

Analysis of the Relation Among Global Stocks Using Fuzzy/Linguistic IF-THEN Rules

Sina Mirshahi^a and * Soheyla Mirshahi^b and Vilém Novák^b

^aDepartment of Informatics and Artificial Intelligent, Faculty of Applied Informatics, University of Tomas Bata, Nad Stráněmi 4511 , 760 05 Zlín, Czech Republic, Mirshahi@utb.cz

^bInstitute for Research and Applications of Fuzzy Modeling, NSC IT4Innovations, University of Ostrava, 30. dubna 22, 701 03 Ostrava 1, Czech Republic, Soheyla.Mirshahi@osu.cz, Vilem.Novak@osu.cz

Abstract

In this paper, using a fuzzy mathematical model, we analyze multiple stocks and produce a description for their interrelation. Our analysis is based on theories of F-transform and fuzzy natural logic. The suggested method consists of two main steps; first, we smooth out time series using F-transform. Afterward, we evaluate their similarities in different lags and provide a linguistic description of their interrelations. A python package is developed to automatize the suggested method fully. The method is tested on 26 global stocks from five different regions.

Keywords: Stocks relation, fuzzy modeling, time series analysis.

1 Introduction

In recent years, many factors, such as economic globalization, politics, internet communication, and lately pandemics, have strengthened the world's integration. Countries and professions are more dependent on each other than ever before, and changes in one business may lead to the other one. Therefore, to understand the dynamic structure of one financial time series, researchers must consider it jointly with other factors, such as different stock markets, unemployment rate, oil, and gold prices. There are two main questions in this direction; the first question is how to evaluate the interconnection among different financial time series. The second is how to model such a relation. These valuable questions interest many researchers from different fields such as physics, economy, computer science, and mathematics. For instance, in a study, Junior et al. evaluate the dependency relations among international stocks using partial Transfer Entropy [6]. In this study, they uncover the flow of information between 83 international market indices. In another study,

researchers investigate the causality relation between crude oil returns and eight stock returns [28]. Similar to the previous study, they applied Transfer Entropy while benefiting the Kernel method Granger Causality index to provide insight into the interrelationship between oil and equity markets. Wen, in 2019, adopts Pearson's correlation coefficient and the symmetrized Joe-Clayton copula model to uncover the tail dependence networks of global stock markets [27]. In another paper, the authors examine the return and volatility linkages between oil prices and the Lebanese stock market by applying VAR-GARCH (Vector Autoregressive-Generalized Autoregressive Conditional Heteroskedasticity) model to weekly data [1]. Studies in this area are centered around providing formulas or graphs to show these interrelations. This paper is the first study to provide a linguistic description of the uncovered interrelations to our best knowledge.

The techniques mentioned above are based on statistical theory. However, there are also non-statistical techniques based mostly on fuzzy set theory. An important and successful theory is that of *fuzzy transform* (F-transform) that has been applied to analysis and forecasting of time series (see [24, 23, 26, 22, 16, 20, 18]).

The key to these applications is the decomposition model which assumes that a time series can be additively decomposed into a trend-cycle, a seasonal component, and an irregular fluctuation (noise). The F-transform provides techniques for reducing noise and suppressing the seasonal component and estimation of the trend-cycle. The technique can be completed by methods based on *fuzzy natural logic* (FNL), using which various kinds of linguistically specified information can be mined from the time series.

Our previous paper has shown that the suggested method is well established and efficient [10]. Current paper extends the previous paper by analyzing the interrelation among 26 global stocks from different regions of the world and describes their uncovered relation lin-

guistically. The obtained results are well interpretable for human understating, which is an essential factor for financial experts [9]. Moreover, for the first time, we have developed a python package that automates the suggested method fully.

2 Fundamental theoretical tools

2.1 Fuzzy transform

Due to the lack of space, we will only briefly outline the main principles of the Fuzzy transform (F-transform) and refer the reader to the extensive literature, e.g., [19, 20] and many others.

2.1.1 Fuzzy partition

The F-transform is an approximation procedure applied, in general, to a bounded real continuous function $f : [a, b] \rightarrow [c, d]$ where $a, b, c, d \in \mathbb{R}$. It is based on the concept of a *fuzzy partition* that is a set $\mathcal{A} = \{A_0, \dots, A_n\}$, $n \geq 2$, of fuzzy sets fulfilling special axioms. The fuzzy sets are defined over nodes $a = c_0, \dots, c_n = b$ in such a way that for each $k = 0, \dots, n$, $A(c_k) = 1$ and $\text{Supp}(A_k) = (c_{k-1}, c_{k+1})^1$. The nodes are usually (but not necessarily) uniformly distributed, i.e., $c_{k+1} = c_k + h$ where $h > 0$ is a given value. To emphasize that the fuzzy partition is formed using the distance h , we will write \mathcal{A}_h .

