the composition of the best performing indicator set. Full=all available indicators, selection=per economic indicator only the most relevant.

	Producer confidence	Consumer confidence	Consumption	IP	Exports
tree	Selection	Selection	Full	Selection	Selection
random forest	Selection	Selection		Selection	Selection
boosting	Full	Full	Full		Full
Bagging	Selection	Full	Selection	Selection	Selection
SVM linear	Full	Full	Full	Selection	Selection
SVM sigmoid		Selection	Selection	Selection	Full
LDA	Selection	Selection		Full	
model selection OLS	Full	Selection	Full	Selection	Full
model selection GLM	Selection	Selection	Selection	Full	Full
DFM 1st static factor OLS	Selection	Selection	Full		
DFM 1st static factor GLM	Selection	Full	Full		
DFM common OLS	Selection	Full	Selection	Full	Full
DFM common GLM	Selection	Full	Selection	Full	
DFM 3 static factors OLS	Selection		Full	Full	Selection
DFM 3 static factors GLM	Full	Full		Full	Full

The outcome is an overall tie. The number of times that the full data set performed better is about equal to the number of times that using the selection worked better. And there is a not negligible number of ties. On the one hand, this is a difficult to interpret result. On the other, it means that it is possible to get good results with these methods without spending too much time on variable selection. It is interesting though that quite a few machine learning techniques perform better when using a pre-selected indicator set.

The final question is whether these methods can be useful in practice. If one wants to qualify the current realisation by testing what next periods change might be, or if one wants to base an analysis on the predicted direction of change, the prediction needs to be reliable to be of value. Weather an expected accuracy rate of around 70% is enough, is debatable. It is a clear improvement over the naïve benchmarks, but is still means that the analysis will be incorrect in 30% of cases, or around 4 months per year. Given the high variability and large noise component of the series considered here, it is questionable whether a higher success rate is achievable.

5. Conclusions

The ever increasing need for current information on economic conditions and the accompanying shift in focus on short-term developments lead to ever increasing demands on analysts of short-term economic developments. This paper proposes as a supportive tool using predictions for the next month of the sign of the change in five key economic indicators. Month on month changes tend to be rather volatile, with frequent alterations between months with positive and negative growth rates. A reliable prediction of the sign of next month's change can be used to assess whether the current month's change is likely to persist or is just transitory.

A range of prediction techniques were tested here, from standard regression type models to advanced machine learning techniques. The five key economic indicators targeted were consumer confidence, household consumption, producer confidence, manufacturing production, and exports, all for the Netherlands. The choice of techniques to be considered was based on a desire for accuracy and robustness. These had to be powerful techniques, but also fast and easy to use in a time constrained operations environment, and they had to be robust to missing data and changes in the relationships between the target indicators and the explanatory variables. This ruled out most classical forecasting model construction techniques, which rely heavily on extensive model evaluation and variable selection. Instead, techniques were selected which either relied on averaging, or on automatic variable selection/weighting.

The models were evaluated on their performance in predicting the sign of the month on month changes in a test set, consisting of the last 30 months of the data set. In this, they were compared to simple naïve prediction strategies. For all five key economic indicators, model-based prediction strategies outperformed naïve ones, though the difference in performance was sometimes only a few percentage points. The success rates varied between 65.5% and 80%. In general, machine learning and regression based techniques performed with similar accuracy. The best regression models were GLM models based on either automatic variable selection or dynamic factor models. There is one problem concerning the machine learning models tested here.

These exhibited widely varying performance, and when a robustness test was performed, it was found that in all cases the average expected performance of the best performing techniques was markedly below the results found for the test sample. This is probably due to the fact that the available data set (~230 observations), whilst respectable for economic forecasting exercises, is rather small for the application of machine learning techniques. These tend to be developed for much larger datasets, meaning that estimation and performance were not stable here, but depended heavily on the division into training and test sample. The overall best technique to use is a GLM combined with dynamic factors. These tend to uniformly perform well, and are quite robust to all sorts of problems one could encounter.

References

Breiman, L. (1984) Classification and regression trees. Chapman & Hall/CRC .

Breiman, L. (1996). "Bagging predictors". Machine Learning 24 (2) p. 123–140.

Breiman, L. (2001). "Random Forests". Machine Learning 45 (1): p. 5–32.

Chauvet, M. and Piger, J.M. (2003) “Identifying business cycle turning points in real time” Review of the Federal Reserve Bank of St. Louis march/april p. 47-61.

Cornec, M and Mikol, F. (2011) “Nowcasting GDP directional change with an application to French business survey data”.

