Clustering and Prediction of Risk Spread Curves

Ventsislav Nikolov

Abstract: *In the paper a combination of clustering, detection of risk spread curves in the clusters and prediction is presented. The risk curves are identified and analyzed by two different types of neural networks using supervised and unsupervised learning. Such an analysis allows estimation of future effects about issuers with currently increasing risk in the market.*

Key words: *Clustering, Prediction, Time Series, Spread Curves*

INTRODUCTION

Given is a set of financial spread curves from credit default swap (CDS) markets with their maturities and historical values which can be considered as two-dimensional data vectors. The objectives are:

- organizing them into clusters in order to reduce their number;
- identification of the riskiest curve which tends to deviate from the cluster;
- prediction of the future development of this curve;

First the spread curves with similar historic behaviour are grouped into clusters and synthetic representative curve for every cluster is generated. The main goal of this subtask is to reduce a large set (several thousands) of curves to a small set (several dozens) of synthetic curves. Second, the risk analysis is performed. The curves which belong to a given cluster but in some moment, begin to increase should be detected. Thus, the curves that represent issuers with increasing risk can be analyzed more closely. Next a prediction of the future development of the risk curves is performed by analyzing their historical behaviour. Finally, estimation could be made for some future moment in order the effects of the development of detected risk curves to be understood if preventive measures are not applied.

RISK CURVES IDENTIFICATION AND PREDICTION

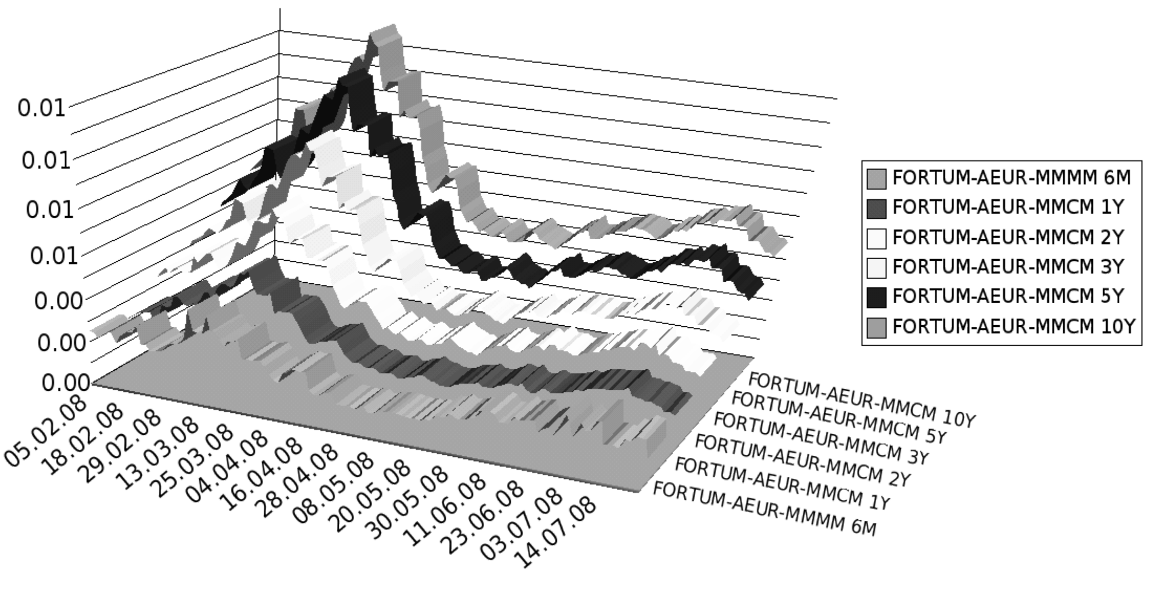

Fig.1 A spread curve with its historical dates and maturities

As some researchers have shown, clustering is a useful pre-processing operation in many efficient prediction algorithms [1]. Clustering allows analysis to be focused on a particular local data area in which a particular task should be performed. In addition to this clustering is also useful in many other tasks like defining benchmarks for time-consuming calculations. In our case the initial data are defined by historical development of spread curves with several maturities along specified working days which can be interpreted as two-dimensional series. This historical development gives useful information about the curves behaviour from which some patterns could be obtained for the sake of prediction.