2.1.2 Direct F-transform

The F-transform has two phases: direct and inverse. The *direct* F-transform assigns to each $A_k \in \mathcal{A}$ a component $F_k[f]$. The results is a vector of components

$$\mathbf{F}[f] = (F_0[f], \dots, F_n[f]). \quad (1)$$

We distinguish *zero degree* F-transform whose components $F_k^0[f]$ are numbers and first degree² F-transform whose components have the form

$$F_k^1[f](x) = \beta_k^0[f] + \beta_k^1[f](x - c_k).$$

The coefficient $\beta_k^1[f]$ provides estimation of an average value of the tangent (slope) of f over the area characterized by the fuzzy set $A_k \in \mathcal{A}$.

2.1.3 Inverse F-transform

From the direct F-transform (1) of f we can form a function $\tilde{f} : [a, b] \rightarrow [c, d]$ using the formula

$$\tilde{f}(x) = \sum_{k=0}^n (F_k[f] \cdot A_k(x)), x \in [a, b]. \quad (2)$$

¹Of course, certain formal requirements must be fulfilled. They are omitted here and can be found in the cited literature.

²In general, higher degree F-transform.

The function \tilde{f} is called the *inverse F-transform* of f (w.r.t. the fuzzy partition \mathcal{A}) and it approximates the original function f . For applications of the F-transform to the time series analysis see [20, 11].

2.2 Fuzzy natural logic

The paradigm of this theory is to become a formal logic aiming at modeling of natural human reasoning that necessarily proceeds in natural language. It is a class of theories of *mathematical fuzzy logic* that extends the classical concept of *natural logic* suggested by Lakoff in [8]. FNL is, so far, a class of the following formal theories:

- Formal theory of evaluative linguistic expressions [13].

Evaluative linguistic expressions are special expressions of natural language, such as *small, medium, big, roughly one hundred, very short, more or less deep, very tall, roughly warm or medium-hot, quite roughly strong*, etc. Note that such expressions are used by people regularly whenever they need to evaluate phenomena around them, to make decisions, and in various other activities based on their evaluation. A special formal theory of FNL makes it possible to capture semantics of evaluative expressions. The details can be found in [13] and also in [19].

- Formal theory of fuzzy/linguistic IF-THEN rules and approximate reasoning [12, 15, 17].

These rules have the form

$$\text{IF } X \text{ is } \mathcal{A} \text{ THEN } Y \text{ is } \mathcal{B} \quad (3)$$

where X and Y are antecedent and consequent variables, respectively, and \mathcal{A}, \mathcal{B} are evaluative linguistic expressions. The rules are taken as conditional clauses of natural language. Finite sets of fuzzy/linguistic IF-THEN rules are called *linguistic descriptions*. They represent a piece of text describing various decisions, control, and other kinds of situations. The details can be found in [15, 19].

- Formal theory of intermediate and generalized quantifiers [3, 14].

The semantics of evaluative expressions is based on the concept of (linguistic) *context*. The general idea is to specify what does it mean “small, medium”, or “big” in a given situation.

Let $v_L, v_S, v_R \in \mathbb{R}$ be three elements such that $v_L < v_S < v_R$. Then a *context* is a set $w = [v_L, v_S] \cup [v_S, v_R]$ where



Figure 1: Workflow pipe line.

v_L is the *most meaningful left bound*, v_S is the *most typical central point*, and v_R is the *most meaningful right bound*. The interval $[v_L, v_S]$ contains all small values and $[v_S, v_R]$ all big ones. Because the context w is uniquely determined by the three distinguished points v_L, v_S, v_R , we will usually write it as $w = \langle v_L, v_S, v_R \rangle$.

As an example, consider *high* or *small* sales of a company with five people. Then the context can normally be set to $w = \langle 10, 25, 90 \rangle$ (thousand \$), while for a big company with 1000 people, it can be $w = \langle 10, 40, 100 \rangle$ (million \$). In general, we can specify v_L as minimal and v_R as maximal value occurring in the data. A reasonable heuristic rule says that the middle value v_S should be nearer to v_L than to v_r which suggests to set $v_S = v_L + (v_R - v_L) * 0.4^1$.

