A Fuzzy Model of the MSCI EURO Index Based on Content Analysis of European Central Bank Statements

Viorel Milea, *Graduate Student Member, IEEE*, Rui J. Almeida, *Graduate Student Member, IEEE*,
Uzay Kaymak, *Member, IEEE* and Flavius Frasincar

Abstract— In this paper we investigate whether the MSCI EURO index can be predicted based on the content of European Central Bank (ECB) statements. We propose a new model to retrieve information from free text and transform it into a quantitative output. For this purpose, we first identify all adjectives in an ECB statement by using the Stanford Part-of-Speech Tagger and feed these to the General Inquirer (GI) content analysis tool. From GI we obtain a matrix that provides for each document and for each content category the percentage of words in the document that fall under each category. After normalizing the data, we develop a Takagi-Sugeno (TS) fuzzy model using fuzzy c-means clustering. The TS fuzzy system is used to model the levels of the MSCI EURO index. To determine the performance of the model, we focus on the accuracy of predicting upward or downward movement in the index, and obtain, on average, an accuracy of 66%, that corresponds to an improvement of 16% over a random classifier.

I. INTRODUCTION

The Internet provides access to more and more sources of textual information. From this context, an increasing need emerges of being able to deal with these large volumes of untagged content. In an era where information can be directly translated to power, more and more efforts to extract and represent meaning from free textual representations are being undertaken in a variety of contexts. One such context is the economic one, where increased attention is devoted to the automated processing of news messages with the goal of predicting stock prices. The difficult task of extracting economic meaning from free text in the form of news messages can focus on different aspects, such as, for example, the extraction of sentiment or annotation of events. Additionally, the focus could consists of word categories and the frequencies of words across these categories. For the latter purpose, content-analysis tools such as the General Inquirer (GI) [1] are available.

In this paper, we use the monthly statements of the European Central Bank (ECB) [2] for the prediction of movements of the MSCI EURO index [12]. For this purpose, we employ a fuzzy model defined on a subset of the different content categories supported by GI and train the model for predicting whether, on the month of the ECB statement, the index will move up or down relative to the previous month.

The motivation for this study is two-fold. Our main contribution is to show how information can be retrieved from free

Viorel Milea, Rui J. Almeida, Uzay Kaymak and Flavius Frasincar are with the Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, the Netherlands email: {milea, almeida, kaymak, frasincar}@ese.eur.nl. U. Kaymak is also with the Eindhoven University of Technology, the Netherlands.

text and transformed into some quantitative output. Secondly, we aim at designing a more general approach that will enable prediction of financial variables, such as market prices, based on the content of textual information, such as financial news. Devising such approaches can prove very useful in assisting financial traders with the investment decisions they have to make.

Each month, the ECB publishes a statement for the public, titled 'Introductory Statement with Q&A' [2]. In this statement the levels of key interest rates under the control of the ECB are mentioned, as well as an overview of different aspects of the economy. This overview encompasses the economic and monetary analysis, as well as the fiscal policies and structural reforms. This statement has been published each month, at various dates, since June 9th, 1998. Rather than only focussing on aspects from the past, the ECB statements provide projections into the future of the expected states of the different indicators discussed in the statements. For this reason, it is realistic to assume that such statements have an impact on the expectation formation of the different participants in the economy, such as the financial traders.

The MSCI EURO index has been brought to life in 1998, and has since provided a good overview of the European equity markets. Initially intended as basis for derivative contracts, the index can also be employed as an indicator of the European economies it represents. The index covers a set of 38 industry groups, such as Energy, Finance and Consumer Goods. Based on liquidity, a large number of European companies are incorporated in the index, such as, among the largest, Glaxo Wellcome and Royal Dutch Petroleum Co. In our research, the index is employed as a barometer for the European economy and as a benchmark for the fuzzy model we employ for assessing the impact of ECB statements on the economy.

The General Inquirer is a content analysis tool based on the Laswell Value Dictionary [3], [4] and Harvard-IV Dictionary [5], [6]. It provides over 300 word categories, among which the categories Positive, Negative, Weak, and Strong. A word may find itself indexed under one or more categories. Based on where the individual words in a document are indexed, General Inquirer is able to generate, amongst others, a document fingerprint consisting, for each content category, of the percentage of words in the document that fall under that category.

