UNIVERSITY OF OSTRAVA
FACULTY OF SCIENCE
DEPARTMENT OF INFORMATICS AND COMPUTERS

A FUZZY MODEL FOR MULTIPLE TIME SERIES ANALYSIS AND FORECASTING USING SIMILARITY

PhD. Thesis

AUTHOR: SOHEYLA MIRSHAHI
SUPERVISOR: PROF. ING. VILÉM NOVÁK, DRSC.

2021

OSTRAVSKÁ UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA INFORMATIKY A POČÍTAČŮ

Fuzzy model pro analýzu a predikci více časových řad pomocí podobnosti

Disertační práce

Autor práce: SOHEYLA MIRSHAHI
Vedoucí práce: Prof. Ing. VILÉM NOVÁK, DrSc.

2021

OSTRAVSKÁ UNIVERZITA

Přírodovědecká fakulta

Akademický rok: 2020/2021

ZADÁNÍ DISERTAČNÍ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	**Mgr. Soheyla MIRSHAHI**
Osobní číslo:	**P16243**
Studijní program:	**P1802 Aplikovaná informatika**
Studijní obor:	**Aplikovaná informatika**
Téma práce:	**Fuzzy model pro analýzu a predikci více časových řad pomocí podobnosti**
Téma práce anglicky:	**A fuzzy model for multiple time series analysis and forecasting using similarity**
Zadávající katedra:	**Katedra informatiky a počítačů**

Zásady pro vypracování

Tato práce se zaměřuje na aplikaci teorie F-transformace k analýze vztahu mezi více časovými řadami a poskytnutí jejich prediktivního jazykového popisu. Tohoto cíle je dosaženo nejprve vyvinutím dvou nových metod podobnosti založených na komponentách F-transformace pro identifikaci finančních časových řad (např. ceny akcií, návratnost akciového trhu) s podobným chováním. Tyto metody jsou založeny na aplikaci fuzzy transformace a přizpůsobené metriky. V návaznosti na to představujeme systematický přístup k poskytování prediktivních jazykových komentářů založených na předcházejících jevech. Dvě základní teorie navrhované metody jsou metoda F-transformace a vybraná teorie fuzzy přirozené logiky. Kromě poskytnutí matematického důkazu o navrhovaných metodách podobnosti se provádí několik rozsáhlých experimentů, které ilustrují použitelnost navrhované metody v praxi. Vyvinuli jsme komplexní pracovní postup pro navrhovaný přístup k analýze a předpovědi více časových řad. Navrhovaný pracovní postup zahrnuje čtyři hlavní fáze: příprava dat, zpracování dat, hodnocení podobnosti a fuzzy modelování a predikce. Při přípravě dat se zkoumá kvalita dat týkajících se chybějících hodnot a různých délek časových řad. Výstupem této fáze jsou čistá data připravená k další analýze. Fáze zpracování dat je zodpovědná za manipulaci a odhad místního trendového cyklu na základě F-transformace. Výsledkem je snížení šumu dat při zachování hlavních vlastností. Následující fáze je věnována vyhodnocení podobnosti mezi časovými řadami. Tento krok je srdcem navrhované metody poskytující přehled o různých úkolech, jako je výběr portfolia nebo analýza závislostí. Poslední krok se používá k prediktivnímu jazykovému popisu více finančních časových řad. Navrhovaný pracovní postup je naprogramovaný a je přístupný v úložišti GitHub. K ověření použitelnosti navrhované metody jsou ilustrovány různé příklady, které vysvětlují výhody navrhované metody a porovnávají ji s přístupy, které se v současnosti používají v praxi. Tento přístup je intuitivní, snadno interpretovatelný a nabízí další pohled na chování časových řad.

Rozsah pracovní zprávy:	**82**
Rozsah grafických prací:	
Forma zpracování disertační práce:	**tištěná/elektronická**
Jazyk zpracování:	**Angličtina**

Seznam doporučené literatury:

1. Mirshahi, Soheyla, and Vilém Novák. „A fuzzy method for evaluating similar behaviour between assets."Soft Computing (2021): 1-11.
2. Novák, Vilém, and Soheyla Mirshahi. „On the Similarity and Dependence ofTime Series."Mathematics 9, no. 5 (2021): 550.
3. Nguyen, Linh, Vilém Novák, and Soheyla Mirshahi. „Trend-cycle EstimationUsing Fuzzy Transform and Its Application for Identifying Bull and Bear Phases in Markets."Intelligent Systems in Accounting, Finance, and Management27, no. 3 (2020): 111-124.
4. Mirshahi, Soheyla, and Vilém 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, pp. 567-577. Springer, Cham, 2020
5. Novák, Vilém, Soheyla Mirshahi, and Viktor Pavliska. „LFL Forecaster: Analysis, Forecasting and Mining Information from Time Series."In 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-6. IEEE, 2019.

Školitel disertační práce: **prof. Ing. Vilém Novák, DrSc.**
Centrum excelence IT4Innovations

Datum zadání disertační práce: **8. června 2021**
Termín odevzdání disertační práce: **31. srpna 2022**

V. Z.

doc. RNDr. PaedDr. Eva Volná, PhD.
garantka oboru

prof. Ing. Vilém Novák, DrSc.
vedoucí doktorské práce

doc. RNDr. PaedDr. Hashim Habiballa, PhD., Ph.D.
vedoucí katedry

V Ostravě dne 9. června 2021

Já, níže podepsaná studentka, tímto čestně prohlašuji, že text mnou odevzdané závěrečné práce v písemné podobě i na CD nosiči je totožný s textem závěrečné práce vloženým v databázi DIPL2.

Prohlašuji, že předložená práce je mým původním autorským dílem, které jsem vypracovala samostatně. Veškerou literaturu a další zdroje, z nichž jsem při zpracování čerpala, v práci řádně cituji a jsou uvedeny v seznamu použité literatury.

V Ostravě dne ....

podpis

Abstract

The thesis focuses on applying F-transform theory to analyse the relations among multiple time series and provide a predictive linguistic description for them. This objective is accomplished by first developing two new similarity methods based on F-transform components for identifying financial time series (e.g., stock prices, stock market return) with similar behaviour. These methods are based on the application of the fuzzy transform and a customized metric. Following that, we present a systematic approach to providing predictive linguistic explanations based on the previous discovery. Fundamental theories for the suggested method are F-transform method and selected theory of fuzzy natural logic. Besides providing mathematical proof of the suggested similarity methods, several extensive experiments are conducted to illustrate the applicability of the suggested method in practice. Finally, we develop a comprehensive workflow for the suggested approach for multiple time series analysis and forecasting. The suggested workflow includes four main stages: data preparation, data processing, similarity assessment, and Fuzzy natural modeling and forecasting. Briefly, in data preparation, the quality of data regarding missing values and varying time series lengths are examined. The output of this phase is clean data ready for further analysis. The data processing stage is responsible for manipulating and estimation of local trend-cycle based on F-transform. As a result, the noise of the data is reduced while the main features are preserved. The following stage is dedicated to evaluating similarity among time series for lag zero and shifted lags. This step is the heart of the suggested method providing insight for various tasks such as portfolio selection or dependency analysis. And finally, the last step is used to provide a predictive linguistic description for multiple financial time series. The suggested workflow is fully automated and is accessible in a GitHub Repository. To validate the applicability of the proposed method, different examples are illustrated, explaining the benefits of the proposed method and comparing it to approaches currently used in practice. This approach is intuitive, easily interpretable, and offers further insight into the behaviour of time series.

Key Words: Fuzzy transform, Multivariate time series forecasting,
Similarity measure, Time series analysis, Trend-cycle similarity.

