Can macroeconomists get rich nowcasting economic turning points with machine-learning?

Thomas Raffinot *

May 2015

Abstract

This paper aims at nowcasting economic cyclical turning points in real time to get useful signals for policymakers and for investors. To nowcast economic turning points, probabilistic indicators are created from a simple machine-learning algorithm known as Learning Vector Quantization (LVQ), introduced in economics by Giusto and Piger (2014). The real-time ability of the indicators to quickly and accurately detect economic turning points in the United States and in the euro area is gauged. To assess the value of the indicators, profit maximization measures based on trading strategies are employed in addition to more standard criteria. A substantial improvement in profit measures over the benchmark is found: macroeconomists can get rich nowcasting economic turning points.

JEL classifications: C53, E32, E37, G11
Keywords: Learning Vector Quantization; Economic turning points detection; Profit maximization measures; Model Confidence Set

**LEDa-SDFi Université Paris Dauphine, Place du maréchal de Lattre de Tassigny, 75016 Paris, France (email: traffinot@gmail.com)*
I thank Anne Epaulard and Marie Bessec for helpful comments.

Introduction

Forecasting economic turning points in real time is a notorious difficult task (see Berge (2013) or Katayama (2010)), but ex-post identification of turning points is also challenging: economists fail to detect if a new economic phase has already begun. For instance, in May 2001, the Survey of Professional Forecasters, conducted by the American Statistical Association and the National Bureau of Economic Research, said there would not be a recession in 2001, even though one had already started. That being said, governments and central banks are naturally sensitive to indicators showing signs of deterioration in growth to allow them to adjust their policies sufficiently in advance. In this respect, timing is important and the earlier the signal, the better.

A significant literature focuses on the business cycle detection (see Hamilton (2011)). The business cycle is meant to reproduce the cycle of the global level of activity of a country[1]. However, Anas and Ferrara (2004) point out the importance for policymakers of the growth cycle, defined as the deviation of the real GDP to its long-term trend[2] and Raffinot (2014) emphasizes the greatest interest of the growth cycle for euro and dollar-based investors. Furthermore, growth cycles are more numerous than business cycles, since all recessions involve slowdowns, but not all slowdowns involve recessions (see Anas and Ferrara (2004)).

One stylised fact of economic cycles is the non-linearity: the behaviour of the series describing the cycle depends on the phase in which it evolves. Real-time regime classification and turning points detection require thus methods capable of taking into account the non-linearity of the cycles. In this respect, many non-linear parametric models have been proposed, such as smooth-transition autoregressive models (see Ferrara and Guegan (2005)), non-linear probit models (see Liu and Moench (2014)) or Markov switching models (see Guidolin (2011)).

Recently, Giusto and Piger (2014) introduce a very simple machine-learning algorithm known as Learning Vector Quantization (LVQ), which appears very competitive with commonly used alternatives. They demonstrate that parametric models are effective if the true data generating process (DGP) linking the observed data to the economic regime is known. Otherwise, in the absence of knowledge of the true DGP, they advocate non-parametric methods, as they do not rely on a specification of the DGP. In particular, the learning vector quantization (LVQ) algorithm introduced by

[1] The turning points of that cycle separate periods of recessions from periods of expansions

[2] the turning points of that cycle separate periods of slowdowns and accelerations

Kohonen (2001) has been successfully applied to various supervised classification tasks such as radar data extraction or spam email detection.

This paper aims at creating indicators, which could quickly and accurately detect growth cycle turning points in real time. Probabilistic indicators are thus created from Learning Vector Quantization to nowcast growth cycle turning points in real time, not only in the United States but also in the euro area. Useful signals for policymakers and for investors are produced: a substantial improvement in profit measures over the benchmark is found, in the United States and in the euro area.

The rest of the paper proceeds as follows. Section 1 introduces Learning Vector Quantization. Section 2 describes the data, the turning point chronology, the model selection and the evaluation of the forecasts. Section 3 analyses the empirical results.

1 Learning Vector Quantization

The LVQ is a prototype-based supervised classification algorithm and the basic idea of this algorithm is to find a natural grouping in a set of data. As supervised method, LVQ uses known target output classifications for each input pattern of the form. LVQ algorithms do not approximate density functions of class samples like Vector Quantization or Probabilistic Neural Networks[3] do, but directly define class boundaries based on prototypes, a nearest-neighbour rule and a winner-takes-it-all paradigm. In other words, LVQ takes both historical data and its classification as an input, which is then used to train the algorithm. Based on this training, new data that has not yet been classified is then labelled as arising from the slowdown and acceleration regime.