The research described in this paper is aimed at employing document fingerprints generated with GI for individual ECB statements for the prediction of future movement directions

in the MSCI EURO index. For this purpose, we employ a Takagi-Sugeno fuzzy model [7] that takes as input the percentage of words that fall under each of the selected GI content categories. The output of this model is a predicted level of the MSCI EURO index, which we transform into an upward or downward movement, thus making the current problem a classification one.

The paper is organized as follows. In Section II we provide an overview of work related to our current endeavor. In Section III we provide an overview of the data used for the current purpose and in Section IV we introduce the proposed model. In Section V we describe the experiments performed and present the results of these experiments. Finally, we conclude in Section VI and provide some directions for further research.

II. CONTENT AND SENTIMENT ANALYSIS

The first attempt at content analysis in an economic context is presented in [8]. Here, the authors investigate the relationship between focus on wealth and wealth-related words in the speeches of the German Emperor and the state of the economy over the period 1870-1914. They find a strong relationship between the focus on wealth and the state of the German economy. More recent research, such as [9], relies on the GI dictionary for explaining market prices and trading volumes. The author finds that a relationship exists between a daily Wall Street Journal column, 'Abreast of the Market', and the market prices and trading volumes of that day for the stocks discussed in the column. In [14] the authors develop a method for the automated extraction of basis expressions that indicate economic trends. They are able to classify positive and negative expressions which hold predictive power over economic trends, without the help of a dictionary.

Other research has focussed on the extraction of sentiment from free text in an economic context. In [10] the authors focus on eight dimensions of sentiment: Joy, Sadness, Trust, Disgust, Fear, Anger, Surprise and Anticipation, and are able to provide visualizations of how these eight sentiments evolve over time on some concept, e.g., Iraq, based on news messages. The results are validated against ratings of human reviewers of the news messages, and the method performs satisfactorily in visualizing the evolution of these sentiments over time regarding the studied concepts.

Staying in the realm of sentiment mining, we signal an approach in [11] related to the extraction of term subjectivity and orientation from text. The approach starts with two training sets consisting of Positive and Negative words, respectively. It extends these two sets with WordNet synonyms and antonyms of the words found in the sets. Next, a binary classifier is generated by a supervised learner that is able to categorize vectorized representations of terms and classify them as Positive or Negative. Extraction of fuzzy sentiment is done in [13], where the authors are able to assign a fuzzy membership of Positive or Negative to a set of words using the Sentiment Tag Extraction Program (STEP).

The approach we present in this paper differs from the above approaches in that it relies on selected content categories from GI, and employs a fuzzy model for the prediction of movements in the MSCI EURO index. Rather than focussing on sentiment, we select a total of thirteen categories from GI and employ the percentages of words that fall under those categories as document fingerprints for the individual ECB statements. By using a fuzzy model, we are able to investigate how each category impacts the index, and draw economic conclusions from here. This is different from the approach that stands closest to the research outlined in this paper, namely [9], as it does not aggregate all content categories into one single indicator, thus losing the ability to question the impact of the different content categories on the explanandum.

III. DATA PROCESSING

In this section we provide an overview of the data we employ for our goal and the model we design. In the first part, an overview of the used data is provided. The second part focusses on steps that are needed to prepare the data from which fuzzy model can be developed.

In our approach, we require data from two different sources. On the one hand, we employ ECB statements available from the ECB press website [2]. Additionally, we employ the MSCI EURO index, available from the Thomson One Banker website [15]. As the index is available starting from 31 December 1998, we select both the statements and the index values for the period 1 January 1999 to 31 December 2009.

An ECB statement as employed for our current purpose consists of different parts. The first part deals with the key ECB interest rates and their levels for the coming months. The following four parts deal with the economic and monetary analysis, as well as the fiscal policies and structural reforms. These first five parts are considered relevant for the question at hand. Finally, approximately the second half of an ECB statement consists of questions and answers from the press directed towards the president of the ECB. For the current scope, we consider the Q&A part of an ECB statement relevant only indirectly, and only focus on the part describing the current and expected future state of the economy.

The relevant parts of the ECB statements for the selected period are extracted by using an HTML wrapper from the ECB press website. Upon successful extraction, each statement is annotated for parts of speech using the Stanford POS Tagger [16], [17]. Based on the part of speech annotation, we extract only the adjectives from the text. It should be noted that all ECB statements, at least in the part we consider relevant for the current purpose, follow a similar structure. For this reason, we believe that the adjectives in the text could provide a good discrimination among the different statements.