3 The methodology

In this section, we provide a methodology for preparing stocks, assessing their similarity, and providing a natural language predictive linguistic description. This workflow which is shown in Figure 1, is based on the fuzzy transform (F-transform) and selected methods of fuzzy natural logic. The procedure described above is automated and is accessible in a Github repository ². The package is written in python and integrated with an open source R package *lfl R* implemented for the algorithms of linguistic fuzzy logic controller available at CRAN repository [21].

1. **Data acquisition:** the first step is to gather the stocks data. We use Yahoo finance for collecting the selected stocks of interests and extract their adjusted closing price.
2. **Data cleaning:** to evaluate the relationship between time series, it is necessary to examine the quality of the data set. Some stocks may miss some data points due to a variety of factors such as system faults, human mistakes, various working days, or unexpected closing days (e.g., Covid-19 pandemic). For some methods such as dynamic time warping (DTW), this task is not necessary,

¹This value can be justified by the golden section principle.

²https://github.com/Soheyla_Mirshahi/FuzzyMultiple-TimeseriesAnalysis

but for the purposed method, time series must have the same length size to measure their relationships. As a result, at this step, we concentrate on dealing with missing values and leveling the length of the time series using methods such as uniform scaling [7] or padding.

3. **F-transform:** financial time series (stocks, forex, cryptocurrencies, oil and gold prices) are among the most challenging time series to analyze [25, 29] Many interconnected factors, such as the economic indicators, politics, unexpected natural shocks such as a pandemic, features of the enterprises, and even investors psychology, highly affect the financial time series [30, 2].

As a result, we encounter a very dynamic non-linear series, which contains a lot of noise and chaos, and its behavior is highly uncertain. The F-transform, offers techniques for suppressing noise of time series while keeping the main features of the stocks. Thus at this stage, we assign to each stock an adjoint time series that is smoothened by the fuzzy transform.

4. **Similarity assessment:** after cleaning the data and reducing the noise using F-transform, it is possible to evaluate similarity among the time series of the given data set. The relation among stocks is evaluated using the similarity method according to the following definition.

Definition 3.1 Let $X = \{X(t) | t = 1, \dots, n\}$ and $Y = \{Y(t) | t = 1, \dots, n\}$ be two time series of the length n . Let us determine a fuzzy partition \mathcal{A} and construct inverse F-transforms \tilde{X}, \tilde{Y} of the time series X, Y . Then we define the similarity between these two time series as follows:

$$S(X, Y) = 1 - \frac{1}{n} \sum_{t=1}^n \frac{|\tilde{X}(t) - E(\tilde{X}) - (\tilde{Y}(t) - E(\tilde{Y}))|}{|\tilde{X}(t) - E(\tilde{X})| + |\tilde{Y}(t) - E(\tilde{Y})|}, \quad (4)$$

where $E(\tilde{X})$ and $E(\tilde{Y})$ are mean values (averages) of \tilde{X} and \tilde{Y} , respectively and $|\cdot|$ denotes absolute value. It is easy to show that $S(X, Y) \in [0, 1]$ and it is fuzzy symmetry (i.e., it is reflexive and symmetric). The \tilde{X}

and \tilde{Y} are smoothed out time series in which the noise is reduced (cf. [4, 19]). The output of this stage is a heat map containing similarity coefficients. At this point. The competency of the suggested method has been shown before, see [10].

5. **Linguistic evaluation:** this phase is responsible for evaluating the relationship among time series in natural language based on the theory of evaluative linguistics description that we mentioned in section [2.2]. The proposed method provides a comprehensive, interpretable description of the relationships between multiple stocks. The output of this step is a set of fuzzy linguistic If-Then rules describing the dependency among stocks.

The suggested methodology can be used for measuring the similarity between any number of stocks. Moreover, this approach has several advantages, especially in the context of financial time series. One of the objectives of the fuzzy model is to construct a model based on data that can be expressed in numbers and, more imprecisely, in the form of natural language expressions. Thus one of the key benefits of the suggested method is that it is closer to the way that investors think and reason in practice and is highly interpretable. Furthermore, the proposed approach is resistant to outliers. Since financial markets are susceptible to many outliers, this feature is critical. In the following section we apply the method on real world data.

