ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Forecasting Financial Markets with Classified Tactical Signals

Patrick Kouontchou[1], Amaury Lendasse[2], Yoan Miche[2] and Bertrand Maillet[3]

1- Variances and University de Lorraine (CEREFIGE).
Ile du Saulcy, 57045 Metz cedex 01 - France

2- Aalto University School of Science, ICS Department
Konemiehentie 2, FI-00076, Finland

3- AAAdvisors-QCG (ABN AMRO) and LEO/CNRS
Rue de Blois, F-45067 Orléans Cedex 2 - France

Abstract. The financial market dynamics can be characterized by macro-economic, micro-financial and market risk indicators, used as leading indicators by market professionals. In this article, we propose a method to identify market states integrating two classification algorithms: a Robust Kohonen Self-Organising Maps one and a CART one. After studying the market's states separation using the former, we use the latter to characterize the economic conditions over time and to compute the conditional probabilities of related market states.

1 Introduction

Financial markets evolve in a complex dynamics characterized by variations around bullish or bearish trends, punctuated by severe crashes episodes such as in September 2008. Under these conditions, identifying the market's state is crucial to characterize the level of risk in the system. Maillet and Michel in [1] provide a first characterization of financial crises with an indicator of market shocks (IMS). In this article, we now propose to generalize their approach by studying a series of macro- and micro-financial indicators using an unsupervised neural classification algorithm: Kohonen Self-Organizing Maps (SOM) in its robust version [2] and [3], known as R-SOM, in order to reduce the variability of the neighbourhood structure. The former algorithm will allow us to identify three different market states. Thereafter, we use a supervised statistical learning method proposed by Breiman et al. [4] called Classification and Regression Trees (CART), using a bootstrap aggregating technique (bagging). The former algorithm provides us with the tool to determine the relevant and stable variables for the economic analysis, and tactical indicators which best characterize the market conditions and regimes [5], and also to compute related conditional probabilities of future market states. The aim of this paper is to search for a method for identifying the financial markets' states and their conditional probabilities. The basic idea is to combine two classification algorithms. On one hand, the SOM algorithm will provide us class labels upon which the CART algorithm will learn, and on the other hand, it will provide information about the relevant financial indicators and the future conditional probabilities of market states.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

2 Robust Self Organizing Maps Principles

The Self-Organizing Maps (SOM) is a neural classification technique introduced by Teuvo Kohonen in 1982 [6]. The SOM is an unsupervised learning algorithm used to map a real space, to study the distribution of data in a large dimensional space. The stochastic nature of the algorithm does not guarantee a systematic convergence. In this paper, we use a robust version using a method proposed by Guinot et al. [2]. This approach provides a two-step stochastic method based on a bootstrap process to increase the reliability of the underlying neighbourhood structure. The increase in robustness is relative to the sensitivities of the output to the sampling method and to some of the learning options (initialisation and order of data presentation). At the first step, a bootstrap process is used to build a table (noted P) of probability for any pair of individuals to be neighbours. At the second step, we choose between several maps the one which exhibits the greatest similarity with this table. For each resulting map M_i, we construct a table P_{M_i} (whose elements values are 1 if two individuals are neighbours and 0 otherwise). The robust map selected (denoted below RMap) is that which minimizes the distance between the two neighbouring structures:

$$\text{RMap} = \arg\min_{M_i, i\in I} \{\|P - P_{M_i}\|_{\text{Frob}}\} \qquad (1)$$

where P is the table containing the empirical probability of two individuals to be in the same group, P_{M_i} is the table associated with the map M_i, $I = [1, \ldots, N_{\text{boot}}]$ is the set of maps generated with N_{boot} bootstrapped samples and $|| \cdot ||_{\text{Frob}}$ is the Frobenius norm. Finally, the RMap gives a summary of the data and a neighbourhood structure between classes that is less sensitive to the sampling (due to the first step treatment), to the initialisation and the order of the data presentation (thanks to the second step treatment). We present hereafter the classification on the economics indicators. The study period runs from 26 February 2003 to 20 June 2012. The observation frequency is monthly returns.

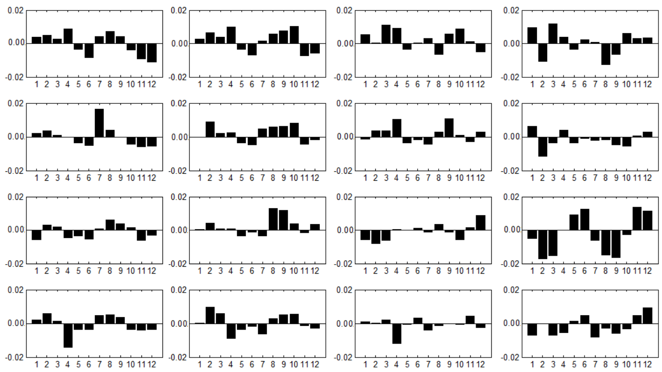

Fig. 1: Kohonen robust map of the tactical indicators series.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Figure 1 represents the 4×4 map classification with sixteen neurons. The numbers (in order, from 1 to 12) on the X-axis stand for: the lagged returns of the MSCI Index, the up/down ratio (UD), trend (BB), risk appetite (RA), market shocks (IMS), volatility (VIX), liquidity (GNREPO), economic sentiment (ESI), firms' valuation (VALO), economic news (NewsG10), alphas risk (SAR) and correlation of pairs in the SP500 (CORRsp). We find that there are significant differences between different classes (dates). Some are associated with low values (classes 3, 9, 10 and 15) and others to higher values (classes 4, 8 and 14).