Upon creating, for each ECB statement from the relevant period, the set of all adjectives contained in the text, the data are fed to the General Inquirer web service. Based on this

input, GI is able to generate a document fingerprint consisting of the percentages of words from the document that fall under each category supported by GI. GI currently supports over 300 content categories, but for our current purpose we only focus on 13 of them, namely [1]:

- *Positiv*, consisting of 1045 positive words,
- *Negativ*, made up of 1160 negative words,
- *Strong*, consisting of 1902 words implying strength,
- *Weak*, containing 755 words implying weakness,
- *Ovrst*, consisting of 696 words indicating overstatement,
- *Undrst*, containing 319 referring to understatement,
- *Need*, made up of 76 words related to the expression of need or intent,
- *Goal*, consisting of 53 words referring to end-states towards which muscular or mental striving is directed,
- *Try*, containing 70 words indicating activities taken to reach a goal,
- *Means*, made up of 244 words denoting what is utilized in attaining goals,
- *Persist*, 64 words indicating endurance,
- *Complet*, consisting of 81 words indicating the achievement of goals,
- *Fail*, which consists of 137 words that indicate that goals have not been achieved.

By feeding the adjectives from each relevant ECB statement to GI, we obtain a matrix of percentages that indicate for each document, for each content category, the percentage of words in that document that fall under that category. Upon generating this matrix, we normalize using min-max normalization across each content category. This resumes to applying Equation 1 individually to each datapoint, for each variable in the dataset (note that the *min* and *max* operations are applied over all values of each variable).

$$x_i = \frac{x_i - min(\mathbf{x})}{max(\mathbf{x}) - min(\mathbf{x})}, \tag{1}$$

where x_i is the i-th datapoint for variable $\mathbf{x}$.

Finally, we obtain the data on the MSCI EURO index from Thomson One Banker (T1B). We extract monthly, end-of-month data for the period January 1st 1999 until December 31st 2009. An overview of the all the data processing steps is provided in Figure 1.
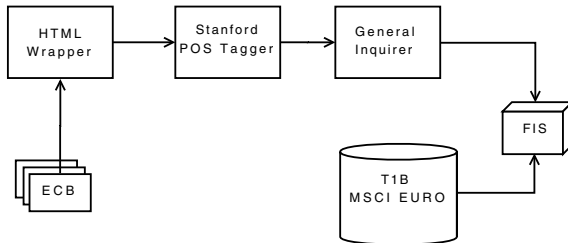

Fig. 1. Data processing steps.

IV. Fuzzy Model

In this section we outline the basics of the adopted fuzzy model for the prediction of the MSCI EURO index based on the content of ECB statements.

Several techniques can be used in fuzzy identification. One possibility is to use identification by product-space clustering to approximate a nonlinear problem by decomposing it into several [19], [20] subproblems. The information regarding the distribution of data can be captured by the fuzzy clusters, which can be used to identify relations between various variables regarding the modeled system.

Takagi and Sugeno (TS) [7] fuzzy models are suitable for identification of nonlinear systems and regression models. In this work, we address the prediction of the MSCI EURO index as a regression model. A TS model with affine linear consequents can be interpreted in terms of changes of the model parameters with respect to the antecedent variables as well as in terms of local linear models of the system. An affine TS model has the following structure:

$$R^k : \text{If } \mathbf{x} \text{ is } A^k \text{ then } y^k = (\mathbf{a}^k)^T \mathbf{x} + b^k, \tag{2}$$

where R^k is the k-th rule in the model rule base, $\mathbf{x} = [x_1 \ldots, x_n]^T$ is the antecedent variable and $A^k = A_1^k, \ldots, A_n^k$ are the fuzzy sets corresponding to the antecedent variables. The rule consequent y^k is an affine combination of $\mathbf{a}^k$, a parameter vector, and b^k, a scalar offset. The consequents of the affine TS model are hyperplanes in the product space of the inputs and the output.