Cortes, C.; and Vapnik, V. (1995). "Support-vector networks” Machine Learning 20 (3) p. 27.

Hamilton, James D. (1989)“ A new approach to the economic analysis of non stationary time series and the business cycle” Econometrica 57(2) p. 357-84.

Forni, M., Halin M., Lippi M., and Reichlin L. (2005) “The Generalized Dynamic Factor Model one-sided estimation and forecasting” Journal of the American Statistical Association vol. 100(471) p. 830-840 .

Reijer, A.H.J. den (2005) “Forecasting Dutch GDP using large scale factor models” DNB working paper Working Paper No. 028/2005.

Schapire, R. E. (1990). "The Strength of Weak Learnability". Machine Learning 5 (2) p. 197–227.

Freund, Y and Schapire, R.E. (1997) “A decision-theoretic generalization of on-line learning and an application to boosting” Journal of Computer and System Sciences, 55(1) p.19-139.

Stock, J.H.. and Watson, M.W. (2002) Journal of Business and Economic Statistics 20(2) p.147-162.

US Census Bureau (2011), X-12-ARIMA reference manual, version 0.3

6. Appendix

Appendix A: Data Set

Indicator	Format	adjustments
Production of manufacturing industry	Index, volume	sa, working days
Household consumption	Index, volume	sa, shopping days
Exports	Index, volume	sa, working days
Consumer confidence	Balance (%)	sa
Producer confidence manufacturing industry	Balance (%)	sa
CS-Economic situation past 12 months	Balance (%)	sa
CS-Economic situation next 12 months	Balance (%)	sa
CS-Personal financial situation past 12 months	Balance (%)	sa
CS-Personal financial situation next 12 months	Balance (%)	sa
CS-Unemployment expectations	Balance (%)	sa
CS-Large purchases	Balance (%)	sa
BS-manufacturing change in production last 3 months	Balance (%)	sa
BS-manufacturing expected change in production next 3 months	Balance (%)	sa
BS-manufacturing assessment order book	Balance (%)	sa
BS-manufacturing assessment foreign order book	Balance (%)	sa
BS-manufacturing employment expectations	Balance (%)	sa
BS-manufacturing price development expectations	Balance (%)	sa

BS-manufacturing assessment level of stocks	Balance (%)	sa
Production Chemical industry	Index, volume	sa
Production Machine industry	Index, volume	sa
Production food industry	Index, volume	sa
Production textile industry	Index, volume	sa
Production Construction materials industry	Index, volume	sa
Production Paper industry	Index, volume	sa
Imports of goods	Index, volume	sa, working days
Hourly wages manufacturing	Index	
Stock market (corporations)	Index	
Stock market (total market)	Index	
Consumer confidence Germany	Balance (%)	sa
Producer confidence manufacturing Germany	Balance (%)	sa

sa=seasonally adjusted, CS=consumer survey, BS=business survey, Balance=%positive-%negative

Appendix B: Detailed results

Table B1; Outcomes for regression model variants per key economic indicator. Accuracy in percentage of correctly predicted signs of month of month change in t+1. Out of sample on test set of last 30 months of total dataset.

		% of signs correct (test sample)	
Model type	Target indicator	Whole set	Selection
Automatic model selection (LM)	Consumer confidence	56.7	60
Automatic model selection (GLM)	Consumer confidence	53.3	60
Ensemble GLM	Consumer confidence	46.7	
Dynamic factor 1st PC LM	Consumer confidence	50	63.3
Dynamic factor 1st PC GLM	Consumer confidence	66.7	63.3
Dynamic factor common factor LM	Consumer confidence	63.3	50
Dynamic factor common factor GLM	Consumer confidence	66.7	63.3
Dynamic factor 3 PC LM	Consumer confidence	53.3	53.3
Dynamic factor 3 PC GLM	Consumer confidence	66.7	63.3
Automatic model selection (LM)	Producer confidence	56.7	43.3
Automatic model selection (GLM)	Producer confidence	60	63.3
Ensemble GLM	Producer confidence	60	
Dynamic factor 1st PC LM	Producer confidence	60	63.3
Dynamic factor 1st PC GLM	Producer confidence	60	66.7
Dynamic factor common factor LM	Producer confidence	63.3	66.7
Dynamic factor common factor GLM	Producer confidence	63.3	73.3
Dynamic factor 3 PC LM	Producer confidence	56.7	70
Dynamic factor 3 PC GLM	Producer confidence	70	63.3
Automatic model selection (LM)	Consumption	80	63.3
Automatic model selection (GLM)	Consumption	63.3	70
Ensemble GLM	Consumption	36.7	
Dynamic factor 1st PC LM	Consumption	46.7	43.3
Dynamic factor 1st PC GLM	Consumption	46.7	36.7
Dynamic factor common factor LM	Consumption	46.7	50
Dynamic factor common factor GLM	Consumption	43.3	53.3
Dynamic factor 3 PC LM	Consumption	46.7	43.3
Dynamic factor 3 PC GLM	Consumption	43.3	43.3
Automatic model selection (LM)	Manufacturing production	63.3	66.7
Automatic model selection (GLM)	Manufacturing production	60	56.7
Ensemble GLM	Manufacturing production	43.3	
Dynamic factor 1st PC LM	Manufacturing production	46.7	46.7
Dynamic factor 1st PC GLM	Manufacturing production	46.7	46.7
Dynamic factor common factor LM	Manufacturing production	63.3	46.7
Dynamic factor common factor GLM	Manufacturing production	56.7	46.7
Dynamic factor 3 PC LM	Manufacturing production	63.3	46.7
Dynamic factor 3 PC GLM	Manufacturing production	63.3	46.7
Automatic model selection (LM)	Exports	60	53.3