3 Identifying Financial Markets' States

The purpose of the classification by robust Kohonen maps is to identify, through a series of financial indicators, different market states by observing cumulative returns of the MSCI World index associated with each class. We choose to group classes associated with cumulative returns that seem to have the same behaviour. We construct three sets of cumulative returns conditional on each state of the market.

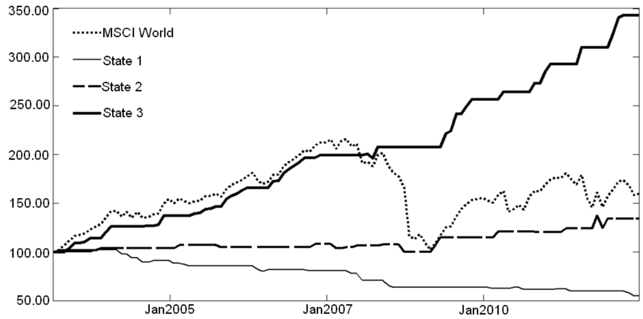

Fig. 2: Series of conditional cumulative returns of the MSCI World.

Figure 2 shows that the RSOM algorithm is able to distinguish the different market's states because after grouping classes, the third state leads to an increase of 348.11% of the MSCI World over the entire period (against 60.08% for the latter). The first state leads in turn to a decrease of 44.9% in the index. Moreover, we observe that the map obtained after classification is divided into two distinct parts, each corresponding to a different market condition: the top of the map (Class 4, 8, 12 and 16) corresponds to a bull market while the bottom (classes 1, 5, 9 and 13) is associated with bear markets systematically. This partitioning suggests that the classification by SOM identified code vectors, typical values of macroeconomic indicators correlated with financial market conditions.

4 CART Algorithm and Market Conditions

Using the above results needs to be able to calculate the probability of occurrence of future market states. We then use a supervised statistical learning method:

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Classification and Regression Trees (CART) developed by Breiman et al. [4]. Thanks to its robust version using the bootstrap technique (called Bootstrap Aggregating), we can determine the relevant variable for the economic analysis over time, i.e. which characterizes best the market conditions, but also we can calculate the conditional probabilities of future states.

4.1 CART Principles

Classification and Regression Trees [4] is a non-parametric method of building decision trees by binary recursive partitioning in order to distinguish different sets or classes to which the observations belong. The algorithm can be described as follows:

Starting from a parent node t containing a set of observations $\left\{x_j, y_j^q\right\}_{j\in J, q\in Q}$, with $J = \{1, \ldots, N\}$ the set of N vectors of explanatory variables and Q the set of all labels (in this paper, this set corresponds to three possible market states), the partitioning of a parent node t consists in choosing a value α of the explanatory variable x_j, noted x_j^α. Then, the vectors of observations contained in a node t are divided into two child nodes t_1 and t_2 (with fractions p_1 and p_2 respectively) according to the following rule: $x_j \leq x_j^\alpha$. The partitioning is optimal if it minimizes the heterogeneity of the observations contained in the child nodes, following a certain indicator of the degree of homogeneity (the Gini index here in our approach):

$$s^* = \arg\min_{s\in S} \{p_1\{i[t_1(s)]\} + p_2\{i[t_2(s)]\}\} \tag{2}$$

where S is the set of all possible partitioning points, and $i[\cdot]$ the Gini index calculated for a sub-set of observations. This principle is repeated a sufficient number of times to obtain pure terminal nodes, i.e. each vector of observations belong to a unique label.

The CART algorithm has the advantage of not requiring any specification for the functional form of the relationship between dependent variable and explanatory variables. The algorithm allows the use of all types of variables (continuous, discrete and categorical) and does not require pre-selection of variables: it selects itself the most significant variables. However, its main drawback is that it can produce trees with unstable structure: a change in the training set can change the size of the tree or the decision rules, thus leading to different conclusions. In this paper, we use a robust version of the algorithm using the technique of Bootstrap Aggregating (bagging) developed by Breiman [7].

4.2 Characterizing Market Conditions with Tactical Indicators

The first information that can be provided by CART concerns changes over time of the relevant tactical indicators, highlighting the fundamental forces that characterize the evolution of international financial markets. In Figure 3, we show that it's the indicator of growth of repos (GNREPO) which characterizes

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

the period from April 2006 to July 2007 (collapse of the quality of subprime mortgages). Subsequently, it is an indicator of volatility and market stress, the VIX, which characterizes the collapse of the MSCI between March 2008 and March 2009. The decrease of -34% of the index was accompanied by a rise in volatility rather characteristic.