Giusto and Piger (2014) emphasize that LVQ has some computational advantages over parametric methods: the algorithm is very simple to implement and can be easily modified to incorporate data series that arrive with different reporting lags, as well as data series of mixed reporting frequency. Moreover, LVQ can be implemented when there is a large number of indicators with little increase in computational complexity.

The main idea of the algorithm is to cover the input space of samples with "codebook vectors" (CVs), each representing a region labelled with a class. Once these relevant points are singled out, data is classified to belong to the same class as the closest codebook vector. In this paper, the Euclidean metric is used as it is the dominant metric in the literature.

[3]LVQ can be understood as a special case of an artificial neural network: LVQ is a feedforward net with one hidden layer of neurons, fully connected with the input layer

Learning means modifying the weights in accordance with adapting rules and, therefore, changing the position of a CV in the input space. Since class boundaries are built piecewise-linearly as segments of the mid-planes between CVs of neighbouring classes, the class boundaries are adjusted during the learning process. The tessellation induced by the set of CVs is optimal if all data within one cell indeed belong to the same class.

Several versions of the prototype update rule are known as LVQ1, LVQ2, LVQ3, or OLVQ, which differ in robustness and the rates of convergence. Venkatesh and Thangaraj (2010) conclude that LVQ1 produced better grouping of genes compared to other variants.

Giusto and Piger (2014) describe LVQ1 algorithm as follows. Let X be a collection of N observations $x_n \in \mathbb{R}^m, n = 1, ..., N$ for which the classification in the set $\{C_k\}_{k=1}^K$ is known. Let there be $\bar{N} \in [K, N]$ codebook vectors $m_i \in \mathbb{R}^m, i = 1, 2, ..., \bar{N}$ with given initial locations. Finally, let $g = 1, 2, ..., G$ denote iterations of the algorithm and let α^g be a decreasing sequence of real numbers bounded between zero and one. Given the initial location of the $\bar{N}$ codebook vectors, the LVQ1 algorithm makes adjustments to their location through these steps:

Step 1 Let $g = 1$ and $n = 1$

Step 2 Identify the codebook vector m_g^c closest to the data point x_n

$$c = argmin_{i \in \{1,...,\bar{N}\}} \|x_n - m_i^g\|$$

Step 3 Adjust the location of the codebook vector with index c according to the following rule:

$$\begin{cases} m_c^{g+1} & = m_c^g + \alpha^g [x_n - m_c^g] \text{ if } x_n \text{ and } m_c^g \text{ belong to the same class} \\ m_c^{g+1} & = m_c^g - \alpha^g [x_n - m_c^g] \text{ otherwise} \end{cases}$$

Step 4 If $n + 1 \leq N$ let $n = n + 1$ and repeat from step 2. Otherwise let $n = 1$ and $g = g + 1$ and if $g \leq G$ repeat from step 2; stop otherwise.

Classification after learning is based on a presented sample's vicinity to the CVs: the classifier assigns the same class label to all samples that fall into the same tessellation. Giusto and Piger (2014) emphasize that this process is analogous to the in-sample parameter estimation and out-of-sample prediction steps employed with parametric statistical models and they describe the classification process for new data as follows.

Let x_{N+1} a new point for which the classification is unknown. Its class is predicted by first finding the codebook vector m_c that is closest to x_{N+1} in the Euclidean metric:

$$c = argmin_{i \in \{1,\ldots,\bar{N}\}} \|x_{N+1} - m_i\|$$

and then x_{N+1} is assigned to the same class as is assigned to codebook vector m_c.

To define parameters $\bar{N}$, α, G, Kohonen (2001) recommendations, which are based on a survey of a large number of empirical implementations of the LVQ algorithm, are followed. $\bar{N}$, the total number of codebooks, is set to 70 and the same number of codebooks is assigned to each class. The parameter α equals 0.3, while the number of algorithm iterations, G, is set to 40 times the total number of codebook vectors.

At last, Giusto and Piger (2014) have tested several approaches to initialize the codebook vectors (k-means, k-medoids, and self-organizing maps) and recommend to initialize the codebook vectors to a randomly drawn set of data vectors in their training sample.

2 Model selection

2.1 Data set

The real-time detection of turning points faces the difficult issues of late release dates and data revision. As a matter of fact, key statistics are published with a long delay, are subsequently revised and are available at different frequencies. For example, gross domestic product (GDP) is only available on a quarterly basis with a time span of one to three months, and sometimes with significant revisions.

However, a range of monthly economic series are released giving indications of short-term movements. Among them, business surveys provide economists for timely and reliable pieces of information on business activity. They are subject to very weak revisions, are usually less volatile than other monthly series. They are published before the end of the month they relate to or just a few days after. In the euro area, surveys published by the European Commission have been proven to be very effective (see Raffinot (2007)). In the United States, the surveys published by the Institute for Supply Management(ISM), the Conference Board and the National Association of Home Builders (NAHB) are often tested in the literature (see Liu and Moench (2014)).