Abstrakt

Tato práce se zaměřuje na aplikaci teorie F-transformace k analýze vztahu mezi více časovými řadami a poskytnutí jejich prediktivního jazykového popisu. Tohoto cíle je dosaženo nejprve vyvinutím dvou nových metod podobnosti založených na komponentách F-transformace pro identifikaci finančních časových řad (např. ceny akcií, návratnost akciového trhu) s podobným chováním. Tyto metody jsou založeny na aplikaci fuzzy transformace a přizpůsobené metriky. V návaznosti na to představujeme systematický přístup k poskytování prediktivních jazykových komentářů založených na předcházejících jevech. Dvě základní teorie navrhované metody jsou metoda F-transformace a vybraná teorie fuzzy přirozené logiky. Kromě poskytnutí matematického důkazu o navrhovaných metodách podobnosti se provádí několik rozsáhlých experimentů, které ilustrují použitelnost navrhované metody v praxi. Vyvinuli jsme komplexní pracovní postup pro navrhovaný přístup k analýze a předpovědi více časových řad. Navrhovaný pracovní postup zahrnuje čtyři hlavní fáze: příprava dat, zpracování dat, hodnocení podobnosti a fuzzy modelování a predikce.

Při přípravě dat se zkoumá kvalita dat týkajících se chybějících hodnot a různých délek časových řad. Výstupem této fáze jsou čistá data připravená k další analýze. Fáze zpracování dat je zodpovědná za manipulaci a odhad místního trendového cyklu na základě F-transformace. Výsledkem je snížení šumu dat při zachování hlavních vlastností. Následující fáze je věnována vyhodnocení podobnosti mezi časovými řadami.Tento krok je srdcem navrhované metody poskytující přehled o různých úkolech, jako je výběr portfolia nebo analýza závislostí. Poslední krok se používá k prediktivnímu jazykovému popisu více finančních časových řad. Navrhovaný pracovní postup je naprogramovaný a je přístupný v úložišti GitHub. K ověření použitelnosti navrhované metody jsou ilustrovány různé příklady, které vysvětlují výhody navrhované metody a porovnávají ji s přístupy, které se v současnosti používají v praxi. Tento přístup je intuitivní, snadno interpretovatelný a nabízí další pohled na chování časových řad.

Klíčová slova: Fuzzy transformace, Vícerozměrné predikce časových řad, Měření podobnosti, Analýza časových řad, Podobnost trendového cyklu.

I would like to express my deepest gratitude to professor Novák, my supervisor as well as my mentor in many areas of life. He always guided me through the ups and downs of my studies while giving me enough freedom to find my interest. Without professor Novák, I would not have discovered my passion for financial data analysis and forecasting and could not finish this thesis.

Furthermore, I want to express my heartfelt appreciation to my dear husband, Pavel, who was very supportive during the long emotional journey of completing this thesis, whether it was enthusiasm, disappointment, uncertainty or pride.

Last but not least, I would like to dedicate this thesis to my parents, who were not scientists themselves but still valued science and made many sacrifices to help me on this journey.

Contents

1 Introduction

In recent years, many factors, such as economic globalization, politics, internet communication, and lately pandemics, have strengthened world integration. Countries and professions are more dependent on each other than ever before. Changes in one business may lead to the other one, and therefore to understand the dynamic structure of one, researchers must consider it jointly with other factors. It is clear knowing how these factors are interrelated is of great importance. These indicated data are often represented in the form of time series, and luckily, today, large datasets of time series are available in many fields, providing comprehensive data. Time series analysis is not a brand-new topic. Many researchers have already studied this area (-see [1, 2, 3, 4, 5]) and Yang and Wu [6] rate it as one of the top ten challenging data mining problems due to its particular properties. Mining time series data includes many subareas; however, assessing time series similarity, i.e., the degree to which a given time series resembles another, is the core of many of them, such as retrieval and clustering, classification, and even forecasting [7].

The fuzzy transform (or simply F-transform) is a universal technique introduced by Perfilieva in [8] and further elaborated in [9] and several other papers. Its fundamental idea is to map a bounded continuous function $f : [a, b] \longrightarrow \mathbb{R}$ to a finite vector of numbers and then to transform it back. The former is called a direct F-transform and the latter an inverse one. The result of the inverse F-transform is a function $\hat{f}$ that approximates the original function f. Parameters of the F-transform can be set in such a way that the approximating function $\hat{f}$ has desired properties. In the case of time series, it was proved in [10] that the F-transform has the ability to filter out high frequencies and minimize noise. Simultaneously, average values of first and second derivatives of f over a specified area can be estimated.

With the first publication [11], the F-transform was successfully applied to time series analysis and forecasting. Then, in [12, 13, 14, 15, 16, 17] it was further elaborated. From all of these studies, it is clear that the F-transform is a reliable technique for time series processing. The key to these applications is the additive decomposition model of time series. The F-transform, offers techniques for suppressing noise and estimating the trend-cycle under the assumption that a time series can be additively decomposed into a trend-cycle, a seasonal variable, and an irregular fluctuation (noise). Moreover, it can also be used to forecast the trend-cycle and seasonal components by using fuzzy natural logic (FNL) techniques. Nevertheless, most of these investigations are limited to univariate time series in which the potential dependency of the variable with other possible variables has been neglected. It will be a significant deficiency if we do not apply the F-transform to multivariate time series. As we mentioned earlier, the F-transform application to the univariate time series has been developed and improved later. In these approaches, the time series analysis is considered to be decomposed into several components followed by characterization and prediction of each component separately. There is a significant advantage: its results are well interpretable [18], which is an essential factor for financial experts [19]. Therefore our focus in this thesis is to extend the approach for multivariate time series. However, to our best knowledge, there are only a few researches [20, 21] about analysis with fuzzy natural logic in this field. Therefore the necessity in focusing our research in this area is definite.

This thesis is focused on proposing a fuzzy method for discovering the relationships among multiple financial time series and to provide a predictive linguistic description for it. Our method is based on the fuzzy transform (F-transform) and selected methods of fuzzy natural logic. There are two main parts of the thesis: the first part focuses on time series data mining. To accomplish this, we establish two fuzzy similarity measurement methods. The aim is to find such time series that behave similarly. In the second part, using fuzzy natural logic, we propose a linguistic forecasting method based on this interrelation among similar time series.

2 Author's contribution

The two key questions of the thesis are "How to identify financial time series which behave similarly " and "how to provide a complete transparent predictive model using the dependency between financial time series." Based on a comprehensive study of literature, I believe this topic is brand-new, and I am the first to recognize this viewpoint. In conclusion, I have addressed the thesis key questions with the following contributions:

- I proposed two powerful similarity methods to measure similar behaviour of time series based on estimated local trend-cycle in section [6]. I validated these methods theoretically using mathematical proof. In addition, I conducted several extensive experiments to demonstrate their capability and power to assess the similarity degree in practice.

- I suggested a straightforward workflow in section [7] for analysis and forecasting one time series based on multiple similar time series using the F-transform method and selected methods of Fuzzy natural Logic.

- To improve the accessibility of the suggested workflow I automatized the method using the most popular programming language in data science. The code in the GitHub repository -a cloud service for versioning and sharing open-source codes- for all researchers to customize and apply according to their needs. Using Github, the future improvement of the code is much easier and agile.

- In section [7] I suggested practical guidance for adjusting parameters of the F^m-transform (the degree of the fuzzy transform and the bandwidth of the fuzzy sets forming the fuzzy partition) to get a precise estimation of the financial time series trend-cycle.

3 List of author's publications

3.1 Journal articles

- Mirshahi, Soheyla, and Vilém Novák. "A fuzzy method for evaluating similar behaviour between assets." Soft Computing (2021): 1-11.
- Novák, Vilém, and Soheyla Mirshahi. "On the Similarity and Dependence of Time Series." Mathematics 9, no. 5 (2021): 550.
- Nguyen, Linh, Vilém Novák, and Soheyla Mirshahi. "Trend-cycle Estimation Using Fuzzy Transform and Its Application for Identifying Bull and Bear Phases in Markets." Intelligent Systems in Accounting, Finance and Management 27, no. 3 (2020): 111-124.
- Khadivar, Ameneh, Soheyla Mirshahi, and Sara Aghababaei. "Modeling and knowledge acquisition processes using case-based inference." Iran. J. Inf. Process. Manag 32, no. 2 (2017): 467-490.