4 Illustration

In this section, we provide a large-scale experiment to demonstrate the efficacy of the proposed method. Further investigation will focus on two key objectives. The major goal is to identify appropriate relationships among various stocks and then provide a linguistic description of their interactions. Various stock markets from six distinct world regions are chosen to investigate the relationships between stocks that differ geographically, politically, and in terms of trading volumes.

4.1 Data cleaning

Our data set consists of the daily price of 26 global stock indices from six different regions, North America and Canada, Latin America, Europe, Oceania, East Asia, and southwest Asia, from 2016 to 2020. The first step of the suggested workflow is data cleaning. As we mentioned earlier, these markets operate on different working days with various holidays; therefore, our data set contains various lengths and a significant amount of missing values. Omitting all days with a missing value

will lead to considerable data loss. To solve this problem, we define a condition applied in other researches -see [5, 6]. We solely remove the days on which less than 60% of stock markets operate on that specific day. We repeat the previous day's value to fill the missing values for markets that did not operate in the remaining days. This approach primarily affects the Israel market deeply, which have different weekends, so many working days were removed. Even after removing data and filling in missing values, it appears that some markets have a shorter length. Using the padding technique, we level all 26 time series. Our goal is to investigate the dependency among the current markets and their previous day (lagged values). Therefore, by adding their lagged values, we obtain a data set of 52 stocks, consisting of 26 price values and their previous day.

4.2 F-transform

Using the F-transform, we assign to each stock a smoothed time series which exhibits the local behavior of each stock market. An example of one stock from 2016 to 2020 and its smoothing is given in Figure 2. The sudden drop in 2020 is due to the pandemic's shock. The black squares in the illustration illustrate that the F-transform minimizes noise while preserving the essential feature of stock prices. For more details, see our previous paper [10].

4.3 Similarity assessment

In this stage, we calculate the return values, and using the method explained in the previous section, we build an enlarged similarity matrix between the original and their previous day. Figure 3 illustrates the results where the original return values were arranged from 0 to 25 and their lagged values from 26 to 52. Increased fuzzy similarity values are represented in darker shades, and lighter shades manifest lower similarity.

Figure 4 shows the fuzzy similarity for each region. Besides the dark main diagonal, representing the similarity of a market with itself, which is always 1, other clear clusters of substantial similarity are visible in this Figure. The whole European region and North America region form a relatively big similarity cluster; moreover, some small cluster is within East Asia. To gain more insight, we narrow down our further investigation to these regions. As we stated earlier, the stock market is very dynamic; thus, the dependency relation among two stocks might change. One critical question to answer is that if the discovered dependency remains the same in different years, or it will disappear. To answer this question, we evaluate the similarity between market indices over five years from 2016 to 2020. Figure 5 demonstrates the fuzzy similarity for the Euro-

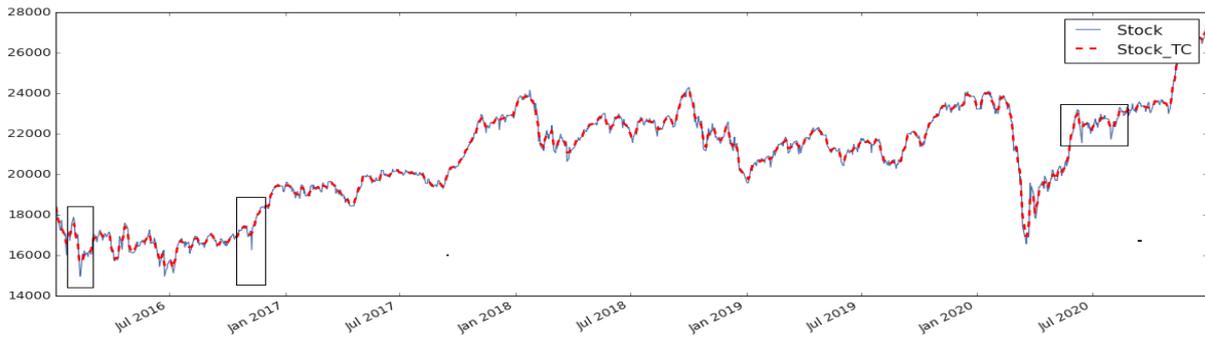


Figure 2: A stock time series and its smoothing.