Automatic model selection (GLM)	Exports	66.7	63.3
Ensemble GLM	Exports	50.0	
Dynamic factor 1st PC LM	Exports	53.3	53.3
Dynamic factor 1st PC GLM	Exports	53.3	53.3
Dynamic factor common factor LM	Exports	60	53.3
Dynamic factor common factor GLM	Exports	53.3	53.3
Dynamic factor 3 PC LM	Exports	53.3	60
Dynamic factor 3 PC GLM	Exports	53.3	43.3

Table B2; Outcomes for machine learning techniques per key economic indicator. Accuracy in percentage of correctly predicted signs of month of month change in t+1. Out of sample on test set of last 30 months of total dataset.

	Target indicator	% of signs correct (test sample) Whole set	Selection
decision tree	Consumer confidence	46.6666667	56.6666667
random forest	Consumer confidence	46.6666667	50
boosting	Consumer confidence	56.6666667	53.3333333
Bagging	Consumer confidence	56.6666667	43.3333333
SVM linear	Consumer confidence	60	53.3333333
SVM sigmoid	Consumer confidence	50	63.3333333
LDA	Consumer confidence	40	60
decision tree	Producers confidence	46.7	56.7
random forest	Producers confidence	66.7	73.3
boosting	Producers confidence	66.7	53.3
Bagging	Producers confidence	63.3	70
SVM linear	Producers confidence	56.7	53.3
SVM sigmoid	Producers confidence	63.3	63.3
LDA	Producers confidence	53.3	63.3
decision tree	Consumption	60	50
random forest	Consumption	46.7	46.7
boosting	Consumption	63.3	53.3
Bagging	Consumption	46.7	50
SVM linear	Consumption	80	70
SVM sigmoid	Consumption	70	73.3
LDA	Consumption	70	70
decision tree	Manufacturing production	63.3	66.7
random forest	Manufacturing production	63.3	70
boosting	Manufacturing production	66.7	66.7
Bagging	Manufacturing production	63.3	70
SVM linear	Manufacturing production	56.7	63.3
SVM sigmoid	Manufacturing production	53.3	60
LDA	Manufacturing production	60	56.7
decision tree	Exports	60	63.3
random forest	Exports	56.7	63.3
boosting	Exports	66.7	56.7
Bagging	Exports	56.7	60
SVM linear	Exports	60	63.3
SVM sigmoid	Exports	73.3	60
LDA	Exports	56.7	56.7

Table B3; Outcomes for repeated sampling based machine learning techniques per key economic indicator. Accuracy in average of percentage of correctly predicted signs of month of month change in t+1. In each case, 100 different splits in training/test set were performed, and the test results averaged.