Since recognizing and defining of curve external factors dependencies is complex and non-trivial it is not considered.

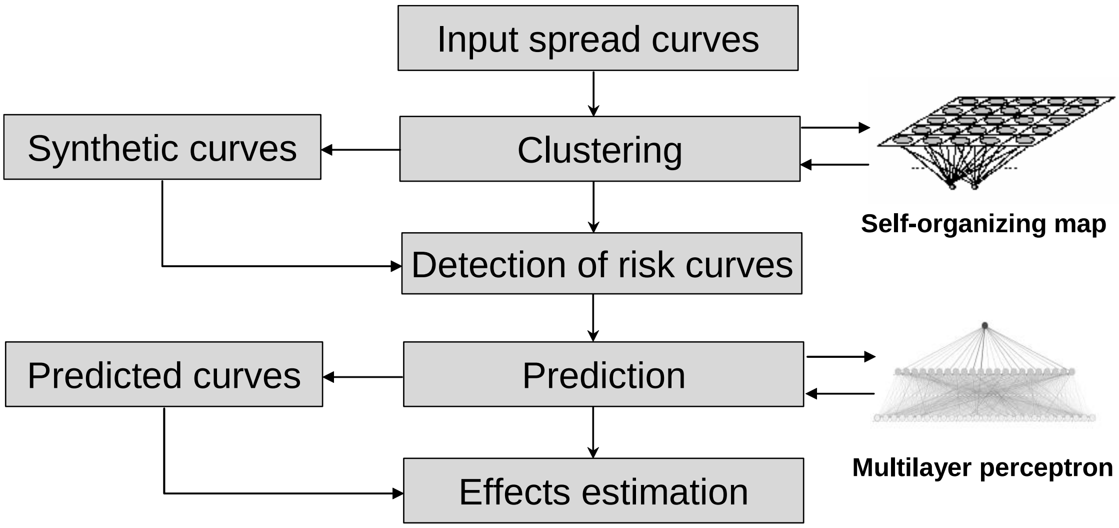

Fig.2 Stages of analysis

Having a big number of curves, the first task is to separate them into small number of clusters in such a way so the most similar curves to be combined and associated to a generated synthetic curve which represents the cluster and has the most important characteristics of the curves in it. Generally, in one cluster are curves whose Euclidean distance of all historical values to all historical values of the synthetic cluster curve is smallest. Clustering is done by means of self-organizing map which is based on the ordering of input data in an optimal way according to the Euclidean distance [4]. Each output unit of the grid represents a cluster. The synthetic representative curve for a given cluster is the prototype vector of its output unit after network training. Thus, it is possible to control the number of clusters by implementing self-organizing map with given size of the output grid. In fact, this is not always true because if the number of input curves is near to or bigger than the number of output units then in the output layer can exist some so called dead neurons. They are such neurons that have not mapped input data and thus they are not used as clusters. In the practical situations this case is naturally avoided, because the main task of clustering is to reduce a very large set of curves into as smaller set as possible with preserving many of their characteristics.

Having initial data organized into clusters, a close analysis of the curves can be done. To make this their historical behaviour is considered. If some curve is moving along or near to all other curves in the cluster then it must have mapped its main characteristic into the synthetic cluster curve behaviour. However, if one of the curves becomes to deviate from all other cluster curves then it must deviate also from the synthetic curve. If the number of the curves is big and the number of the clusters is small (which is the real situation in the most cases) then the number of input curves in every cluster is also big.

Thus some of the curves could be in a new cluster, but because of the small specified number of clusters they are mapped to some of the available clusters. This is to say that some of the curves in a given cluster would tend to deviate from the synthetic cluster curves. Such situation is shown on fig. 4.