3.2 Conference articles

- Mirshahi, Soheyla, and Vilém 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, pp. 567-577. Springer, Cham, 2020.
- Novák, Vilém, Soheyla Mirshahi, and Viktor Pavliska. "LFL Forecaster: Analysis, Forecasting and Mining Information from Time Series." In 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-6. IEEE, 2019.
- Mirshahi, Soheyla, and Nhung 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, pp. 28-39. Springer, Cham, 2018.
- Mirshahi, Soheyla, Sina Mirshahi and Vilém Novák. "Analysis of the relation among global stocks using Fuzzy/Linguistic IF-THEN rules." The 12th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), 2021 (Accepted)

4 Literature overview

4.1 Financial time series

Time series data comprises data points obtained from a real-valued, continuous process across time. There are several properties of time-series data that make it distinct from other types of data. Firstly, the sampled time series are data often high dimensional and contain a considerable amount of noise. Thus, dimensionality reduction techniques are necessary to remove noise, reduce dimensionality, and extract important features. However, the feature extraction application has many disadvantages [22]; there is the danger of losing valuable information; thus, the choice of features and signal processing techniques is vital. Secondly, there is uncertainty about whether enough information is available to understand the process or not. There is no guarantee that the selection of the past data is sufficient for prediction. For instance, in the case of financial time series, a single stock only covers a small aspect of a much more complex system. Therefore, there is most likely non-adequate information to predict the future [23]. This can be another explanation for why multiple time series analysis is essential. Financial time series (stocks, forex, cryptocurrencies, oil and gold prices) are among the most challenging time series to analyze and forecast [24, 25]. 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 [26, 27].

As a result, we encounter a very dynamic non-linear series, which contains a lot of noise and chaos, and its behaviour is highly uncertain. [28, 29, 30, 31]. Therefore, the analysis of financial time series is a highly empirical discipline, but the theory forms the foundation of making an inference like any other scientific field. Stock financial time series forecasting is a challenging task. Through the years, several theories have been proposed. Their objective was to either explain the complexity of asset markets or determine whether they could be beaten. In 1970 Fama [32] introduced one of the most influential and fiercely debated theories named *efficient market hypothesis*. This hypothesis argues that the current price of any asset reflects all available information; thus, price movements are unpredictable. However, many researchers argue otherwise; for instance, in [33] the presence of numerous price patterns in financial markets and the serial correlations between actual events and economic values affecting the markets are mentioned as reasoning against the efficient market hypothesis. Thanks to the challenging and ever-changing landscape of stock markets, many studies from various directions have questioned the validity of this hypothesis and introduced new approaches over the years [34].

4.2 Analysis and forecasting models

Time series analysis contains methods for investigating time series data in order to extract meaningful statistics and other aspects of the data. Simultaneously, the use of a model to predict future values based on previously observed values is called time series forecasting. We can divide these methods into three categories in a broad sense: classical methods, machine learning models, and combinations of these two. These techniques will be briefly explained in the following sections.

4.2.1 Classical models

One of the first classical approaches which are surprisingly common is *judgmental forecasting*. In many cases, such as a complete lack of historical data, introducing a new product, having a new competitor to the market, or occurrence of entirely distinctive and unusual market conditions, judgemental forecasting has a firm place [35]. For example, in March 2020, the World Health Organization (WHO) declared the coronavirus outbreak a global pandemic leading to a complete panic as the oil price for the first time in history dropped to a negative. There were no historical precedents in this case; thus, a judgment must be exercised in predicting the impact of such conditions on financial markets. In other situations, data might be incomplete; therefore, some judgmental prediction is required. For instance, GDP is only available every quarter, and central banks use judgment when predicting the current economic activity level, a process known as *nowcasting*. The judgmental approach is subjective, and it has limitations such as being bias or inconsistent [36]. Therefore a better approach can be to apply systematic and well-structured forecasting methods and adjust them with judgmental forecasting afterward. Clearly, for this purpose, the forecasting method should be interpretable and easy to understand for experts.

Another important classical approach dates back to the 1920s. It is a straightforward procedure that is considered as the foundation for most other time series decomposition techniques. An additive decomposition and a multiplicative decomposition are the two forms of classical decomposition. It is not easy to model a time series as a whole since it manifests a large number of components influencing each other. Therefore, it is beneficial to divide a time series into important categories of patterns that are more or less understandable. Hence, it is more common to classify constituents appearing in a time series into four components characterizing important properties of the time series, namely, *trend*, *cycle*, *seasonality* and *noise* (irregular fluctuation). A time series have a trend when there is a long-term increase or decrease in data. The cycle exists when data exhibit rises and falls are not of a fixed period. In practice, the trend and the cycle usually occur together, and it is not easy to distinguish one from the other. Thus, these two components usually are joined and addressed as one component called *trend-cycle*. A continuous function called $TC(t)$ models this component that smoothly changes in its course.

The third component, named seasonality, represents the possible effects of seasonal factors (e.g., daily, weekly, monthly, the quarter of the year, the month, or annually) to a time series. These influences repeat themselves in a fixed and known period. Consequently, this component is usually modeled by a periodic function $S(t)$ with a known period. The noise is the last component that reflects irregular properties exhibiting in a time series. This component is usually assumed to be a realization $R(t)$ of a stationary process with bounded variance, and zero mean. Therefore, in classical approaches, usually time series analysis consists of its decomposition into several components before the characterization and prediction of each component individually. To estimate the trend-cycle, the first step in a classical decomposition is to use a moving average process [35]. Among the statistical approaches are several models that gain more popularity. Auto-Regressive Moving Average (ARMA), the Auto-Regressive Integrated Moving Average (ARIMA), the Generalized Auto-Regressive Conditional Heteroscedastic (GARCH) and the smooth transition autoregressive volatility are among these models [26]. Some researcher have reported the ARIMA model as the most used tool for stock market analysis [37]. George Box and Gwilym Jenkins, in 1970, developed a very influential method for describing, fitting, measuring, and using integrated autoregressive moving average (ARIMA) time series [38]. A very influential method for defining, fitting, testing, and using integrated autoregressive, moving average (ARIMA) time series is introduced by George Box, and Gwilym Jenkins in 1970 [38]. They formed a systematic methodology showing how to identify and estimate the models incorporating both autoregressive and moving average approaches.

4.2.2 Machine learning methods

The methods mentioned earlier, such as moving averages, autoregressive models, were among the primary prediction techniques used in stock markets [39, 29]. In practice, two approaches are widespread. First is the technical analysis, where standards of support and resistance and other indicators are determined from past prices [40]. The so-called fundamental analysis is the second approach that focuses on exploring the impact of business, and economic factors on the market trends [41]. Nevertheless, some traditional machine learning modeling methods such as support vector machine (SVM) [42], random forest, and fuzzy analysis [40, 43] has also been applied in the finance sector. Note that, in general, these techniques ignore the time aspect of the data and model the time series similar to other data types. However, In 2010, Novák et al. suggest a fuzzy method for time series analysis, considering the time dependency among data [44]. This method was later extended and applied to the finance sector in another paper in 2020 [45]. Some researchers [46, 47] applied ANN and SVM for predicting stock price index movement and compared the performances of both models. Their conclusion shows that the average performance of the ANN model was significantly better than the SVM model. However, some researchers report the otherwise. For instance, in [48] Patel et al. predict the direction of the Indian stock market using several machine learning techniques, including the neural network, support vector machine, and random forest. They primarily select ten technical parameters from stock data and transform them into trend deterministic data as inputs. They conclude, the random forest gave the highest prediction performance. The following section will describe the neural networks and their application in financial time series analysis in detail.