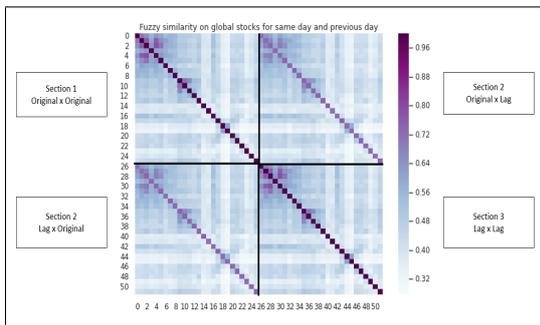


Figure 3: A heat map of the enlarged similarity matrix of both original and lagged indices, representing same-day similarity in Section 1 and previous-day similarity in Section 2. likewise, Sections 3 and 4 are representative of same day-previous day and previous day-previous day similarity, respectively.

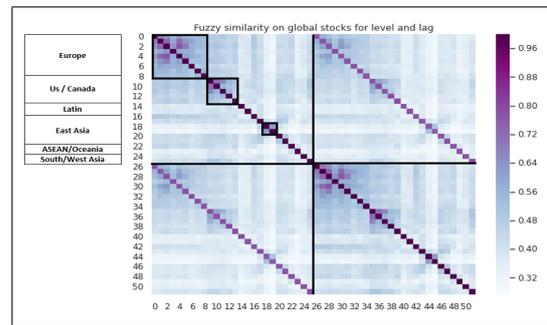


Figure 4: Heatmap of fuzzy similarity for different regions.

pean/North American region over five years. The figure shows that their similarity changes over time, where in 2016, 2018, and 2020 the similarity increases while 2017 and 2019 decrease. There might be many reasons for that, but note that the US-China trade war started from the beginning of January 2018, when the US began to impose sanctions on Chinese companies, affecting European markets non symmetrically. Besides, in 2020 the pandemic affects the whole European region dramatically, so it is no surprise that their similarity has increased. However, it seems that N100 (Paris market) and STOXX50E (stock index of Eurozone), shown by black squares, have kept their similarity over the five years. We will analyze this relation in a Fuzzy-linguistic form in the next section.

4.4 Linguistic evaluation

Using fuzzy natural logic methods, we form a linguistic description characterizing the relation between N100 and STOXX50E (for short STXE). The output of this step is a system of fuzzy/linguistic IF-THEN rules with

more than 100 rules. Selecting rules with a higher confidence degree is one way to reduce the size of the rule base. Another alternative is to ask experts to validate the rule base and select the most relevant rules based on their expertise. Table 1 displays some examples of the rules that experts have chosen. They reveal interesting information about today's and yesterday's prices (lag-value) for the mentioned stocks. For instance, the first rule tells that if yesterday, N100 and STXE had a *more or less medium* and a *more or less big* decrease respectively, and today STXE continues to have a *more or less big* decrease, then N100 is experiencing a *very big* decrease today. Similarly, by the second rule, we conclude that if yesterday, N100 and STXE had a *typical medium and medium* decrease sequentially, and today, STXE changes the direction and shows a *more or less medium* increase, then N100 has a *medium* increase today.

4.5 Conclusion

In continuation of our previous paper, we introduced a methodology to analyze the relationships among multiple stocks using fuzzy/linguistic IF-THEN rules. We illustrated our method on a large data set composed of