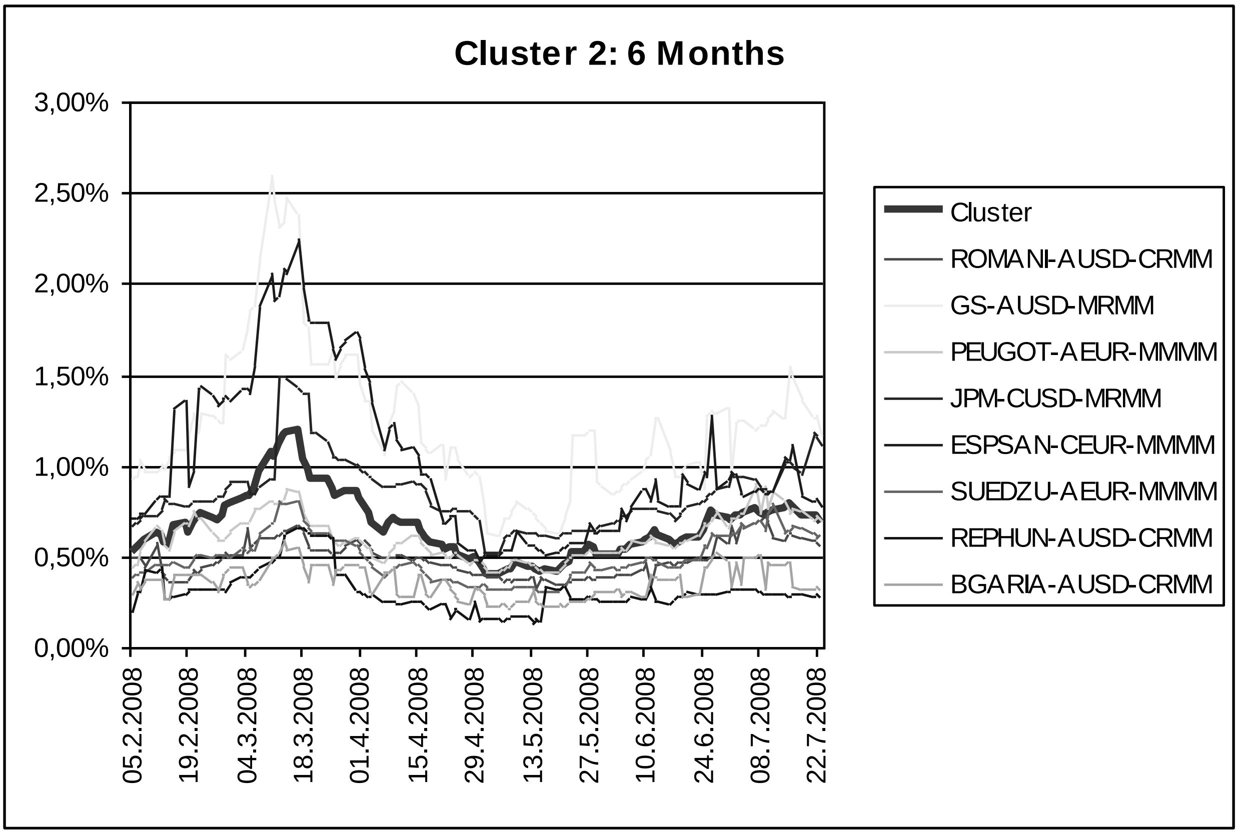

Fig.3 Historical development for a cluster

Because of spread curves meaning, the dynamic of their values can be used for risk measurement. Their increasing values mean increasing risk. A curve with increasing values represents an issuer in the financial domain which is becoming risky. The task of the risk analysis stage is to detect the curve with such behaviour. To do this the generated synthetic curve is considered as a benchmark and new (pseudo-synthetic) curves are generated for all real curves as differences between real and synthetic curve of the cluster they belong to. Then the slope information of the pseudo-synthetic curves is analyzed in order to be found the curve which most quickly deviates from its cluster synthetic curve. If the pseudo-synthetic curve is sloped up then apparently it tends to deviate from the cluster synthetic curve. A prediction of such a curve should be done in order the effect to be estimated if preventive measures are not taken.

The prediction is made as a univariate time series prediction. Each maturity of every curve is considered as a time series and it is predicted without giving an account of external factors.

There are different well-known prediction methods. For example, for non-stationary time series can be used autoregressive method, Holt-Winter's exponential smoothing, Box-Jenkins ARMA and ARIMA methods, Kalman filtering, etc. Under standard statistical conditions there is a slight difference in prediction accuracy between the most of these methods and neural network prediction, but some authors report that even simple neural network can outperform alternative traditional prediction algorithms when the complexity of time series varies and the neural network is optimally tuned [6].