4.2.3 Support Vector Machines

A support vector machine (SVM) is a machine learning method that performs classification and regression problems. In addition to linear classification, SVM can perform nonlinear classification. A linear model is used to achieve nonlinear class boundaries by some nonlinear mapping of the input vector x into the high-dimensional feature space. The constructed linear model in the new space, known as the optimal separating hyperplane, can describe a nonlinear decision boundary in the original space. In other words, SVM is an algorithm that finds a special kind of linear model, the maximum margin hyperplane, which provides the maximum separation between the decision classes. The support vectors are those training examples that are closest to the maximum margin hyperplane. The rest of the training examples are assumed to be trivial for determining the binary class boundaries [42]. Kim [42] makes one of the first attempts to use SVM in stock markets in 2003. He used technical analysis indicators as predictive variables and SVM to classify the Korean stock market index. He then performs a comparison between his method and case-based reasoning. Similarly, Huang et al. [49] use SVM to classify the market direction and compare its performance with the Elman backpropagation neural network, linear discriminant analysis and quadratic discriminant analysis. Both these researchers claim to have better accuracy compared to other methods. In 2017, X. Qian [50] conducted a comparative analysis between support vector machines, a simple ARIMA, a multi-layer perceptron to predict financial stock market data. He claimed to receive a slightly higher accuracy of SVM models over the other models.

4.2.4 Random forest

Random forest [51], is a type of ensemble learning method utilized for classification and regression analysis. By generating a multitude of decision trees and combining each of their predictive value, random forest aims to obtain a classifier with high precision and stability. Developing multiple decision trees through bagging and learning by picking variables randomly for each tree makes this method more advantageous than a single decision tree. Considering random forest is more robust towards noise and outliers as well as it overcomes the disadvantage of overfitting that occurs in the learning process for a single decision tree [52]. Researchers typically aim to minimize prediction error in the stock market literature by treating the forecasting problem as a classification one. For instance, in [53], researchers use some technical indicators such as Relative Strength Index (RSI), stochastic oscillator, and some other indicators as inputs to train their random forest and to predict the returns of a stock.

4.2.5 Artificial neural networks

Artificial Neural Networks (ANN) are a nonlinear modeling approach that gives a reasonably accurate universal approximation to any functions. Furthermore, their learning ability makes neural networks very versatile for generalization and prediction tasks. ANN incorporates a set of threshold functions. Adaptive weights form a connection among these functions, and afterwards, they can be trained based on historical data [54]. In the case of time series forecasting tasks, the prediction model

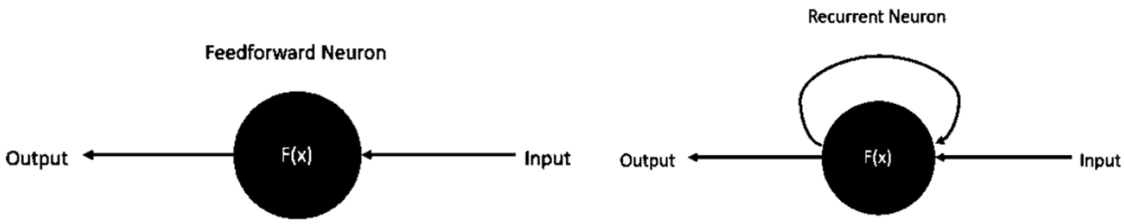


Figure 1: Recurrent and feed-forward neurons [57]

has the general form as follows:

$$X_t = \mathbf{f}(X_{t-1}, X_{t-2}, X_{t-3}, \cdots, X_{t-n}) + e_t$$

meaning that X_t is a function of its past values and error term e_t [55]. Among different neural networks, related works propose recurrent neural networks as the most suitable models for predictions and time series forecasting. A recurrent neural network is a specific type of artificial neural network where connections between neurons create a directed cycle. Recurrent neural networks are suitable for sequential data due to their features. Therefore, it makes them a great candidate model for predicting financial assets, such as daily stock market prices. The fundamental difference between them and a simple artificial neutral network (feed-forward) network is in their nodes connections. In addition to the connection from input neurons to output, there exists an extra connection from output to input. This additional connection is utilized to store information over time, leading to a network with dynamic memory [56]. Figure 1 illustrates the difference between recurrent and feed-forward neurons.

Additionally, in recurrent neural networks, the network current state is a function of its previous states, enabling them to use past information in order to handle and predict sequential data. Another feature of them is that they can be remarkably deep. This feature can trigger a drawback during training a neural network because they are more susceptible to the vanishing gradient problem [58].

To address this problem in 1997 [58], a specific type of recurrent neural networks was introduced. The long short-term memory neural network can learn the long dependency among data and cope with the vanishing gradient problems that can occur when learning through traditional recurrent neural networks [52]. Long short-term memory neural network solves this problem by introducing a concept known as a gate. There is an input, output, and forget gate. The input gate controls which portion of a new value should flow into the cell and be added. The forget gate is responsible for deciding the information that the network should forget for the new state by multiplying 0 to a position in the matrix. Furthermore, the output gate determines which portion of the value should be applied to compute the output activation of the network unit or, in other words, what will be the following output. This way, information flows in recurrent neural networks can be controlled. These gates and a memory cell allow networks to forget, memorize, and reveal data in a flexible manner. This way, long short-term memory neural network is intended to remember long sequences [57]. Hence, it is suitable for time series prediction, and consequently, it opened its space in financial time series prediction, as well [59, 60].

In 2016, S. McNally[61] applied a bayesian optimized recurrent neural network and a long short-term memory neural network to predict the direction in the price of Bitcoin. He claimed a better accuracy and a lower regression error while using the long short-term memory neural network. In 2018, Fischer and Krauss employed the long short-term memory network on some major stocks of the $S\&P$ 500 from 1992 to 2015 to predict their direction [62]. Bao et al. [63] proposed an ensemble approach. They apply wavelet transforms to remove the noise from stock data, stacked autoencoders to extract high-level features for forecasting, and finally, long short-term memory neural network to predict stock price indices. They reported the long short-term memory neural network to provide high accuracy in these studies.

However, because of its lack of transparency and complex architecture, it may not be the most efficient network structure, especially in the stock market which is too ambiguous and clarity is therefore essential.

4.3 Multivariate time series analysis in finance

Time series analysis in theoretical and practical aspects is an inseparable part of studying stock markets. Empirical research started in 1933 [64], focusing on analyzing the stock market as a single independent time series, often referred to as univariate time series analysis. In the univariate analysis, the financial time series consists of single observations recorded sequentially over equal time increments. However, within the last 50 years, worldwide economies became more related to each other. Therefore, various phenomena such as politics, technology, and even pandemics can influence a set of financial time series in a similar manner. Nowadays, there are likely to be groups of stocks that follow similar time-based behaviour patterns simultaneously or with some time delay. This covariation of time series is a source of information that can improve forecast accuracy in certain situations [65].

When speaking about multivariate time series analysis in finance, two main models are the most popular ones: the *vector autoregressive* model and the *vector error correction* model. Their focus is to model the relational dependency among economic variables in a nonstructural manner. The vector autoregressive model was introduced in 1980 by Sims [66] for the economic field, becoming a widespread application in dynamic financial system analysis. This model assumes that each endogenous variable includes the lagged value of all endogenous variables in the system; thus, it extends the univariate autoregressive model to the "vector" autoregressive model consisting of k multivariate time series. In other words, the vector autoregressive model regress a vector of time series variables on lagged vectors of these variables. As for autoregressive (p) models, the lag order is denoted by p, so the equations give the vector autoregressive (p) model of two variables (k=2) for X_t and Y_t is as follows:

$$
\begin{aligned}
Y_t &= \beta_{10} + \beta_{11} Y_{t-1} + \cdots + \beta_{1p} Y_{t-p} + \gamma_{11} X_{t-1} + \cdots + \gamma_{1p} X_{t-p} + u_{1t}, \\
X_t &= \beta_{20} + \beta_{21} Y_{t-1} + \cdots + \beta_{2p} Y_{t-p} + \gamma_{21} X_{t-1} + \cdots + \gamma_{2p} X_{t-p} + u_{2t}.
\end{aligned}
$$

The β and γ values can be estimated using ordinary least squares on each equation [67]. It is important to note that, vector auto regressive model consists of more than one equation and it assumes that each time series must be stationary. In practice, stationarity is achieved by first differencing the time series, which may result in the potential loss of information about the relationship of the time series at the level. Engle and Granger [68] combined cointegration and error correction models to establish the trace error correction model to solve this problem. If there is a cointegration relationship between time series at level, the error correction model can be derived from the autoregressive distributed lag model. In other words, the vector error correction model is a vector autoregressive model with cointegration constraints that combines levels and differences.