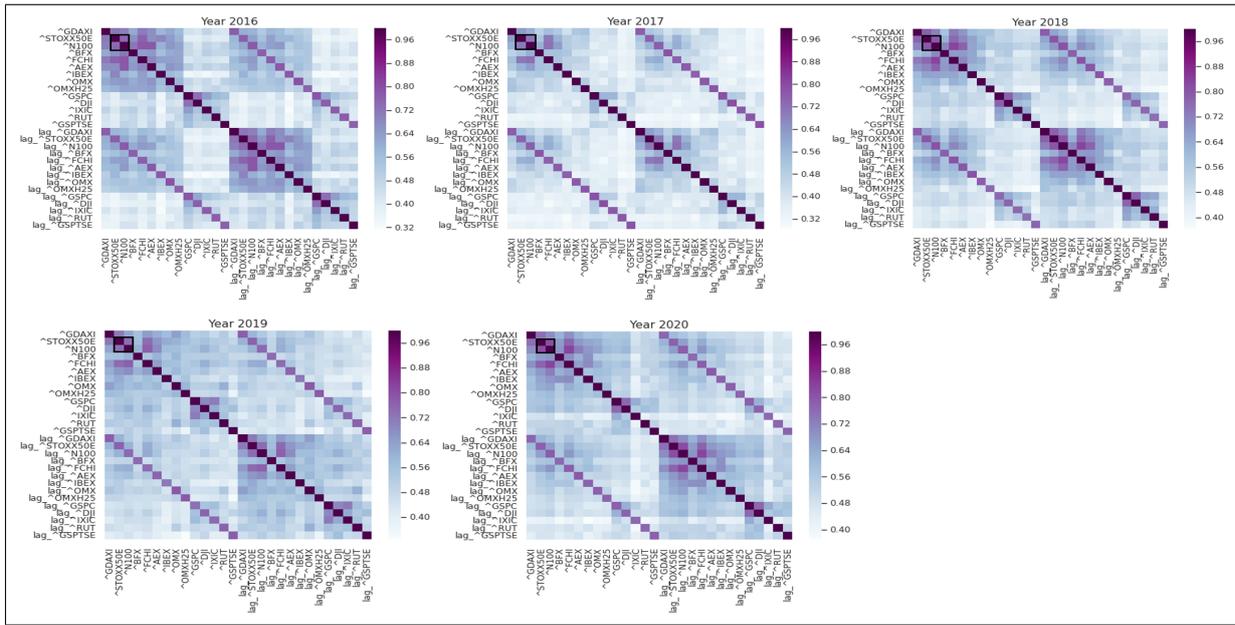


Figure 5: Heatmap of fuzzy similarity from 2016 to 2020

lag-N100	&	lag-STXE	&	STXE	→	N100
-ml me	&	-ml bi	&	-ml bi	→	-ve bi
-ty me	&	-me	&	ml me	→	me
vr sm	&	vr sm	&	-ro sm	→	-ml sm
-vr sm	&	-ml me	&	me	→	ml me
vr bi	&	ml bi	&	ml bi	→	vr bi

Table 1: A linguistic description describing relation among N100 and STOXX50E. *Vr*, *ml*, and *ro* are abbreviations for very, more or less, and roughly, and *sm*, *me* and *bi* stand for small, medium, and big, respectively.

26 global stocks over five years. In addition, a python package for calculation is created in order to automate the process. To the best of our knowledge, we are the first to propose a language description for the study of multiple stock relationships. Further research will compare our model to others, such as the vector autoregressive model. Another addition to this work can be extending the method for other financial time series such as commodities

Acknowledgement

This work was supported by Internal Grant Agency of Tomas Bata University under the Project No. GA/CebiaTech/2021/001 and by resources of A.I.Lab research group at Faculty of Applied Informatics, Tomas Bata University in Zlin (ailab.fai.utb.cz). Another support was given by ERDF/ESF project “Centre for the development of Artificial Intelligence Methods for the Automotive Industry of the region” No. CZ.02.1.01/0.0/0.0/17/ -049/0008414.

References

- [1] E. Bouri, Return and volatility linkages between oil prices and the lebanese stock market in crisis periods, *Energy* 89 (2015) 365–371.
- [2] H. Chen, K. Xiao, J. Sun, S. Wu, A double-layer neural network framework for high-frequency forecasting, *ACM Transactions on Management Information Systems (TMIS)* 7 (4) (2017) 1–17.
- [3] A. Dvořák, M. Holčapek, L-fuzzy quantifiers of the type $\langle 1 \rangle$ determined by measures, *Fuzzy Sets and Systems* 160 (2009) 3425–3452.
- [4] M. Holčapek, V. Novák, I. Perfilieva, Noise reduction in time series using F-transform, in: *Proc. Int. Conference FUZZ-IEEE’2013, Hyderabad, India, 2013*.
- [5] L. S. Junior, I. D. P. Franca, Correlation of financial markets in times of crisis, *Physica A: Statistical Mechanics and its Applications* 391 (1-2) (2012) 187–208.
- [6] L. S. Junior, A. Mullokandov, D. Y. Kenett, Dependency relations among international stock market indices, *Journal of Risk and Financial Management* 8 (2) (2015) 227–265.
- [7] E. Keogh, Efficiently finding arbitrarily scaled patterns in massive time series databases, in: *European Conference on Principles of Data Mining and Knowledge Discovery, Springer, 2003*, pp. 253–265.