Moreover, financial series, which are not revised and often available on a daily basis, have also been considered: the yield curve, defined as the difference between the ten-year and the three-month yield, the level and curvature of the yield curve (see Chauvet and Senyuz (2012) or Berge (2013)). Other financial series are the investment-grade and high-yield corporate spreads, stock markets (S&P500, Eurostoxx), stock markets volatility (see Chauvet et al. (2010)), the VIX index and the VSTOXX index, which is the VIX equivalent for the euro area. It should be added, that this papers uses end of month values to match stock index futures and options contracts settlement prices.[4]

Finally, some real economic data have been tested, such as the four-week moving average of initial claims for unemployment insurance, which is a weekly measure of the number of jobless claims filed by individuals seeking to receive state jobless benefits.[5]

To detect the turning points in real-time, not only original series are screened, but also differentiated series (to underline the phases of low and high pace of growth). Because of the classical trade-off between reliability and advance, different lags of differentiation were considered: 3 months, 6 months, 9 months, 12 months. The large dataset of predictors consists of more than 180 monthly variables in the euro area and 110 in the United States.

2.2 Turning point chronology

To implement the LVQ classifier, a classification of economic regimes is needed. This paper employs the turning point chronology established in Raffinot (2014) (see AppendixA). In the euro area, the number of periods identified as slowdown equals the number of periods identified as acceleration. In the United States, 71 % of the data are classified as acceleration. Over the period from January 1999 to December 2013, there were 9 turning points in the growth cycle in the euro area and in the United States.

In real time, the complete chronology is not available, but the monthly GDP introduced by Raffinot (2007)[6] allows to quickly refine the turning point chronology. In this paper, for the recursive estimation of the models, the LVQ classifier is trained each month on a sample that extends from the beginning of the sample through month $T-12$, over which the turning point chronology is assumed known.

[4]http://www.cmegroup.com/trading/equity-index/fairvaluefaq.html

[5]All series are provided by Datastream.

[6]A temporal disaggregation based on business surveys of the non revised values of gross domestic product GDP is used to develop a monthly indicator of GDP

2.3 Model evaluation

2.3.1 Classical criteria

Two metrics are computed to evaluate the quality of classification of a model.

The first one is the famous Brier's Quadratic Probability Score (QPS), defined as follows:

$$QPS = \frac{1}{T}\sum_{t=1}^{T}(P_t - R_t)^2$$

where $t = 1, ..., T$ is the number of forecasts from the LVQ classifier. The best model should strive to minimize the QPS.

The second one is the area under the Receiver Operating Characteristics (ROC) curve. The ROC curve describes all possible combinations of true positive and false positive rates that arise as one varies the threshold c used to make binomial forecasts from a real-valued classifier: as c is varied from 0 to 1, the ROC curve is traced out in $T_p(c), F_p(c)$ space that describes the classification ability of the model. Accuracy is measured by the area under the ROC curve (AUROC). An area of 1 represents a perfect test, an area of 0.5 represents a worthless test. A general rule of thumb is that an AUROC value > 0.85 indicates a useful prediction performance. The ROC curve is a useful measure, because it precisely captures the ability of each model to accurately categorize economic phases. In particular, by using the area under the ROC curve (AUROC), one can evaluate the categorization ability of the model over an entire spectrum of different cutoffs for determining a chosen economic phase, instead of evaluating predictive power at any one arbitrary threshold.

2.3.2 Profit maximization measures

In addition to more standard criteria, profit maximization measures are employed. Indeed, Kollar (2013) and Raffinot (2014) point out that various asset classes behave differently during different phases of the economic cycles. These applications should illustrate that profit maximization measures are helpful to select the best model and that investors could have earned money based on economic regimes induced by the probabilistic indicator.

In order to frame the concept of active portfolio management, a specified investment strategy is required. We first consider an equity portfolio manager investing 100€ or 100$ on January 1, 2001. Each month, the investor decides upon the fraction of wealth to be invested based on the current state of the economy induced by the indicator that would have been available at the time the decision was made. If the probabilistic indicator classifies the period as acceleration, then the investor can leverage his portfolio (120% of his wealth is invested on the asset and 20% of cash is borrowed), otherwise he only invests 80% of his wealth and 20% is kept in cash.

Moreover, if different asset classes perform differently during different stages of the growth cycle, it might be reasonable to rebalance the portfolio (shifting allocation weights) based on the stage of the growth cycle. The second strategy aims at beating the classic asset allocation for an institutional portfolio, *i.e.* 60% of the portfolio allocated to equities and 40% to fixed income securities (bonds). The investor decides each month to rebalance his portfolio. If the probabilistic indicator indicates acceleration, then 80% of the portfolio is allocated to equities and 20% to bonds, otherwise 40% of the portfolio is allocated to equities and 60% to bonds.