Multivariate time series analysis and forecasting are not bounded to these two models. Because of the significant importance of multivariate time series analysis many models have emerged. They intend to incorporate various interconnection analysis approaches into predictions. It is, however, much more complicated than univariate time series analysis, particularly when dealing with a large number of series[69]. One way to analyze the interconnection among multiple time series is by assessing their similarity. Devising a proper measurement to detect similar behaviour among time series is a trivial task [70]. For example, in stock markets; finding the answer to the following questions is vital.

a) What are all the stocks that behave “similar” to stock A?

b) What are all the stocks that have “heads and shoulders pattern” for a month?

Besides Euclidean distance measures, many others can be found in the literature [71, 72, 73, 74, 75, 76, 77]. In 1999, Mantegna suggested a methodology known as the standard methodology that has been adopted and followed by many researchers in different areas. As of 2021, his book [78] has been cited more than 5100 times. To calculate the distance (similarity) among assets, Mantegna recommends using correlation among price differences. In continuation of his work, many researchers aim to improve his suggested method concerning the clustering algorithm or the distance measure itself, briefly described in the following. Let N be the number of assets, $P_i(t)$ be the price at time t of asset i, 1≤i≤N, then the logarithm of price difference $r_i(t)$, is calculated as:

$$r_i(t) = \log P_i(t) - \log P_i(t-1).$$

To determine the distance (similarity) between each pair i, j of assets, he suggests computing their correlation as follows:

$$\rho_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\left(\langle r_i^2 \rangle - \langle r_i \rangle^2\right)\left(\langle r_j^2 \rangle - \langle r_j \rangle^2\right)}}.$$

Having the correlation, it is possible to convert the correlation coefficients into distance using the following equation:

$$d_{i,j} = \sqrt{2(1 - p_{i,j})}.$$

Afterwards, he employs a minimum spanning tree to cluster the most similar assets in the form of a tree. A distinctive indexed hierarchy is another product of the resulting tree, corresponding to the one given by the dendrogram obtained using the single linkage clustering algorithm. However, there are few concerns regarding this standard methodology. The biggest problem is instability, which can be partially caused by the minimum spanning tree algorithm or using the correlation coefficient. It is known that Pearson linear correlation is sensitive towards outliers and generally not suitable for other distributions except Gaussian ones. Also, it is difficult to interpret the changes of linkage during the time since there is a high level of statistical uncertainty associated with the correlation estimation [79]. Interestingly, as one might expect, the higher the correlation coefficient between an asset pair, the more reliable their link should be, but in a study in [80], research shows that this hypothesis is not always satisfied in practice. Mantegna [81] concludes that a better approach is needed to use a distance measure other than the expected squared deviation and is distribution-free. Thus, researchers have investigated different distant measures with specific viewpoints.

Besides similarity analysis, there are other methods to evaluate the linkage among multiple time series. In 2017, Gautier Marti et al. reviewed over a two-decade of data mining tasks in finance (-see [79]). Some studies focus on the lead-lag effect among assets to evaluate if one financial instrument provides relevant information about another one. These quantifying linear relationships methods are as following:

- the Granger causality analysis [82]
- correlation/partial correlation analysis [83]
- asynchronous, lead-lag relationships analysis using mutual information instead of Pearson's correlation coefficient [84, 85] ,
- normalized correlation matrix using the affinity transformation.

Other researches, focused on nonlinear analysis methods to evaluate the connection between financial time series such as:

- the Brownian distance [86],
- mutual information, mutual information rate, and other information-theoretic networks,
- tail dependence distances [87] and copula-based [88].

These methods have been used in many fields, including finance. For instance, Mok [89] applied the autoregressive integrated moving average and Granger causality analysis to investigate the connection among the daily stock prices, exchange rates, and interest rates of Hong Kong. Based on the results, there were occasional unidirectional Granger causality from stock price to the interest rate as well as a weak bidirectional Granger causality between exchange rate and stock price.

In another study, Marschinski and Kantz [90] applied effective transfer entropy to investigate the linking between the Dow Jones Industrial Average from the United States and the German Stock Index. They discovered that the Dow Jones stock market transfer more information to the German market than they receive from it. The result proves the asymmetric information flow between them. This study does not append the research to prediction.

Kwon and Yang [91] investigate the relationship between Twenty-five global financial market indices using a transfer entropy analysis. They provided two main findings. Firstly, the US stock markets have the most influence on the global stock market, and secondly, the Asian and Pacific markets received utmost information. It is worth noting that none of these studies expand their results to forecasting financial time series.

Another category of researchers does not evaluate the dependency among variables, and they apply multivariate forecasting based on a preexisting assumption. For instance, in 2018, Torres et al. [57] use recurrent neural networks for multivariate financial time series. To forecast bitcoin's price, he utilizes Google trends as an indicator of interest over the keyword "Bitcoin". He assumes that the Bitcoin google trend is related to its price value. One of the methods to quantify the relationships among multiple time series is by estimating their similarity. Note that this similarity measurement is approximate. In [92] researchers used closing price returns together with trading volumes as two dependant variables to obtain a structure for the US stock market. Alternatively, another example can be using low, high, open, and close prices together [93] to define the structure between multivariate time series.

Kim et al. [52] used efficient transfer entropy and a variety of machine learning algorithms to forecast the course of US stock prices. First, they examine the association between financial crises and Granger-causal relationships among stocks to see whether the efficient transfer entropy can be regarded as a market explanatory variable. Later, they include effective transfer entropy as a new function in machine learning algorithms such as random forest and long short-term memory networks. They conclude that incorporating the efficient transfer entropy-driven variable as a new function improves prediction efficiency on stock price direction. Using Granger-causal relationship between financial news data and the Korean market, Nam and Seong [94] applied various machine learning algorithms to predict stock market price direction. In a recent study in 2020, Sidi [95] uses similarity measurements to improve the prediction of the stocks direction. Using random forest and gradient boosting Trees, he handles the prediction problem as a classification one where he classifies the trend values as increase or decrease. He tested different similarity measurement methods such as Euclidean distance, Pearson correlation, Johansen cointegration, etc., to find the similarity among seven S&P stocks. His primary focus is to train the models on similar stocks to determine if utilizing similarity can improve the prediction accuracy. His results show that the prediction model that was trained on similar stocks had significantly better results.

Table 1 summarizes the methodology of previous studies on multivariate time series analysis. In general, we summarized earlier studies into three categories. The first category contains those researches investigating the interrelation among

Table 1: Outline of the previous literature on multivariate time series

Reference	**Analysis method**	**Prediction**	**Data**
Mok	ARIMA and Granger causality analysis	No	Daily stock prices, exchange rates and interest rates of Hong Kong.
Marschinski and Kantz	Effective transfer entropy	No	Dow Jones Industrial Average from the United States and the German Stock Index
Kwon and Yang	Transfer entropy	No	25 Global financial markets
Torres et al	Recurrent neural network	Yes	Bitcoin and google trend
Juan and Risso	Minimal Spanning and Hierarchical trees	No	Return values and traded volume
Lee and Djauhari	Minimal spanning tree	No	Malaysian stock exchange
Kim et al.	Effective Transfer Entropy and machine learning algorithms	Yes	US stock markets
Sidi	Random forest and gradient boosting	Yes	US stock market
Nam and Seong	Granger causality analysis and machine learning algorithms	Yes	Financial news data and the Korean market

stocks without extending the analysis further into prediction. The second category assumes linkage between financial time series without analysis, and their focus is only on improving the forecast. The third category has analyzed the multivariate relation as well as prediction.

In this thesis, we propose two powerful methods to evaluate the relation among financial time series, and we extend this to forecasting. The suggested methods are based on the classical decomposition approach and two main fuzzy theories namely F-transform and fuzzy natural logic. Therefore, in the following section, we will provide a mathematical definition of necessary concepts, the decomposition model, and fundamental theories.