To form the fuzzy system model from the data set with N data samples, given by the regressor $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^T$ and the regressand $Y = [y_1, y_2, \ldots, y_N]^T$ where each data sample has a dimension of n ($N >> n$), the structure is first determined and afterwards the parameters of the structure are identified. The number of rules characterizes the structure of a fuzzy system. In this study the number of rules is the same as the number of clusters. Fuzzy clustering in the Cartesian product-space $X \times Y$ is applied to partition the training data into K clusters. The partitions correspond to the characteristic regions where the systems behavior is approximated by local linear models in the multidimensional space.

In this work, we use the fuzzy c-means (FCM) [18] algorithm. As result of the clustering process, we obtain a fuzzy partition matrix $U = [\mu_i^k]$. The fuzzy sets in the antecedent of the rules are identified by means of the matrix U that have dimensions $[N \times K]$. One dimensional fuzzy sets A_j^k are obtained from the multidimensional fuzzy sets by projections onto the space of the input variables x_i. This is expressed by the point-wise projection operator of the form

$$\mu_{A_j^k}(x_i) = \text{proj}_j(\mu_i^k), \tag{3}$$

after which the pointwise projections are approximated by Gaussian membership functions.

When computing the degree of fulfillment $\beta_k(x)$ of the k-th rule, the original cluster in the antecedent product space

is reconstructed by applying the intersection operator in the cartesian product space of the antecedent variables:

$$\beta_k(x) = \mu_{A_1^k}(x_1) \wedge \mu_{A_2^k}(x_2) \wedge \ldots \wedge \mu_{A_p^k}(x_p). \quad (4)$$

Other t-norms, such as the product, can be used instead of the minimum operator. The consequent parameters for each rule are obtained by means of linear least square estimation, which concludes the identification of the classification system. After the generation of the fuzzy system, rule base simplification and model reduction could be used [21], but we did not consider this step in our current study.

The focus of the current research is the prediction upwards of downwards movements in the MSCI EURO index. This movement is specified in the output data in the following way. With the exception of the first observation from the dataset, all output values are set to 1 if the predicted value for the index in period $t+1$ is higher than or equal to the predicted value of the index in period t, and to 0 if the predicted value of the index is lower in period $t+1$ compared to the same value in period t. The same procedure is applied to the actual values of the index.

V. EXPERIMENTS AND RESULTS

This section provides an overview of the experiments and the results we obtained by running our analysis on the collected dataset. The first part of the section describes the experimental setup, while the second part focusses on the obtained results.

For training of the model we employ 70% of the data, and leave the remaining 30% for testing purposes. For the training dataset, we generate a random permutation of indexes of data points covering 70% of the complete dataset. In this way, every run of the system will be using different, randomly selected data. We do this in order to test the accuracy of the system regardless of economic cycles, as training the system on the first 70% of the data cannot account for the economic crisis from 2008 onwards. By using multiple runs on randomly selected datapoints we aim at reducing this effect.
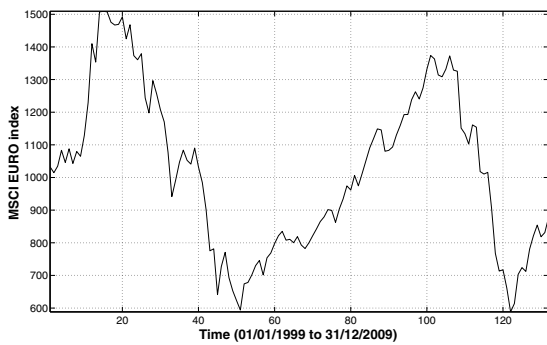

Fig. 2. The MSCI EURO index.

The training data are also randomized for a different reason. If the choice would have been made to chronologically

divide the dataset into two parts, from which the 70% training part would represent the first 70% of the data, we would face the risk of implicitly using the trend in the index as input, which we consider undesirable, as we only want to select the different content categories as determinants for movements in the index. Additionally, due to the evolution of the index, by choosing a chronological division of the data, it can be argued that the training set would not contain enough variation as illustrated in Figure 2.

Based on the randomized data, we run 100 experiments, and for each experiment the data are randomly drawn again from the dataset. For all 100 experiments, we maintain 70% of the dataset for training and 30% of the dataset for testing. Although different types of fuzzy systems have been tested, the best results have been obtained with a Takagi-Sugeno fuzzy system based on fuzzy c-means clustering [18]. Several initial numbers of clusters have been attempted for the algorithm, but the best results were obtained when 3 clusters have been employed. Before presenting the results, a few more comments on the data are in place.