Berkowitz et al. (1988), Pesaran and Timmermann (1994), Johanning (2002) and Han et al. (2013) demonstrate that the total cost of transactions appears to be low, less than 1% (around 50 basis points when trading in stocks while the cost for bonds is 10 basis points). To simplify, no transaction costs are considered.

The active strategies are then compared with the buy-and-hold strategy, henceforth a passive strategy. For conventional comparison of the portfolio performances, annualized average returns, annualized standard deviation (volatility), performance to volatility ratio, max drawdown (MDD) are computed. The performance to volatility ratio compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) is an indicator of permanent loss of capital. It measures the largest single drop from peak to bottom in the value of a portfolio. In brief, the MDD offers investors a worst case scenario.

2.4 Model selection

Two economic phases are considered: slowdown and acceleration. Applied to the context of nowcasting, it can be summarized as follows:

$$R_t = \begin{cases} 1, & \text{if in acceleration} \\ 0, & \text{otherwise} \end{cases}$$

The training sample runs over the period from January 1988 to December 1998. The performance of the LVQ classifier is then evaluated over the period from January 1999 to December 2013. A recursive estimation is done: each month the model is estimated with the data and the chronology that would have been available at the time the nowcasting is done. For instance, in

January 2012, the chronology that would have been available to implement the LVQ classifier runs over the period from January 1988 to January 2011.

In the empirical analysis, the turning point identification procedure is based on 100 runs of the LVQ algorithm with different random initializations for codebook vectors, therefore providing 100 results of either 1 or 0. The mean of those results is computed and is assumed to be the probability P_t of being in the regime 1 (acceleration). For a given covariate x_n:

$$P_t^x = E[R_t = 1|x_t]$$

A threshold of 0.5 appears logical to classify regimes:

$$\hat{R_t^x} = \begin{cases} 1, & \text{if } P_t^x \geq 0.5 \\ 0, & \text{otherwise} \end{cases}$$

Ng (2014) suggest that cross-correlation of regressors in large datasets might result in inaccurate forecasts and hence a smaller set is more likely to provide a smaller average forecast error. To narrow down the predictors to only those that are relevant, the identification of a first candidate set of time series of interest is done, by examining the series one by one. The best 15 series according to the QPS are retained.

All possible combinations of 4 variables from the selected predictors are computed. We get thus a probability for each model: P_t^{model}. It should be noted that 4 is an arbitrary choice, consistent with the number of series used in Stock and Watson (1989) to estimate parameters of the model.

To avoid data snooping (see White (2000)), which occurs when a given set of data is used more than once for purposes of inference or model selection, the model confidence set (MCS) procedure proposed by Hansen et al. (2011) is computed. The MCS procedure is a model selection algorithm, which filters a set of models from a given entirety of models. The resulting set contains the best models with with a probability that is no less than $1 - \alpha$ with α being the size of the test (see Hansen et al. (2011)).

An advantage of the test is that it not necessarily selects a single model, instead it acknowledges possible limitations in the data since the number of models in the set containing the best model will depend on how informative the data are.

More formally, define a set M_0 that contains the set of models under evaluation indexed by: $i = 0, ..., m_0$. Let $d_{i,j,t}$ denotes the loss differential between two models by

$$d_{i,j,t} = L_{i,t} - L_{j,t}, \forall i, j \in M_0$$

L is the loss calculated from some loss function (QPS in this paper) for each evaluation point $t = 1, ..., T$. The set of superior models is defined as:

$$M^* = \{i \in M_0 : E[d_{i,j,t}] \leq 0 \, \forall j \in M_0\}$$

The MCS uses a sequential testing procedure to determine M^*. The null hypothesis being tested is:

$$\begin{cases} H_{0,M} : E[d_{i,j,t}] = 0 \, \forall i, j \in M \text{ where } M \text{ is a subset of } M_0 \\ H_{A,M} : E[d_{i,j,t}] \neq 0 \text{ for some } i, j \in M \end{cases}$$

When the equivalence test rejects the null hypothesis, at least one model in the set M is considered inferior and the model that contributes the most to the rejection of the null is eliminated from the set M. This procedure is repeated until the null is accepted and the remaining models in M now equal $\widehat{M}^*_{1-\alpha}$.