- [8] G. Lakoff, Linguistics and natural logic, *Synthese* 22 (1970) 151–271.
- [9] S. Mirshahi, N. Cao, Fuzzy relational compositions can be useful for customers credit scoring in financial industry, in: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Springer, 2018, pp. 28–39.
- [10] S. Mirshahi, V. Novák, A fuzzy approach for similarity measurement in time series, case study for stocks, in: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Springer, 2020, pp. 567–577.
- [11] L. Nguyen, V. Novák, Forecasting seasonal time series based on fuzzy techniques, *Fuzzy Sets and Systems*.
- [12] V. Novák, Perception-based logical deduction, in: *Computational Intelligence, Theory and Applications*, Springer, 2005, pp. 237–250.
- [13] V. Novák, A comprehensive theory of trichotomous evaluative linguistic expressions., *Fuzzy sets and systems* 159 (22) (2008) 2939–2969.
- [14] V. Novák, A formal theory of intermediate quantifiers, *Fuzzy Sets and Systems* 159 (10) (2008) 1229–1246.
- [15] V. Novák, S. Lehmke, Logical structure of fuzzy if-then rules, *Fuzzy Sets and Systems* 157 (15) (2006) 2003–2029.
- [16] V. Novák, V. Pavliska, M. Štěpnička, L. Štěpničková, Time series trend extraction and its linguistic evaluation using F-transform and fuzzy natural logic, in: *Studies in Fuzziness and Soft Computing*, Springer Verlag, 2014, pp. 429–442.
- [17] V. Novák, I. Perfilieva, On the semantics of perception-based fuzzy logic deduction, *International journal of intelligent systems* 19 (11) (2004) 1007–1031.
- [18] V. Novák, I. Perfilieva, Forecasting Direction of Trend of a Group of Analogous Time Series Using F-Transform and Fuzzy Natural Logic, *The International Journal of Computational Intelligence Systems* 8 (2015) 15–28.
- [19] V. Novák, I. Perfilieva, A. Dvořák, *Insight into Fuzzy Modeling*, Wiley & Sons, Hoboken, New Jersey, 2016.
- [20] V. Novák, I. Perfilieva, M. Holčapek, V. Kreinovich, Filtering out high frequencies in time series using F-transform, *Information Sciences* 274 (2014) 192–209.
- [21] V. Novák, I. Perfilieva, N. G. Jarushkina, A general methodology for managerial decision making using intelligent techniques, in: E. Rakus-Anderson, R. Yager, N. Ichalkaranje, L. Jain (Eds.), *Recent Advances in Fuzzy Decision-Making*, Springer, Heidelberg, 2009, pp. 103–120.
- [22] V. Novák, I. Perfilieva, V. Krejnovich, F-transform in the analysis of periodic signals, in: M. Inuiguchi, Y. Kusunoki, M. Seki (Eds.), *Proc. 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty*, Osaka University, Osaka, Japan, 2012, pp. 150–158.
- [23] V. Novák, M. Štěpnička, A. Dvořák, I. Perfilieva, V. Pavliska, L. Vavříčková, Analysis of seasonal time series using fuzzy approach, *Int. Journal of General Systems* 39 (2010) 305–328.
- [24] V. Novák, M. Štěpnička, I. Perfilieva, V. Pavliska, Analysis of periodical time series using soft computing methods, in: D. Ruan, J. Montero, J. Lu, L. Martínez, P. D’hondt, E. Kerre (Eds.), *Computational Intelligence in Decision and Control*, World Scientific, New Jersey, 2008, pp. 55–60.
- [25] F. E. Tay, L. Cao, Application of support vector machines in financial time series forecasting, *omega* 29 (4) (2001) 309–317.
- [26] M. Štěpnička, A. Dvořák, V. Pavliska, A linguistic approach to time series modeling with the help of F-transform, *Fuzzy Sets and Systems* 180 (2011) 164–184.
- [27] F. Wen, X. Yang, W.-X. Zhou, Tail dependence networks of global stock markets, *International Journal of Finance & Economics* 24 (1) (2019) 558–567.
- [28] D. Xiao, J. Wang, Dynamic complexity and causality of crude oil and major stock markets, *Energy* 193 (2020) 116791.
- [29] N. Zhang, A. Lin, P. Shang, Multidimensional k-nearest neighbor model based on eemd for financial time series forecasting, *Physica A: Statistical Mechanics and its Applications* 477 (2017) 161–173.
- [30] X. Zhong, D. Enke, Forecasting daily stock market return using dimensionality reduction, *Expert Systems with Applications* 67 (2017) 126–139.