The training of the system is done on the actual values of the index. Thus, the fuzzy model is initially used to predict the actual level of the MSCI EURO index in the month of the statement that is used as input. However, the focus of the current research, as well as the method of measuring the accuracy of the system, are both based on the prediction of upwards of downwards movements in the index. Thus, rather than focussing on the actual levels of the MSCI EURO index, we aim at predicting whether the index will move up or down in the month of the respective statement. The accuracy of the fuzzy system is measured as the percentage of times that the system is able to correctly predict whether the index will move up or down. The formula for accuracy is presented in Equation 5, where M^+ stands for the number of datapoints correctly predicted as upward movement, M^- stands for the number of datapoints correctly predicted as downward movement, and D stands for the total number of datapoints.

$$ACC = (M^+/D + M^-/D) \times 100\% \quad (5)$$

In Table I we present an overview of the results of 100 experiments on the data described in the previous paragraphs. For the 100 runs, for both the training set as well as the testing set, we provide an overview of the minimum, maximum, mean accuracy obtained and the standard deviation hereof.

TABLE I
RESULTS OF 100 EXPERIMENTS.

	Min (%)	Max (%)	Mean (%)	St. dev.
Training	58.82	77.65	69.18	4.01
Testing	44.44	80.56	63.03	7.88

On the training set, the average accuracy ranges between 58.82% and 77.65%, while reaching a mean of 69.18% with a standard deviation of 4.01. The small standard deviation indicates consistent results, while the average accuracy shows

that in about 2/3 cases, the system is able to correctly identify an increase or decrease in the MSCI EURO index. The average accuracy goes down over the 100 experiments for the test set, but only slightly to 63.03%, indicating that no overfitting occurs. However, the standard deviation nearly doubles to 7.88, and that can also be observed in the much wider range between the minimum and the maximum accuracy. Having a minimum accuracy as low as 44.44% on the test data might indicate that periods are present in the test set when the model does a very poor job at predicting the modification in the index, such as when the movement of the index is solely determined by a crisis period.

In Table II we present the average confusion matrix for 100 fuzzy inference systems that we generate. The results are obtained by averaging 100 confusion matrices, the ones obtained for each of the 100 fuzzy inference systems. The numbers are expressed as percentages. The rows indicate the predicted movement direction of the index, upward or downward change, while the columns indicate the true change in the index value. Thus the first cell indicates the true positives.

TABLE II

CONFUSION MATRIX FOR 100 EXPERIMENTS.

	True Up	True Down
Pred. Up	34.28% (7.83)	16.72% (6.19)
Pred. Down	20.25% (6.31)	28.75% (6.79)

As it can be noticed from Table II, a slight bias can be observed between true positives and true negatives. The system seems to be able to better predict upward movement rather than downward movement. In terms of misclassifications, the same can be stated about the false positives and the false negatives. In Table II we also show the standard deviation for all mean values between parentheses.

In Figure 3 we provide an overview of the fuzzy inference system output surface for selected pairs of inputs. The values of the MSCI EURO index in this figure have been obtained by min-max normalization as described in Section III, and therefore the values for the index range between 0 and 1. From this figure, one can notice the high non-linearity of the relations, such as for example in the case of *Positiv* vs. *Negativ* selected inputs pair. The presence of nonlinear relations supports our choice for a fuzzy inference system to model the relationship between the content of ECB statements and the MSCI EURO index. It can also be observed that small parts of the results are sometimes counterintuitive, as in the case of the *Positiv - Negativ* plot. For example, one could observe that there is a positive correlation between *Negativ* and the MSCI EURO index for very small values of the *Negativ* variable. We consider this a spurious effect, and a direct result of the limited amount of data that is available for training, as well as testing, the model. For higher values of the *Negativ* variable, the relation is as expected: higher values of this variable result in lower values for the index.

An overview of the rule consequents for one of the generated FIS is provided in Table III. It should be noted that the output function takes the form $f(\mathbf{x}) = \mathbf{a^T}\mathbf{x} + b$. For each input n, we present the parameter a_n and the overall parameter b.

TABLE III

FUZZY SYSTEM RULE CONSEQUENTS.