5 Preliminaries

In this section, we briefly review the mathematical preliminaries related to the subject of the thesis.

5.1 Basic concepts of stochastic processes

In order to delineate the time series in detail, we must first define some basic statistical concepts. Therefore, in this thesis **E** and **Var** will denote the expectation and the variance of a random variable, respectively. Moreover, for covariance and the correlation of two random variables, we use **Cov** and **Cor**, respectively. Note that the variance, covariance, and correlation can all be calculated using the expectation as follows. On the probability space $(\Omega, \mathscr{F}, P)$ let X, Y be two complex-valued random variables. Then

$$\begin{aligned}
\mathbf{E}(X) &= \sum_x xP(X = x), \\
\mathbf{Var}(X) &= \mathbf{E}\,|X - \mathbf{E}(X)|^2, \\
\mathbf{Cov}(X, Y) &= \mathbf{E}\left[(X - \mathbf{E}(X)) \cdot \overline{(Y - \mathbf{E}(Y))}\right], \\
\mathbf{Cor}(X, Y) &= \mathbf{E}(X\overline{Y}).
\end{aligned}$$

Consequently, $\mathbf{Var}(X) = \mathbf{Cov}(X, X)$, $\mathbf{Cov}(X, Y) = \overline{\mathbf{Cov}(Y, X)}$, $\mathbf{Cor}(X, Y) = \overline{\mathbf{Cor}(Y, X)}$ and $\mathbf{Cov}(X, Y) = \mathbf{Cor}(X, Y) - \mathbf{E}X \cdot \overline{\mathbf{E}Y}$.

5.1.1 Stochastic processes

Let $T \subseteq \mathbb{R}$ be an index set and $(\Omega, \mathcal{F}, P)$ a probability space. A real valued stochastic process (or random process) is a family $\{\xi(t, \omega) \mid t \in T\}$ where for each $t \in T$, $\xi(t, \omega) \in \mathbb{R}$ is a real valued random variable. The *covariance function* Υ and the *correlation function* Γ of a stochastic process $\xi(t)$, $t \in \mathbb{R}$ are bivariate functions that are sequentially described as follows:

$$\begin{aligned}
\Upsilon(t, s) &= \mathbf{Cov}(\xi(t), \xi(s)), \\
\Gamma(t, s) &= \mathbf{Cor}\left(\xi(t), \xi(s)\right),
\end{aligned}$$

for any $t, s \in \mathbb{R}$. It is easy to see that $\Gamma(t, s) = \overline{\Gamma(s, t)}$, $\Upsilon(t, s) = \overline{\Upsilon(s, t)}$ and $\Upsilon(t, s) = \Gamma(t, s) - \mathbf{E}(\xi(t)) \cdot \overline{\mathbf{E}(\xi(s))}$.

5.1.2 Stationary processes

The so-called, stationary processes are a particular class of stochastic processes that play a key role in time series analysis. Generally, there are two standard categories of stationarities naming *weak stationarity* and *strict stationarity*. Commonly, in the literature, instead of "weak" stationary time series, researchers refer to stationary time series, which is also the case in this thesis. Yaglom at 1962, defined that *a stationary process* is a specific class of a stochastic process [96].

Definition 1 *A stochastic process $\xi(t)$ is said to be* a stationary process *if for any $t \in \mathbb{R}$, the following statements are satisfied:*

(*i*) $\mathbf{E}(\xi^2(t)) < \infty$,

(*ii*) $\mathbf{E}(\xi(t))$ *is constant and independent on t,*

(*iii*) $\mathbf{Cor}(\xi(t), \xi(t+\tau))$ *is independent of t for each τ.*

The class of *a stationary process* can be divided into several significant subclasses. These processes are fundamental in time series analysis; therefore, we briefly introduce them in the following.

- *White noise process:* a white noise, denoted by $\xi(t) \sim WN(0, \sigma^2)$, is a stochastic process $\xi(t)$ that holds the following conditions:

$$\mathbf{E}(\xi(t)) = 0, \mathbf{Var}(\xi(t)) = \sigma^2, \text{ and } \mathbf{Cor}(\xi(t), \xi(s)) = 0 \quad \text{for all} \quad t \neq s.$$

- AR *and* MA *processes: A stochastic process $\xi(t)$ is called an* AR *(autoregressive) process of order $\rho \geq 1$ (or* MA *(moving average) process of the order $\varrho \geq 1$), indicated by* AR(ρ) *(or* MA(ϱ)*), if calculated by* ARMA($\rho, 0$)*, i.e., $\xi(t) = \varphi + \varphi_1 \xi(t-1) + \cdots + \varphi_\rho \xi(t-\rho) + \varepsilon(t)$ (or* ARMA($0, \varrho$)*, i.e., $\xi(t) = \mu + \varepsilon(t) + \phi_1 \varepsilon(t-1) + \cdots + \phi_\varrho \varepsilon(t-\varrho)$).*

- *Autoregressive and moving average process* (ARMA): an ARMA process of order $\rho, \varrho \geq 1$, denoted by $\xi(t) \sim$ ARMA(ρ, ϱ), is a stochastic process $\xi(t)$ with the following form:

$$\xi(t) = \varphi + \varphi_1 \xi(t-1) + \cdots + \varphi_\rho \xi(t-\rho) + \varepsilon(t) + \phi_1 \varepsilon(t-1) + \cdots + \phi_\varrho \varepsilon(t-\varrho)$$

where $\varepsilon(t) \sim WN(0, \sigma^2)$ and

$$\varphi = \left(1 - \sum_{j=1}^{\rho} \varphi_j\right) \mu$$

with μ indicating the process mean.

5.2 Time series

Many phenomena accrued within time order. A time-ordered sequence of observations is called a time series. Examples of such data sets are diverse, for instance, daily closing stock prices, monthly unemployment rates, quarterly crime rates, and annual death rates [97]. In this thesis, we suppose that a time series X is a realization of the stochastic process ξ, i.e.,

$$X(t) = \xi(t, \omega), \quad t \in T,$$

for some fixed elementary event $\omega \in \Omega$. We usually assume that T is a finite set of natural numbers $T = \{1, \dots, n\}$.

5.2.1 Decomposition models

As specified in the previous subsection, let $X(t)$ be a time series with three critical components. Traditionally, there are two approaches to model $X(t)$ from its components. The first approach called the *additive decomposition model*, and as the name suggests aims to model $X(t)$ as a sum of the trend-cycle, the seasonal component, and the noise:

$$X(t) = TC(t) + S(t) + R(t). \tag{1}$$

Time series that the magnitudes of their seasonal fluctuations do not vary clearly during the time domain usually suites this model better. The other approach called *multiplicative decomposition model* model the $X(t)$ as a product of its components, namely,

$$X(t) = TC(t) \cdot S(t) \cdot R(t).$$

This model is suitable for time series whose magnitudes of seasonality clearly vary by the time moments. Nevertheless, it can be easily transformed into an additive one by applying the logarithmic transformation, i.e.,

$$\ln X(t) = TC^\star(t) + S^\star(t) + R^\star(t).$$

In this thesis, we limit our analysis to the time series $X(t)$ that can be decomposed into the additive decomposition model (1).

5.3 Fuzzy transform

Let us remind that our ultimate goal here is to achieve a linguistic relation that estimates the predictional relation among multiple financial time series. Typically fuzzy approximation methods are used to estimate the interesting relations among data. In this approach, the primary assumption is that we can provide an approximation of the relations for various reasons, such as insufficient available information. The essential fuzzy approximation methods are fuzzy IF-THEN rules under relational interpretation, Takagi–Sugeno rules, and the F-transform [98]. In this section, we will overview the F-transform which is the fundamental fuzzy modeling method used in time series analysis.

5.3.1 Basic principles of fuzzy transform

Let U be a set (a universe). By a fuzzy set A in U we understand a function $A : U \longrightarrow [0, 1]$. The element $A(x) \in [0, 1]$ is called *membership degree* of $x \in U$ in the fuzzy set A.