Prediction in the current solution is performed by a three-layer perceptron with back propagation training algorithm. Its inputs get lagged values (each unit gets a value measured on a date) and its outputs produce the result of the prediction. The training is made using well-known sliding window technique (fig. 5) [2].

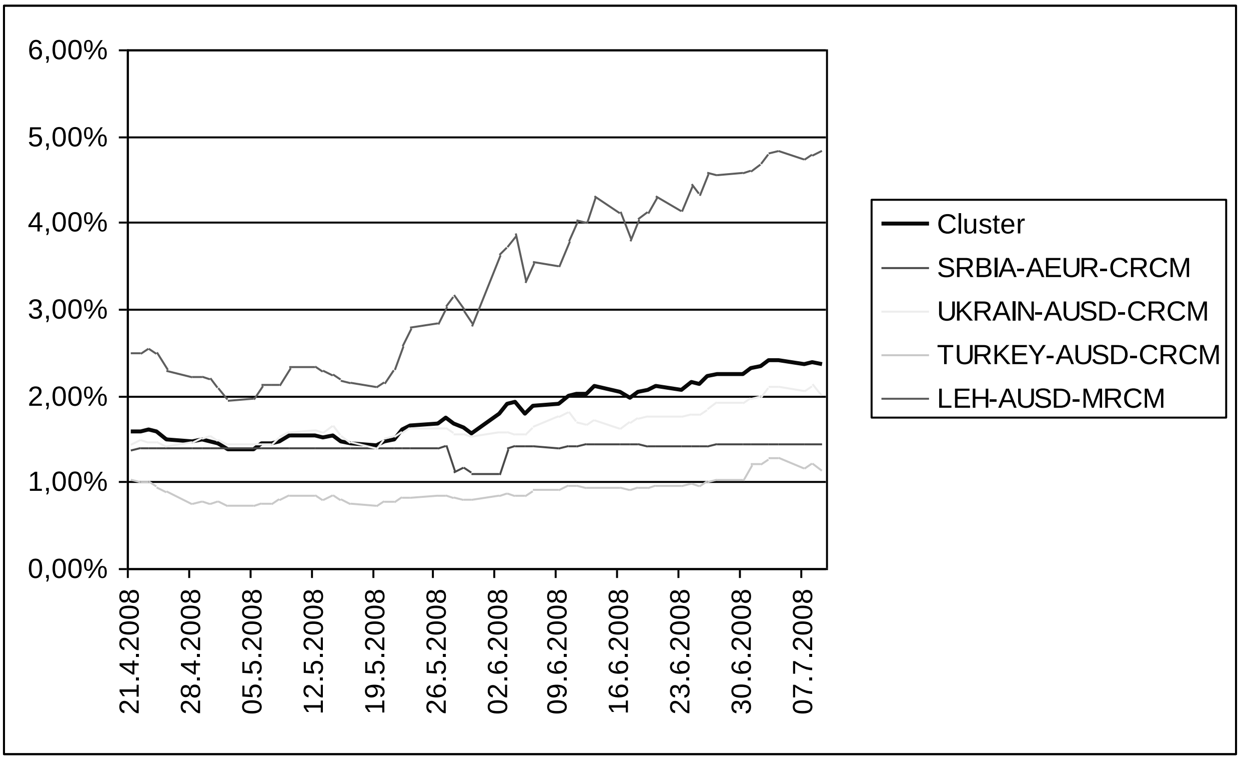

Fig.4 A maturity of a risky curve: the most upper LEH-AUSD-MRCM

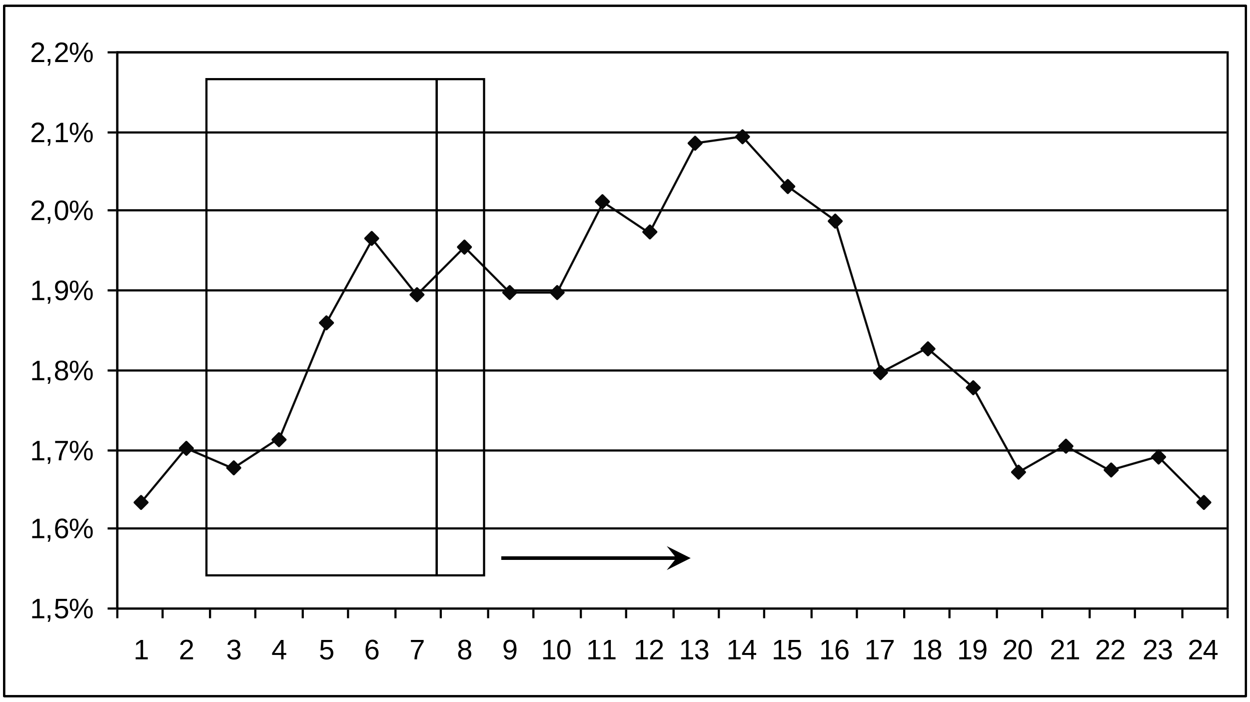

Fig.5 Sliding window used for neural network training

The future value is predicted as a function of the historical values [3].

$$y(t+1) = f(y(t), y(t-1), \dots, y(t-p)) + \varepsilon \quad (1)$$

The size of the window p is calculated based on analysis of autocorrelation function for the series. The size of the window also defines the size of the input layer. On fig.5 a sliding window with length p=5 elements is shown. Next element is used as a target vector.

To find the best suited parameters for analysis the neural network is implemented with three activation functions, possibility of using of known non-random weights initialization procedures, using of adaptive learning rate, different techniques of normalization, trend removing and visual plot of initial time-series correlogram. Moreover, there are convenient methods for cross validation procedures to construct network with effective architecture and after the training it is possible validation in different ranges of the available series to be done. The network parameters determined for maturities 6 months and 1 year are shown on table 1.

Table. 1 Parameters for the first two maturities

Neural network parameters		
Maturity	6M	1Y
Input neurons	30	55
Hidden neurons	22	41
Output neurons	1	1
Learning rate	0,29	0,3
Momentum factor	0	0
Flat spot elimination term	0	0
Epochs	30000	10000
Adaptive learning rate	No	No
Nguyen-Widrow initialization	Yes	Yes
Activation function	Bipolar sigmoid	Bipolar sigmoid
Normalization interval	-0.9 – 0.9	-0.9 – 0.9
Trend removing	First difference	First difference
Window size	22	55
Shuffle in every epoch	Yes	Yes