According to Hansen et al. (2011), the following two statistics can be used for the sequential testing of the null hypothesis:

$$t_{i,j} = \frac{\overline{d}_{i,j}}{\sqrt{\widehat{var}(\overline{d}_{i,j})}} \text{ and } t_i = \frac{\overline{d}_i}{\sqrt{\widehat{var}(\overline{d}_i)}}$$

where m is the number of models in M, $\overline{d}_i = (m-1)^{-1} \sum_{j \in M} \overline{d}_{i,j}$, is the simple loss of the i^{th} model relative to the averages losses across models in the set M, and $\overline{d}_{i,j} = (m)^{-1} \sum_{t=1}^{m} d_{i,j,t}$ measures the relative sample loss between the i^{th} and i^{th} models. Since the distribution of the test statistic depends on unknown parameters a bootstrap procedure is used to estimate the distribution.

Furthermore, to refine the selection, the test proposed by Hanley and McNeil (1982) to compare the AUROC predictive accuracy is then computed. The aim is to test the best models in the selection with another criteria, thereby further reducing the set. The t-statistic for the test of $H_0 : AUROC_1 = AUROC_2$ is given by:

$$t = \frac{AUROC_1 - AUROC_2}{\sqrt{(\sigma_1^2 + \sigma_2^2 - 2r\sigma_1 * \sigma_2}}$$

where, AUROC1 and AUROC2 are the areas under the curve for models 1 and 2 which are being compared. Similarly, σ_1 and σ_2 refer to the variances of the AUROCs for model 1 and model 2, respectively. Finally, r is the correlation between the two AUROCs (see Hanley and McNeil (1982) or Liu and Moench (2014) for more details on the test statistic and its implementation.)

For investors, the usefulness of a forecast depends on the rewards associated with the actions taken by the agent as a result of the forecast. The investment strategies are computed to assess if the investor could get substantial improvement in profit measures over the benchmark. It should be noted that in the "real life", the investor would not have known which models were selected, as the classifier is evaluated over the period from January 1999 to December 2013. The objective is thus to prove that the best models would have generated significant profits. Otherwise, even the best selected models are useless.

3 Empirical results

3.1 United States

Results obtained for the United States are first presented. Table 1 displays classical metrics for the best two selected models[7]. Those models are very close: only one component differs. The components of the first model are:

-Average weekly initial claims for unemployment insurance

-High-yield corporate spreads (BofA Merrill Lynch US High Yield 100 Index)

-Current business conditions (Conference Board)

-S&P500

As regards the second model, instead of the S&P500, the ISM Purchasing Managers Index (manufacturing survey) is introduced.

The choice of the predictors is consistent with other studies. In particular, Ng (2014) concludes that risky bonds and employment variables have been proved to have predictive power, when the interest rate spreads were uninformative.

The performance of the models are impressive and are consistent with the results found in Berge (2013), as regards nowcasting.

[7]In this paper, the confidence level for the MCS is set as low as possible to only select few superior models in the resulting set. Indeed, the aim of this study is to highlight that nowcasting turning points allow to make profits. If the models would have failed to achieve this objective, the resulting set would have been made bigger

Table 1: Classical evaluation criteria in the United States

	AUROC	QPS
Model 1	0.989	0.030
Model 2	0.987	0.034

Note: This table reports three classical metrics used to evaluate the quality of classification of the best two models: the area under the ROC curve (AUROC) and the Brier's Quadratic Probability Score (QPS).

The t-statistic for the test proposed by Hanley and McNeil (1982) is 0.81: no model is more effective than the other. It turns out that it is difficult to chose a better model based on classical metrics.

The ability to produce profits it now tested. Table 2 emphasizes that active investment strategies based on the growth cycle outperform the passive buy-and-hold benchmark. Indeed, they substantially improve the risk-return trade-off. The reduction of the MDD, which focuses on the danger of permanent loss of capital as a sensible measure of risk, is what risk-averse investors value the most. The first model seems to provide more useful signals as it gets better returns and a better performance to volatility ratio. It should be noted that these results have also implications for the risk management and hedging. Especially, in the options market one can utilize the current state of the economy to hedge the portfolio against the possible price declines. For example, besides following one of the previous strategy, writing an out-of-money covered call or buy a put option when the stock market is expected to decrease (slowdown) would limit the losses.

Moreover, dynamic asset allocation delivers a substantial improvement in risk-adjusted performance as compared to static asset allocation, especially for investors who seek to avoid large losses. This time, the selection of the model is quite tricky: the second model improves substantially the MDD, but the first model displays better returns and a better performance to volatility ratio. Risk adversed investor should thus choose the second model, whereas other investors should focus on the first model.