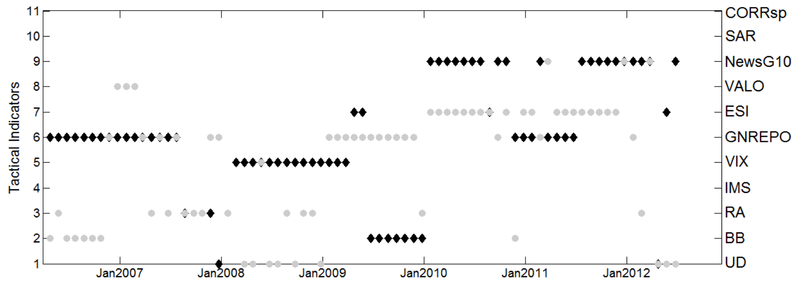

Fig. 3: Evolution in time of the relevant variables.

Between the two periods mentioned above, the ways of diversification in the financial markets began to fail (the CORRsp is the relevant indicator between July 2007 and February 2008). In the following period, we find the results of Gilli and Roko (2006): trend indicators derived from technical analysis are mainly relevant to the following periods of strong upward or downward trend (in fact, during the rebound of the MSCI between June and December 2009 (+15%), the trend indicator Bull/Bear ratio is most relevant). Finally, for the remainder of the sample, the markets seem to react mainly to macroeconomic news (NewsG10 is most relevant). This period, from January 2010 until now covers the sovereign debt crisis in Europe.

4.3 Market's States Conditional Probabilities

The conditional probabilities of market states are out-of-sample forecasts (under prefect information) for one month horizon (one period), obtained by a series of 75 sliding windows with a length of 3 years. For each rolling window, the number of votes for each state reported to 200 (number of bootstrapped samples) gives the value of the conditional probability for the state in question.

Figure 4 shows that in the first part of the sample (March 2006 - November 2007), the MSCI World is in an uptrend and the indicator is constantly in state 3. Between early 2008 and early 2009 (subprimes crisis), the indicator rose to state 2 to state 1 and signalling a major reversal. The rebound experienced by the global index in March 2009 was accompanied by an indicator signalling the state 3. We also note that the period of decline during the year 2011 is always accompanied by an indicator signalling the state 1. Finally, Figure 4 shows that the calculated conditional probabilities reflect a transition mechanism when going from one extreme to another.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

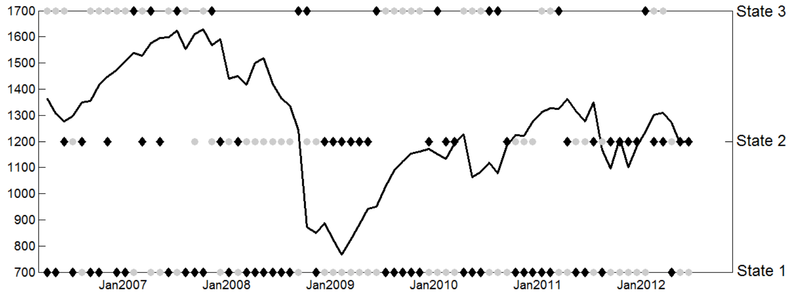

Fig. 4: Conditional probabilities of market's states.

5 Conclusion

We show that the financial markets can be characterized by a series of macro-financial indicators with typical values correlated with the markets' states. The Kohonen map obtained is divided into an upper part (associated to a bullish trend) and a lower part (associated to a bearish trend). We also determined, using the CART algorithm, the most relevant tactical indicators, those that best characterize market conditions: GNREPO, VIX, Bull/Bear ratio, NewsG10 and ESI characterize the different sub-periods of liquidity contraction, crisis (subprimes and the European sovereign debt) and economic recovery on the entire sample. Finally, we constructed an indicator of the conditional probabilities of future states: the predicted states are consistent with the evolution of the MSCI World Index.

References

[1] B. Maillet and T. Michel. An index of market shocks based on multiscale analysis*. *Quantitative Finance*, 3(2):88–97, 2003.

[2] P. Rousset, C. Guinot, and B. Maillet. Understanding and reducing variability of som neighbourhood structure. *Neural Networks*, 19(6-7):838–846, 2006.

[3] A. Sorjamaa, P. Merlin, B. Maillet, and A. Lendasse. SOM+EOF for finding missing values. In M. Verleysen, editor, *ESANN 2007, European Symposium on Artificial Neural Networks*, pages 115–120, Bruges, Belgium, April 25-27 2007. d-side publ. (Evere, Belgium).

[4] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees.* Wadsworth and Brooks, Monterey, CA, 1984.

[5] B. Maillet, M. Olteanu, and J. Rynkiewicz. Non-linear analysis of shocks when financial markets are subject to changes in regime. In M. Verleysen, editor, *ESANN 2004, 12th European Symposium on Artificial Neural Networks*, pages 87–92, Bruges, Belgium, April 28-30 2004. d-side publ. (Evere, Belgium).

[6] T. Kohonen, M. R. Schroeder, and T. S. Huang, editors. *Self-Organizing Maps.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 3rd edition, 2001.

[7] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.