	Rule 1	Rule 2	Rule 3
a_1	-6.37	-2.46	0.89
a_2	-0.64	0.68	0.02
a_3	1.80	0.84	-0.67
a_4	-0.22	4.52	0.25
a_5	-1.17	-2.20	0.79
a_6	-5.23	-4.09	1.47
a_7	5.21	0.32	-0.87
a_8	-3.77	-3.06	0.67
a_9	0.71	-3.02	0.06
a_{10}	-1.93	-1.49	-0.46
a_{11}	-1.34	-9.08	0.75
a_{12}	3.40	-0.12	-0.46
a_{13}	11.29	7.50	-2.92
b	0.59	4.65	0.20

An overview of the membership functions for all 13 inputs is provided in Figure 4. An inspection of the different membership functions for the selected inputs shows that the variable *Try* does not hold any predictive power in discriminating between upward and downward movements in the index. Additionally, it can be stated that the variables *Need, Means, Persist* and *Fail* only hold very small predictive power for the goal pursued in the current paper.

VI. CONCLUSIONS

In this paper we investigate whether the MSCI EURO index can be predicted based on the content of ECB statements. For this purpose, we first identify all adjectives in an ECB statement by using the Stanford POS Tagger and feed this to the GI content analysis tool. From GI we obtain a matrix that, for each document, for each content category, provides the percentage of words in the document that fall under each category. After normalizing the data, we use it to develop a Takagi-Sugeno fuzzy system by using fuzzy c-means clustering. The fuzzy system models the levels of the MSCI EURO index. To determine the performance of the model, we focus on the accuracy of predicting upward or downward movement in the index, and obtain, on average, an accuracy of 66% on the test set, which corresponds to an improvement of 16% over a random classifier. Here, since the classes (increase or decrease) are uniformly distributed, we consider that a random classification will be correct about 50% of the time, and we set the accuracy of a random classifier to this percentage. This accuracy is used as benchmark for our method, and thus measures how much our method improves over a random classifier. Our findings thus indicate that there is sufficient information in the ECB statements, which can be extracted for the prediction of the MSCI EURO index.

Our results lay a solid base for further investigation. One direction would be to focus the efforts on improving the performance of the classifier. This could be achieved by extending the model with additional input variables, such
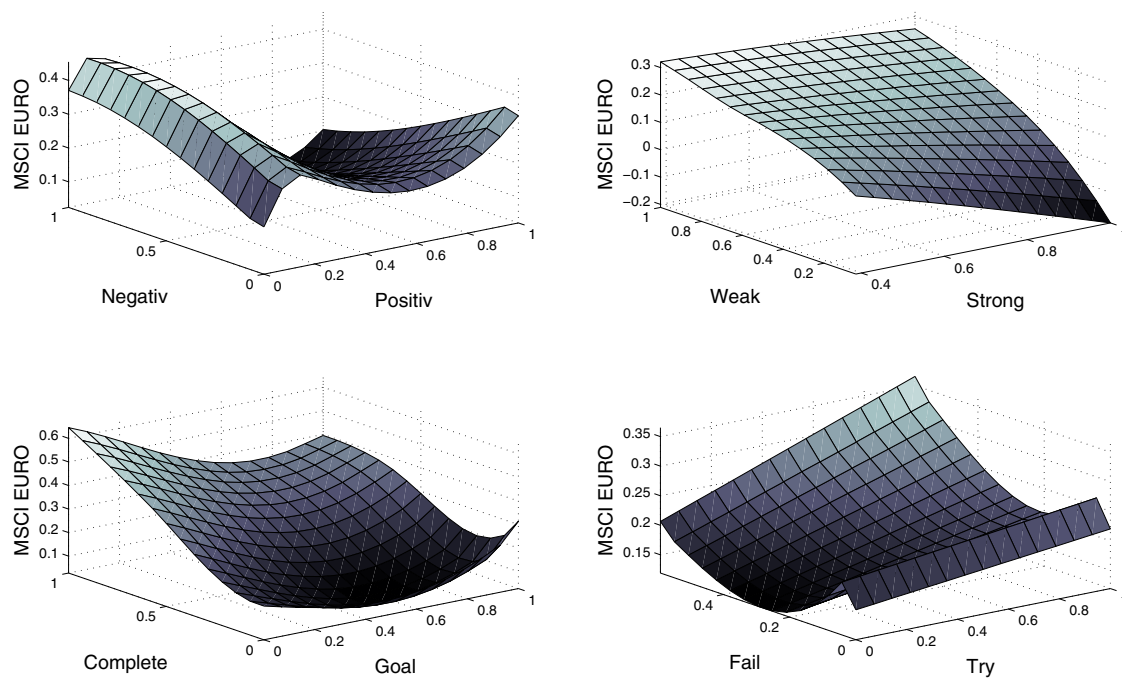
Fig. 3.   Fuzzy inference system output surface for selected pairs of inputs.