		% of signs correct (test sample)	
	Target indicator	Whole set	Selection
decision tree	Consumer confidence	53.4	54.7
random forest	Consumer confidence	57.9	60.1
boosting	Consumer confidence	58.6	60.8
Bagging	Consumer confidence	61.7	61.5
SVM linear	Consumer confidence	63.1	61.9
SVM sigmoid	Consumer confidence	65.5	60.7
LDA	Consumer confidence	57.3	61.1
decision tree	Producers confidence	55.3	57.2
random forest	Producers confidence	54.6	57.5
boosting	Producers confidence	56.1	56.8
Bagging	Producers confidence	57	57.3
SVM linear	Producers confidence	55.6	63.2
SVM sigmoid	Producers confidence	60.3	60
LDA	Producers confidence	55.6	55.6
decision tree	Consumption	56.6	60.2
random forest	Consumption	54.9	60
boosting	Consumption	60.9	61.5
Bagging	Consumption	60.2	64.1
SVM linear	Consumption	64.9	61.8
SVM sigmoid	Consumption	53.3	60.3
LDA	Consumption	65.5	63.1
decision tree	Manufacturing production	57.1	57.8
random forest	Manufacturing production	58.3	58.1
boosting	Manufacturing production	64	64.7
Bagging	Manufacturing production	65.6	66
SVM linear	Manufacturing production	55.4	57.5
SVM sigmoid	Manufacturing production	61	61.8
LDA	Manufacturing production	53.3	53.5
decision tree	Exports	55.8	56.3
random forest	Exports	61.2	62.1
boosting	Exports	63.5	60.9
Bagging	Exports	61.9	62.9
SVM linear	Exports	55.8	61.4
SVM sigmoid	Exports	60.4	65.1
LDA	Exports	53.5	59.1

Appendix C: Composition of selection sets per key economic indicator

Selection Consumer confidence	Selection producer confidence	Selection production manufacturing industry	Selection Household consumption	Selection exports
Household consumption	Production of manufacturing industry	Production of manufacturing industry	Household consumption	Production of manufacturing industry
Consumer confidence	Household consumption	Household consumption	Consumer confidence	Exports
CS-Economic situation past 12 months	Exports	Exports	Producer confidence manufacturing industry	Consumer confidence
CS-Personal financial situation past 12 months	Consumer confidence	Consumer confidence	CS-Economic situation next 12 months	Producer confidence manufacturing industry
CS-Unemployment expectations	Producer confidence manufacturing industry	Producer confidence manufacturing industry	CS-Personal financial situation next 12 months	CS-Economic situation next 12 months
BS-manufacturing assessment order book	CS-Economic situation past 12 months	CS-Economic situation next 12 months	CS-Unemployment expectations	CS-Personal financial situation next 12 months
BS-manufacturing price development expectations	BS-manufacturing change in production last 3 months	CS-Personal financial situation next 12 months	CS-Large purchases	BS-manufacturing change in production last 3 months
BS-manufacturing assessment level of stocks	BS-manufacturing expected change in production next 3 months	BS-manufacturing change in production last 3 months	BS-manufacturing employment expectations	BS-manufacturing expected change in production next 3 months
Production food industry	BS-manufacturing assessment order book	BS-manufacturing expected change in production next 3 months	Production food industry	BS-manufacturing assessment order book
Production textile industry	BS-manufacturing assessment	BS-manufacturing assessment order	Production textile industry	BS-manufacturing assessment

	foreign order book	book		foreign order book
Production Construction materials industry	BS-manufacturing assessment level of stocks	BS-manufacturing assessment foreign order book	Production Construction materials industry	BS-manufacturing assessment level of stocks
Hourly wages manufacturing	Production Chemical industry	BS-manufacturing assessment level of stocks	Imports of goods	Production Chemical industry
Stock market (total market)	Production Machine industry	Production Chemical industry	Hourly wages manufacturing	Production Machine industry
	Production food industry	Production Machine industry	Stock market (total market)	Production Construction materials industry
	Production textile industry	Production Construction materials industry		Imports of goods
	Production Construction materials industry	Imports of goods		Stock market (total market)
	Production Paper industry	Stock market (total market)		Consumer confidence Germany
	Imports of goods	Consumer confidence Germany		Producer confidence manufacturing Germany
	Stock market (corporations)	Producer confidence manufacturing Germany		
	Consumer confidence Germany			
	Producer confidence manufacturing Germany			

Explanation of symbols

.	Data not available
*	Provisional figure
**	Revised provisional figure (but not definite)
x	Publication prohibited (confidential figure)
–	Nil
–	(Between two figures) inclusive
0 (0.0)	Less than half of unit concerned
empty cell	Not applicable
2013–2014	2013 to 2014 inclusive
2013/2014	Average for 2013 to 2014 inclusive
2013/’14	Crop year, financial year, school year, etc., beginning in 2013 and ending in 2014
2011/’12–2013/’14	Crop year, financial year, etc., 2011/’12 to 2013/’14 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress: Statistics Netherlands, Grafimedia
Design: Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information

60083201420 X-10