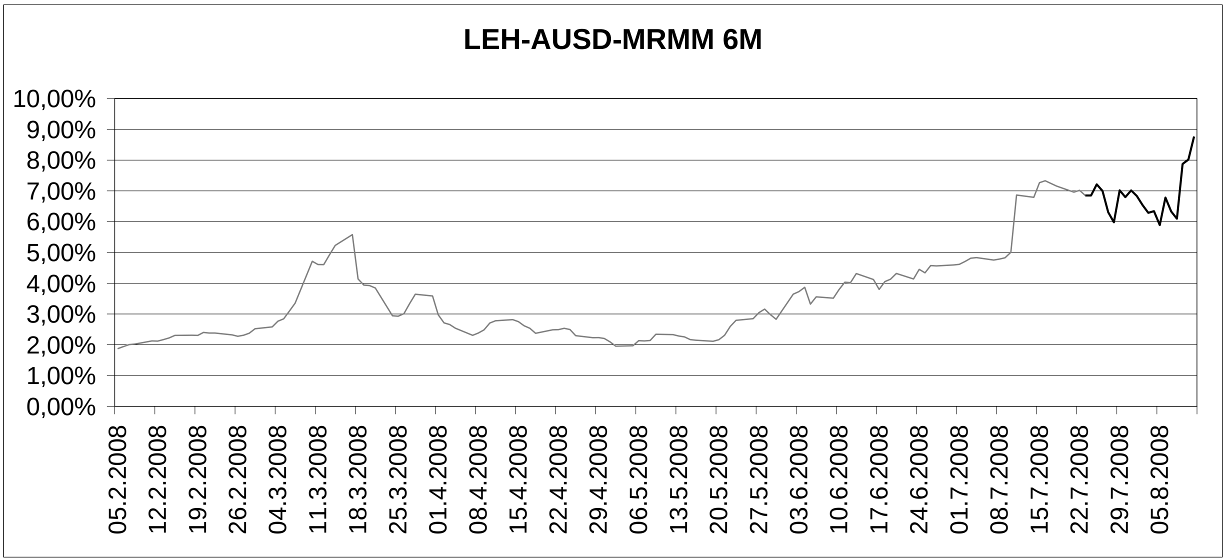

Fig.6 Prediction of detected risk curve for maturity 6 months

When the neural network is trained the prediction is just the calculated output:

$$o_k = f\left(\sum_{j=1}^{B}\left(w_{kj} f\left(\sum_{i=1}^{A}(w_{ji}Y_i + b_i)\right)\right) + b_j\right) \quad (2)$$

where:

o_k is calculated value of an output unit
f is activation function
A is the number of input units that is equal to the size of window in the training stage
B is the number of hidden units
w is weight value
Y_i is time series value
b is bias

Multi-step predictions are made recursively – every predicted value is considered as a real one when the next prediction is estimated, and the procedure continues until the specified number of prediction values is reached [3] [5]. The prediction for the maturity of 6 months is shown on fig. 6 without confidence levels.

On table 2 the current and predicted values for 19 days ahead are shown along with the differences between them.

Table. 2 Actual values, predictions after 19 days and differences

Years	0,5	1	2	3	5	10	
11.08.08	8,6193%	8,7635%	6,2199%	5,2552%	4,3984%	3,2659%	Predicted
23.07.08	6,7011%	6,8482%	5,4481%	4,5185%	3,2897%	2,6982%	Current
Spread	1,9182%	1,9153%	0,7718%	0,7367%	1,1087%	0,5677%	Diffrence

CONCLUSIONS AND FUTURE WORK

In the experimental results is shown that clustering and prediction can be used successfully together in order to identify risky two-dimensional data curve and to estimate its future development.

Because the prediction cannot be always the best, the future work is to use hybrid systems or integration of methods for increasing the performance of prediction.

REFERENCES

[1] Barreto, G. Time Series Prediction with the Self-Organizing Map: A Review. Center of Technology Campus of Pici, Fortaleza, Brazil.

[2] Crone, S. Business Forecasting with Artificial Neural Networks. IBF Tutorial, Institute of Business Forecasting, Boston, 2004.

[3] Hamilton, J. Time Series Analysis. Princeton University Press, 1994.

[4] Kohonen, T. Self-Organizing Maps. Springer, 2001.

[5] McNelis, P. Neural Networks in Finance: Gaining Predictive Edge in the Market. Elsevier Academic Press, 2005.

[6] Palit, A., D. Popovic. Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications. Springer, 2005.

ABOUT THE AUTHOR

Ventsislav Nikolov
Senior Software Developer at Eurorisk Systems Ltd.
31, General Kiselov Str., 9002 Varna, Bulgaria
E-mail: vnikolov at euroriskystems dot com