as variables denoting the sentiment contained in the ECB statements. More attention can be given to the selection of the GI content categories, which have been currently selected by hand based on the intuition of the authors. A more in-depth analysis of these variables could bring to light new content categories, or aggregation of content categories into single variables, that could further improve the accuracy of the model. Finally, the focus could be directed to extending the set of selected words from ECB statements beyond adjectives, and include other parts-of-speech, such as, for example, adverbs.

Further research could also focus on predicting the actual levels of the the MSCI EURO index, rather than only upward or downward movements. Additionally, attention could be given to different economic indicators other than the MSCI EURO index, such as the level of interest rates, GDP, or other Europe-wide index. Focus could also move outside Europe, thus focussing on press releases of the Bank of England or the Federal Reserve, with the belonging UK or US economic indicators.

Finally, the approach can be extended in such a way that makes it applicable on various other sources of information. In this context, we envision systems that are able to extract information from economic news messages in general and aggregate this information in ways that make it easier for traders to asses the impact that a news message, or a group of news messages, could have on the assets of interest.

REFERENCES

[1] http://www.wjh.harvard.edu/~inquirer/ .
[2] http://www.ecb.int/press/pressconf .
[3] H.D. Laswell and J.Z. Namenwirth, *The Laswell Value Dictionary*, vols. 1-3, New Haven: Yale University Press, 1968.
[4] J.Z. Namenwirth and R.P. Weber, *The Laswell Value Dictionary*, 1974 Pisa Conference on Content Analysis, 1974.
[5] E.F. Kelly and P.J. Stone, *Computer Recongnition of English Word Senses*, Amsterdam: Noord-Holland, 1975.
[6] D.C. Dunphy, C.G. Bullard and E.E.M. Crossing, *Validation of the General Inquirer Harvard IV Dictionary*, 1974 Pisa Conference on Content Analysis, 1974.
[7] T. Takagi and M. Sugeno, *Fuzzy Identification of Systems and its Applications to Modeling and Control*, IEEE Trans. Systems, Man, and Cybernetics, vol. 15, no. 1, pp. 116-132, 1985.