The operations with fuzzy sets are defined using the operations in the algebra of truth values (cf. [99, 100]). A binary fuzzy relation is a fuzzy set $R : U \times V \longrightarrow [0, 1]$. If R is a fuzzy relation and $u \in U, v \in V$ then we often write the membership degree of u, v in R as $(uRv) \in [0, 1]$.

The F-transform is an approximation procedure applied, in general, to a bounded real continuous function $f : [a, b] \longrightarrow [c, d]$ where $a, b, c, d \in \mathbb{R}$. It is based on the concept of a *fuzzy partition* that is a set $\mathcal{A} = \{A_0, \ldots, A_n\}$, $n \geq 2$, of fuzzy sets fulfilling special axioms.

The F-transform has two phases: direct and inverse. The *direct* F-transform assigns to each $A_k \in \mathcal{A}$ a component $F_k[f|\mathcal{A}]$. We distinguish *zero degree* F-transform whose components $F_k^0[f|\mathcal{A}]$ are numbers and *higher degree**) F-transform whose components have the form $F_k^1[f|\mathcal{A}](x) = \beta_k^0[f] + \beta_k^1[f](x - c_k)$. 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 $A_k \in \mathcal{A}$.

From the direct F-transform of f

$$\mathbf{F}[f|\mathcal{A}] = (F_0[f|\mathcal{A}], \ldots, F_n[f|\mathcal{A}])$$

we can form a function $\hat{f} : [a, b] \longrightarrow [c, d]$ using the formula $\hat{f}(x) = \sum_{k=0}^{n}(F_k[f|\mathcal{A}] \cdot A_k(x))$, $x \in [a, b]$. The function $\hat{f}$ is called the *inverse F-transform* of f and it approximates the original function f. It can be proved that this approximation is universal. The higher degree one (F^m-transform, $m \in \mathbb{N}$) was introduced in [101]. In the following, we explain the F-transform in detail.

5.3.2 Fuzzy partition

The first step of the F-transform procedure is to form a *fuzzy partition* of the domain $[a, b]$. It consists of a finite set of fuzzy sets

$$\mathcal{A} = \{A_0, \ldots, A_n\}, \qquad n \geq 2, \tag{2}$$

defined over nodes $c_k \in [a, b]$, $k = 0, \ldots, n$, such that

$$a = c_0, \ldots, c_n = b. \tag{3}$$

The properties of the fuzzy sets from $\mathcal{A}$ are characterized by five axioms, namely: *normality, locality (bounded support), continuity, unimodality, and orthogonality.* For their precise formulation (-see [102, 103]).

Each k-th fuzzy set A_k, $k = 1, \ldots, n-1$, has the support (c_{k-1}, c_{k+1}) and $A_k(c_k) = 1$. Otherwise, $A_k(x) = 0$ for all $x \in [a, c_{k-1}] \cup c_{k+1}, b]$.

A fuzzy partition $\mathcal{A}$ is called *h-uniform* if the nodes $c_0, \ldots, c_n$ are *h-equidistant*, i.e., for all $k = 0, \ldots, n-1$, $c_{k+1} = c_k + h$, where $h = (b - a)/n$ and the fuzzy sets $A_1, \ldots, A_{n-1}$ are shifted copies of a *generating function* $A : [-1, 1] \longrightarrow [0, 1]$ such that for all $k = 1, \ldots, n-1$

$$A_k(x) = A\left(\frac{x - c_k}{h}\right), \qquad x \in [c_{k-1}, c_{k+1}].$$

The fuzzy sets A_n have the support (c_{n-1}, c_n)†) and we often call them *basic functions.* Note that $A_1, \ldots, A_{n-1}$ cover the whole domain (a, b), i.e., $(a, b) = \bigcup_{k=1}^{n-1} \mathrm{Supp}(A_k)$

*) In general, higher degree F-transform.

†) We formally put $c_{-1} = c_0 = a$ and $c_{n+1} = c_n = b$.

(since $A_1(a) = A_{n-1}(b) = 0$). For $k = 0$ and $k = n$ we consider only halves of the functions A_k, i.e., A_0 has the support (c_0, c_1) and A_n the support (c_{n-1}, c_n).

The orthogonality is formally defined as[‡)]

$$\sum_{i=0}^{n} A_i(x) = 1, \qquad x \in [a, b]. \tag{4}$$

5.3.3 Zero degree F-transform

Once the fuzzy partition $A_0, \ldots, A_n \in \mathcal{A}$ is determined, we define a *direct F-transform* of a continuous function f as a vector $\mathbf{F}[f] = (F_0[f], \ldots, F_n[f])$, where each k-th *component* $F_k[f]$ is equal to

$$F_k[f] = \frac{\int_a^b f(x) A_k(x)\, dx}{\int_a^b A_k(x)\, dx}, \qquad k = 0, \ldots, n. \tag{5}$$

Clearly, the $F_k[f]$ component is a *weighted average* of the functional values $f(x)$ where weights are the membership degrees $A_k(x)$. The *inverse F-transform* of f with respect to $\mathbf{F}[f]$ is a continuous function $\hat{f} : [a, b] \longrightarrow \mathbb{R}$ such that

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

Theorem 5.1 *The inverse F-transform $\hat{f}$ has the following properties:*

(a) It coincides with f if the latter is a constant function.

(b) The sequence of inverse F-transforms $\{\hat{f}_n\}$ determined by a sequence of uniform fuzzy partitions based on uniformly distributed nodes with $h = (b - a)/n$ uniformly converges to f for $n \to \infty$.

(c) The F-transform is linear, i.e., if $f = \alpha u + \beta v$ then $\hat{f} = \alpha \hat{u} + \beta \hat{v}$.

(d) Let $f(x) = q \in \mathbb{R}$ for all $x \in [a, b]$. Then $F_k[f] = q$ for any $k = 1, \ldots, n - 1$.

All the details and full proofs can be found in [102, 104].

5.3.4 Higher degree F-transform

The F-transform introduced above is F^0-transform (i.e., zero-degree F-transform). Its components are real numbers. If we replace them with polynomials of arbitrary degree $m \geq 0$, we arrive at the higher degree F^m transform. This generalization and its proof have been in detail described in [104]. Let us remark that the F^1 transform

‡) Equation (4) is sometimes called *Ruspini condition*.

enables us to estimate derivatives of the given function f as a weighted average over a specified area.

The direct F^1-*transform* of f with respect to $\mathcal{A}$ is a vector $\mathbf{F}^1[f] = (F_1^1[f](x), \ldots, F_{n-1}^1[f](x))$ where the components $F_k^1[f](x)$, $k = 0, \ldots, n$ are linear functions

$$F_k^1[f](x) = \beta_k^0 + \beta_k^1(x - c_k) \tag{6}$$

with the coefficients β_k^0, β_k^1 given by

$$\beta_k^0 = \frac{\int_{c_{k-1}}^{c_{k+1}} f(x)A_k(x)dx}{\int_{c_{k-1}}^{c_{k+1}} A_k(x)dx}, \tag{7}$$

$$\beta_k^1 = \frac{\int_{c_{k-1}}^{c_{k+1}} f(x)(x - c_k)A_k(x)dx}{\int_{c_{k-1}}^{c_{k+1}} (x - c_k)^2 A_k(x)dx}. \tag{8}$$

Note that $\beta_k^0 = F_k[f]$, i.e. the coefficients β_k^0 are just the components of the F^0 transform given in (5). The F^1 transform has also the properties stated in Theorem 5.1 (-see [104]). In comparison with F^0 transform, it is more precise. Let us remark that in general, we can define n-th degree F-transform. Of course, for higher n it is more complex but also more precise. In practice, it is sufficient to consider only $n \in \{0, 1, 2\}$.

The following theorem is important for mining information from time series (cf. [103]).

Theorem 5.2 *If f is four-times continuously differentiable on $[a, b]$ then for each $k = 1, \ldots, n-1$,*

$$\beta_k^0 = f(c_k) + O(h^2), \tag{9}$$

$$\beta_k^1 = f'(c_k) + O(h^2). \tag{10}$$

Thus, each F^1-transform component provides a weighted average of values of the function f in the area around the node c_k (9), and also a weighted average of slopes (33) of f in the same area.