Table 2: Summary of return and risk measures in the United States

United States (2001-2013)			
	Model 1	Model 2	Benchmark
Equities			
Average returns	9.6	9.2	6.3
Volatility	15.0	14.8	15.2
Performance to volatility ratio	0.64	0.62	0.42
MDD	-43.7	-43.5	-50.9
Asset allocation			
Average returns	8.9	8.6	5.9
Volatility	7.3	7.1	7.1
Performance to volatility ratio	1.22	1.21	0.83
MDD	-22.4	-19.4	-23.5

Note: This table reports profit maximization measures for the active strategies based on the state of the growth cycle induced by the model: a 120/80 equity strategy and an equity-bond asset allocation strategy. Returns are monthly and annualized. The volatility corresponds to the annualized standard deviation. The performance to volatility ratio compares the expected returns of an investment to the amount of risk undertaken to capture these returns. The Max drawdown (MDD) measures the largest single drop from peak to bottom in the value of a portfolio.

The figure 1 illustrates the behaviour of the first model in real time. The indicator gives reliable signals. This indicator is not volatile and display a very high persistence in the regime classification. November 2012 is the only false signal.

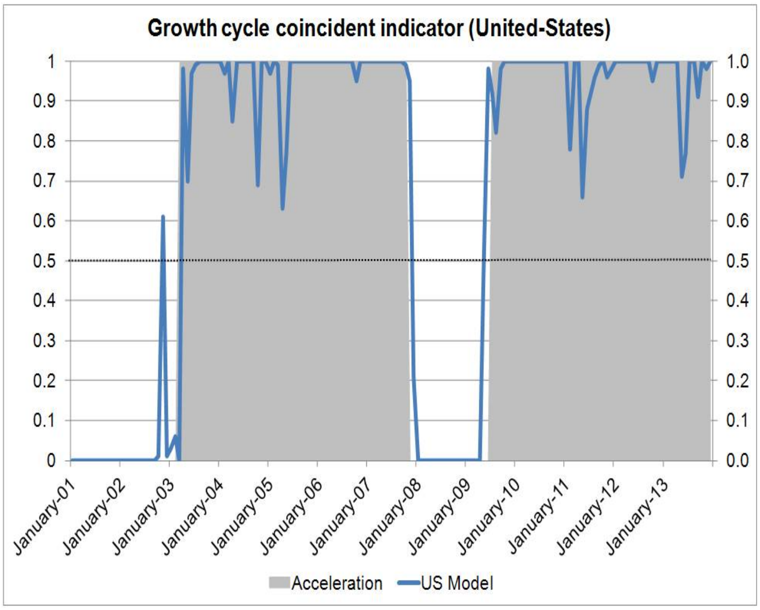

Figure 1: Recursive real time classification of the growth cycle in the United States

To sum up, the real time classification induced by the indicator should be considered as enlightening for policymakers and dollar-based investors.

3.2 Euro area

The results for the euro area are now presented. Table 3 highlights classical metrics for the best models in the euro area. Just like for the United States, only two models are selected. Again, these models are very close and only one component differs. The components of the first model are:

-Employment expectations for the months ahead in the consumer survey

-Financial situation over the last 12 months in the consumer survey

-Production expectations for the months ahead in the industry survey

-Employment expectations for the months ahead in the construction survey

As regards the second model, the last component is the price expectations for the months ahead in the construction survey. It should be noted that financial variables are not introduced in the indicators. This result may stem from the preponderance of bank loans in corporate financing. About 70% of firms' external financing in the euro area comes via the banking system, compared with only 30% in the United States. It implies that the corporate bond market is still small in the euro area. The choice of the predictors is a little bit different but stays in line with other studies, especially Anas et al. (2008)[8].

The choice between models is difficult to perform. The t-statistic for the test proposed by Hanley and McNeil (1982) is 1.12: no model is more effective than the other. It should be noted that metrics in the euro area are less impressive than in the United States. Indeed, the persistence of the regimes is smaller in the euro area growth cycle, the real-time classification is thus harder.

[8] A comparison of LVQ against competing parametric models comes naturally to mind. However, parametric models in academic studies often employ hard data, such as industrial production or unemployment. The delay of publication and the revisions of the data make a "pseudo real-time" comparison very hard to perform. Moreover, a simple transformation of the selected series into a MS-VAR has been performed and the comparison favours the LVQ. However, the results are not judged conclusive, since a proper selection of the covariates for the MS-VAR could have led to another results, or not. A proper comparison is left for future research.

Table 3: Classical evaluation criteria in the euro area

	AUROC	QPS
Model 1	0.960	0.081
Model 2	0.962	0.083

Note: see table1.

Table 4 emphasizes that active investment strategies based on the growth cycle outperform the passive buy-and-hold benchmark. The model 2 seems to provide more useful signals as it reduces the MDD and gets better returns and a better performance to volatility ratio. Naturally, the risk management and hedging implications described for the United States also apply in the euro area.