[8] H.D. Klingemann, P.P. Mohler and R.P. Weber, *Das Reichtumsthema in den Thronreden des Kaisers und die Okonomische Entwicklung in Deutschland 1871-1914*, Computerunterstutzte Inhaltsanalyse in der empirischen Sozialforschung, Kronberg: Athenaum, 1982.
[9] P.C. Tetlock, *Giving content to investor sentiment: The role of media in the stock market*, Journal of Finance, forthcoming.
[10] J. Zhang, Y. Kawai, T. Kumamoto and K. Tanaka, *A Novel Visualization Method for Distinction of Web News Sentiment*, 10th International Conference on Web Information Systems Engineering (WISE 2009), pp. 181-194, 2009.
[11] A. Esuli and F. Sebastiani, *Determining term subjectivity and term orientation for opinion mining*, Proceedings the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2006), pp. 193-200, 2006.
[12] http://www.mscibarra.com/products/indices/international_equity_indices/euro/ .
[13] A. Andreevskaia and S. Bergler, *Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses*, Proceedings the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2006), 2006.
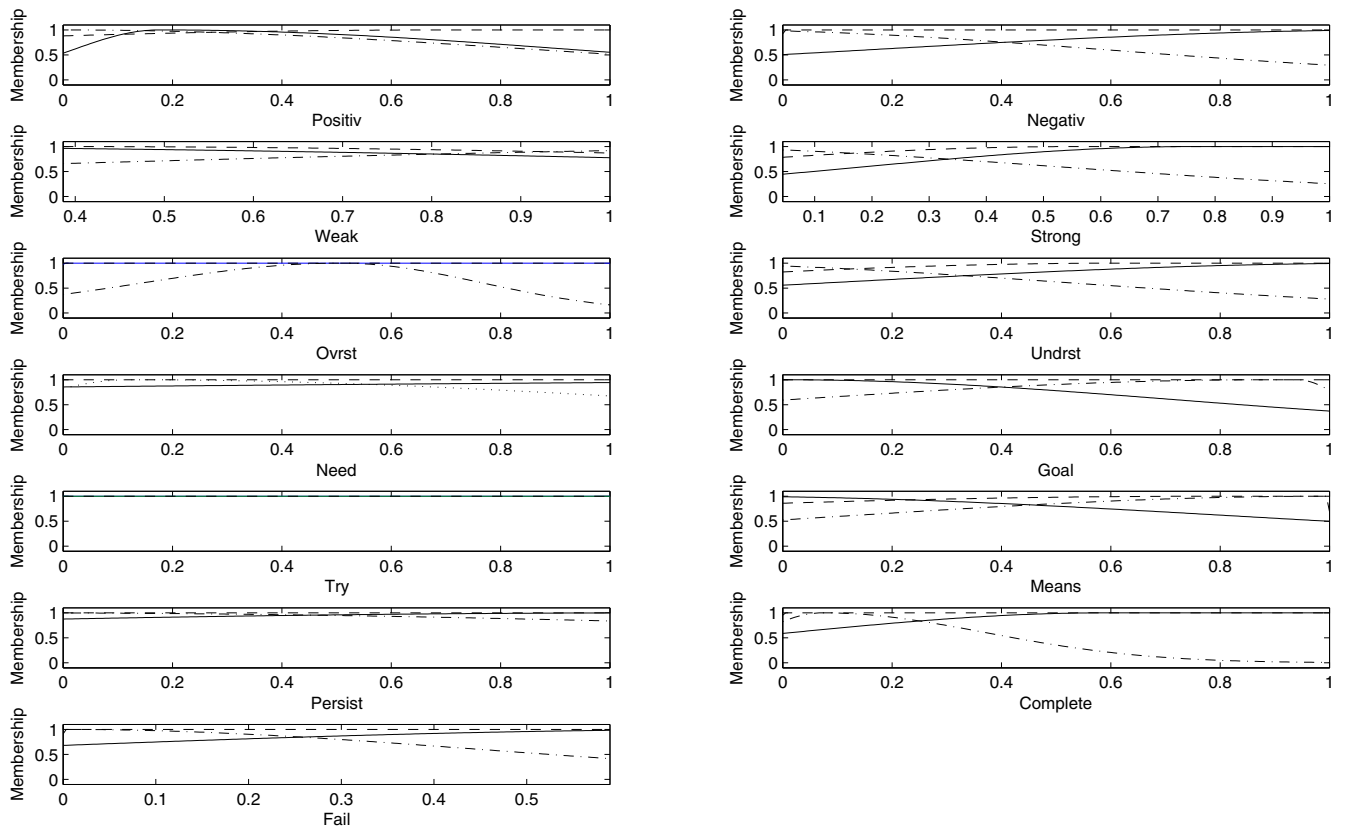[14] H. Sakaji, H. Sakai and S. Masuyama, *Automatic extraction of basis*

Fig. 4.   Input membership functions.

expressions that indicate economic trends, Advances in Knowledge Discovery and Data Mining, pp. 977-984, 2008.

[15] http://banker.thomsonib.com/.

[16] K. Toutanova and C.D. Manning, *Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger*, Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000), pp. 63-70, 2000.

[17] K. Toutanova, D. Klein, C. Manning and Y. Singer, *Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network*, Proceedings of HLT-NAACL 2003, pp. 252-259, 2003.

[18] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, 1981.

[19] R. Babuska, *Fuzzy modeling for control*, Kluwer Academic Publishers Norwell, MA, USA, 1998.

[20] U. Kaymak and R. Babuska, *Compatible cluster merging for fuzzy modelling*, Proceedings of 1995 IEEE International Conference on Fuzzy Systems, 1995.

[21] M. Setnes, R. Babuska, U. Kaymak and H.R. van Nauta Lemke, *Similarity measures in fuzzy rule base simplification*, IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 28, no. 3 pp. 376-386, 1998.