Dynamic asset allocation delivers a substantial improvement in risk-adjusted performance as compared to static asset allocation. One more time, the second model seems more suitable, as it presents a better returns and a better performance to volatility ratio.

Table 4: Summary of return and risk measures in the euro area

Euro area (2001-2013)			
	Model 1	Model 2	Benchmark
Equities			
Average returns	4.9	5.2	2.7
Volatility	17.1	17.2	19.1
Performance to volatility ratio	0.29	0.30	0.14
MDD	-45.9	-45.1	-53.7
Asset allocation			
Average returns	7.0	7.2	4.2
Volatility	7.7	7.7	9.2
Performance to volatility ratio	0.92	0.93	0.46
MDD	-19.5	-19.5	-27.1

Note: see table2

The figure 2 illustrates the behaviour of the second model in real time. The euro area probabilistic indicator appears more volatile than the United States probabilistic indicator. Yet, low phases detected are not erratic and display a very high persistence. There are only three false signals, which do not last for a long time: September 2003, September 2004 and June 2012. Hence real time regime classification should be considered as enlightening for policymakers and euro-based investors.

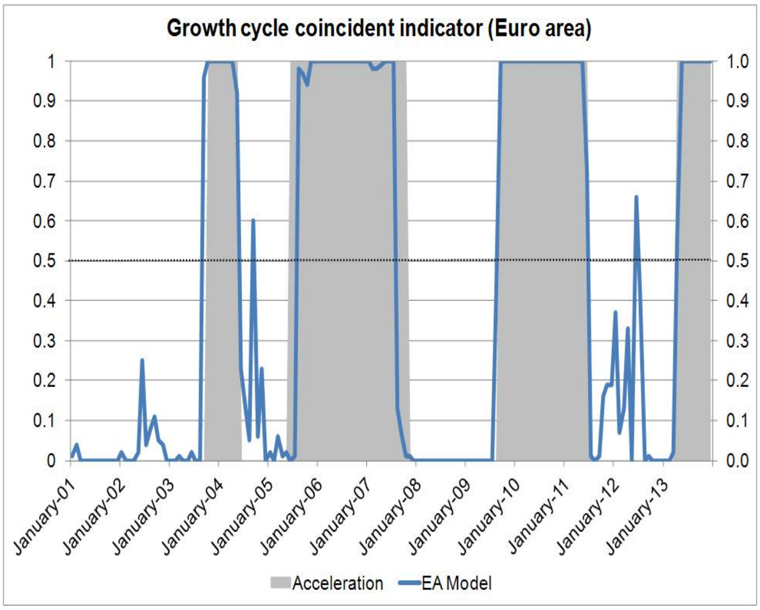

Figure 2: Recursive real time classification of the growth cycle in the euro area

Conclusion

Anas and Ferrara (2004) emphasize the relevance for policymakers of the growth cycle, which seeks to represent the fluctuations around the trend, and Raffinot (2014) highlights the great interest of the growth cycle for euro and dollar-based investors. This paper aims at creating indicators, which could quickly and accurately detect growth cycle turning points in real time. To this end, this study employs a simple machine-learning algorithm known as Learning Vector Quantization (LVQ), recently introduced in economics by Giusto and Piger (2014). To select the best model, profit maximization measures based on trading strategies are used in addition to more standard criteria, such as Quadratic Probability Score and area under the ROC curve (AUROC). The results provide evidence that LVQ is very effective, despite its simplicity, and that some economic and financial indicators can be exploited to quickly identify turning points in real time in the United States and in the euro area. It leads to useful implications for investors practising active portfolio and risk management and for policy makers as tools to get early warning signals.

Last but not least, this article opens the door for further research. A comparison between alternative methods to nowcast the growth cycle comes naturally to mind. An attempt to forecast growth cycle turning points three to twelve months ahead could be very interesting. Finally, evaluating the diversifying power of alternative asset classes based on complex allocation methods in real-time may be newsworthy for investors.

Appendix A: Turning point chronology

The complete chronology is contained in the table 5:

Table 5: **Turning point chronology**

Euro area (Jan 1989-December 2013)		**United States (Jan 1985-Dec 2013)**	
Trough D	March 1991	Peak A	November 1985
Peak A	August 1993	Trough D	April 1987
Trough D	March 1991	Peak A	December 1989
Peak A	March 1995	Trough D	August 1991
Trough D	December 1996	Peak A	January 1993
Peak A	March 1998	Trough D	July 1993
Trough D	February 1999	Peak A	September 94
Peak A	December 2000	Trough D	March 1996
Trough D	September 2003	Peak A	June 2000
Peak A	May 2004	Trough D	February 2003
Trough D	May 2005	Peak A	October 2007
Peak A	October 2007	Trough D	September 2009
Peak B	March 2008	Peak A	June 2011
Trough C	April 2009	Trough D	December 2011
Trough D	August 2009		
Peak A	June 2011		
Peak B	August 2011		
Trough C	November 2012		
Trough D	March 2013		

References

Anas, J., Billio, M., Ferrara, L., and Mazzi, G. L. (2008). A system for dating and detecting turning points in the euro area. *Manchester School*, 76(5):549–577.

Anas, J. and Ferrara, L. (2004). Detecting Cyclical Turning Points: The ABCD Approach and Two Probabilistic Indicators. *Journal of Business Cycle Measurement and Analysis*, 2004(2):193–225.

Berge, T. (2013). Predicting recessions with leading indicators: model averaging and selection over the business cycle. Research Working Paper RWP 13-05, Federal Reserve Bank of Kansas City.

Berkowitz, S., Logue, D., and Noser, E. (1988). The Total Cost of Transactions on the NYSE. *Journal of Finance*, 43(1):97–112.

Burns, A. and Mitchell, W. (1946). *Measuring Business Cycles*. Number burn46-1 in NBER Books. National Bureau of Economic Research, Inc.

Chauvet, M. and Senyuz, Z. (2012). A dynamic factor model of the yield curve as a predictor of the economy. Technical report.

Chauvet, M., Senyuz, Z., and Yoldas, E. (2010). What does financial volatility tell us about macroeconomic fluctuations? MPRA Paper 34104, University Library of Munich, Germany.

Ferrara, L. and Guegan, D. (2005). Detection of the industrial business cycle using setar models. *Journal of Business Cycle Measurement and Analysis*, 2005(3):353–371.

Giusto, A. and Piger, J. (2014). Identifying business cycle turning points in real time with vector quantization. Technical report.

Guidolin, M. (2011). Markov Switching Models in Empirical Finance. Technical report.

Hamilton, J. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2):357–84.

Hamilton, J. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70.

Hamilton, J. (1996). Specification testing in Markov-switching time-series models. *Journal of Econometrics*, 70(1):127–157.

Hamilton, J. (2011). Calling recessions in real time. *International Journal of Forecasting*, 27(4):1006–1026.

Han, Y., Yang, K., and Zhou, G. (2013). A new anomaly: The cross-sectional profitability of technical analysis. *Journal of Financial and Quantitative Analysis*, 48:1433–1461.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36.

Hansen, P., Lunde, A., and Nason, J. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Harding, D. and Pagan, A. (2001). Extracting, Using and Analysing Cyclical Information. MPRA Paper 15, University Library of Munich, Germany.

Johanning, L. (2002). Transaction costs and strategic trading of german investment management firms: Empirical evidence from european stock markets. Technical report.

Katayama, M. (2010). Improving recession probability forecasts in the u.s. economy. Technical report, Louisiana State University.

Kohonen, T. (2001). *Self-Organizing Maps*. Springer.

Kollar, M. (2013). A sketch of macro based asset allocation. *International Journal of Economic Sciences*, 2:101–120.

Liu, W. and Moench, E. (2014). What predicts U.S. recessions? Staff Reports 691, Federal Reserve Bank of New York.

Ng, S. (2014). Viewpoint: Boosting recessions. *Canadian Journal of Economics*, 47(1):1–34.

Pesaran, H. and Timmermann, A. (1994). Forecasting stock returns : an examination of stock market trading in the presence of transaction costs. *Journal of forecasting*, 13(4):335–367.

Raffinot, T. (2007). A monthly indicator of GDP for Euro-Area based on business surveys. *Applied Economics Letters*, 14(4):267–270.

Raffinot, T. (2014). A simple and efficient macroeconomic framework for asset allocation. Technical report.

Stock, J. and Watson, M. (1989). New Indexes of Coincident and Leading Economic Indicators. In *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, pages 351–409. National Bureau of Economic Research, Inc.

Stock, J. and Watson, M. (2010a). Estimating Turning Points Using Large Data Sets. NBER Working Papers 16532, National Bureau of Economic Research, Inc.

Stock, J. and Watson, M. (2010b). Indicators for Dating Business Cycles: Cross-History Selection and Comparisons. *American Economic Review*, 100(2):16–19.

Venkatesh, E. T. and Thangaraj, D. P. (2010). Investigation of micro array gene expression using linear vector quantization for cancer. *International Journal on Computer Science and Engineering*, pages 2114–2116.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica*, 68(5):1097–1126.

Zarnowitz, V. and Ozyildirim, A. (2006). Time series decomposition and measurement of business cycles, trends and growth cycles. *Journal of Monetary Economics*, 53(7):1717–1739.