Lemma 5.3 *If $\mathcal{A}$ is an h-uniform fuzzy partition and each basic function $A_k \in \mathcal{A}$ has a triangular shape then the following can be proved:*

$$\int_{c_{k-1}}^{c_{k+1}} A_k(x)dx = h, \tag{11}$$

$$\beta_k^0 = \frac{1}{h}\int_{c_{k-1}}^{c_{k+1}} f(x)A_k(x)dx, \tag{12}$$

$$\beta_k^1 = \frac{12}{h^3}\int_{c_{k-1}}^{c_{k+1}} f(x)(x - c_k)A_k(x)dx. \tag{13}$$

Lemma 5.4 *Let $\mathcal{A}$ be an h-uniform fuzzy partition with triangular membership functions and let $|f(x)| \leq 1$ for all $x \in [a, b]$. Then* *(a)* $|\beta_k^0| \leq 1$.

(b) If $h \geq 12$ *then* $|\beta_k^1| \leq 1$.

Similar to F^0-transform, the inverse F-transform of higher degree is a continuous function, and using (6), F^1-transform can be defined as follows:

$$\hat{f}^1(x) = \sum_{k=0}^{n} F_k^1[f](x) \cdot A_k(x), \qquad x \in [a, b].$$

5.4 Main concepts of fuzzy natural logic

Fuzzy Natural Logic (FNL) is a group of mathematical theories that extend mathematical fuzzy logic in a narrow sense. In general, this mathematical theory aims to model terms and rules that come with natural language and thus, it provides a model of natural human reasoning. FNL has an essential feature, which is its ability to cope with the vagueness of the semantics of the natural language. In this section, we very briefly remind some of the concepts of FNL used in this thesis. Many details can be found in the book [98].

The components of FNL presented here are the theory of evaluative linguistic expressions, the theory of fuzzy/linguistic IF-THEN rules, and logical inference based on them.

5.4.1 Evaluative linguistic expressions

We call back some fundamental notions from the theory of evaluative linguistic expressions developed by Novák in [105]. *Evaluative linguistic expressions* are expressions of natural language that are used in everyday decision-making processes and various kinds of evaluation. They facilitate the evaluation of phenomena happening around the world (e.g., the course of development of some process, the manifestations of some property) in an interpretable and understandable form. Here, we provide a simplified definition of evaluative expressions that is sufficient for our applications in time series data mining in a subsequent chapter. For more elaborated explanations, please refer to [98]. The general form of evaluative linguistic expressions is the following syntactic structure:

$$\langle \text{linguistic hedge} \rangle \langle \text{atomic evaluative expression} \rangle, \tag{14}$$

where *atomic evaluative expression* is one of the canonical adjectives *small, medium, big (Sm, Me, Bi* for short). However, if the context of the problem is well-defined, we can substitute the vast majority of evaluative adjectives with their equivalent trichotomy *(e. g. young, middle-aged, old* etc.). Moreover, it is possible to make the meaning of evaluative expressions more or less specific by some adverb(e. g. extremely, very, more or less, roughly, etc.). These adverbs are called *linguistic hedges* and can have a *narrowing* or *widening* effect. Tables 2 and 3 contain some examples of such hedges.

Hedge	Abbreviation
extremely	Ex
significantly	Si
very	Ve

Table 2: Linguistic hedges with narrowing effect

Hedge	Abbreviation
rather	Ra
more or less	ML
roughly	Ro
very roughly	VR

Table 3: Linguistic hedges with widening effect

In the following analysis, the evaluative expressions are denoted by script letters $\mathscr{A}$, $\mathscr{B}$, etc. The evaluative expressions are used to characterize the values of variables linguistically. We call such descriptions *evaluative linguistic predications* (in short *evaluative predications*) and take the following form:

$$X \text{ is } \mathcal{A} \tag{15}$$

where X is a variable. The semantics of evaluative expressions is characterized with respect to a set of *contexts*. Examples of evaluative predictions are "volatility is very high", "Return of investment is medium", etc. To model the meaning of evaluative predications, one should bear in mind that the meaning of atomic evaluative expressions changes in different *contexts*.

Definition 1 *Let $v_L, v_M, v_R \in \mathbb{R}$ be numbers such that $v_L < v_M < v_R$. Then the context is a strictly increasing bijection*

$$w : [0,1] \to [v_L, v_M] \cup [v_M, v_R]\,,$$

where $w(0) = v_L, w(0.5) = v_M, w(1) = v_R$.

Hence, it is possible to characterize a context by a triplet $w = \langle v_L, v_M, v_R \rangle$, where v_L, v_M and v_R characterize the minimal, typically middle, and maximal values, respectively. For instance the height of people in the context of central Europe can be $\langle 60, 165, 240 \rangle$ centimeters while in other parts of the world while these numbers can be changed to $\langle 50, 150, 220 \rangle$. Note that the middle value does not need to be the average of the left and right values but an arbitrary value between them. The context $\langle 0, 0.5, 1 \rangle$ is called *standard*. In the case of the financial time series analysis, the concept of context is very vital. When speaking about the values of a variable that stores the return of investment of specific stock, various traders might define the context uniquely. If an investor takes a higher risk, a possible context can be $\langle 0.6,\ 0.76,\ 1 \rangle$, but investors with low tolerance of risk might define the context as $\langle 0.2,\ 0.36,\ 0.6 \rangle$. Note that the middle value in the context does not have to be the average of the values on the left and right but an arbitrary value between them. This feature is one of the most significant advantages of the suggested methods since it empowers each investor to evaluate and invest based on his or her investing strategy.

Definition 2 *The set of all contexts for evaluative expressions is*

$$W = \{w \mid w \text{ is a context according to Definition 1}\}.$$

Definition 3 *Let w be a context from definition 1, then the extended inverse of w is a function $w^{(-1)} : \mathbb{R} \to [0,1]$ defined as*

$$w^{(-1)}(x) = \begin{cases} w^{-1}(x), & \text{if } x \in [v_L, v_R] \\ 0, & \text{if } x < v_L \\ 1, & \text{if } x > v_R \end{cases}$$

In his original paper, Novák, defines the semantics of evaluative predictions via *intension* and *extension*, e.g. *small* is assigned a function from a set of intervals to a set of fuzzy sets, because the meaning of *small* values changes from context to context.

Definition 4 *Let W be a set of contexts and $\mathcal{A}$ be an evaluative expression. The intension of $\mathcal{A}$ is a function*

$$\mathrm{Int}(\mathcal{A}) : W \to \mathcal{F}(\mathbb{R})$$

which assigns to any context $w \in W$ a fuzzy set of $\mathbb{R}$. This fuzzy set is called extension of evaluative expression $\mathcal{A}$ in the context w, i.e.,

$$\mathrm{Ext}_w(\mathcal{A}) = \mathrm{Int}(\mathcal{A})(w) \lesssim \mathbb{R} \tag{16}$$

Not all functions from W to $\mathcal{F}(\mathbb{R})$ are appropriate as intensions as they do not capture the semantics of evaluative expressions adequately. We will show an example of a parametrized family of intensions in the next section. We will start with the definition of *context* and how it applies to the concept of a *horizon*. Each evaluative expression context is defined by an ordered scale with two limit points: left and right bounds, (v_L to v_R). The *horizon* is a sharp line that describes a region of the world that lies ahead of it imprecisely. It can also be moved "along the world" to get closer or further. The only thing we know is that small values run from v_L to a certain horizon of small values, after which they are unquestionably not small. This rationale extends to both big and medium values. Although with medium values, horizons on both sides of the most common value, v_M must be considered. In the following a simple mathematical model for the *horizon* is introduced.

Definition 5 *Let w be a context. Then the horizons are defined as follows*

$$\begin{gathered} LH_w(x) = \tfrac{v_M - x}{v_M - v_L} \text{ for } x \in [v_L, v_M], \text{ else } LH_w(x) = 0 \\ RH_w(x) = \tfrac{x - v_M}{v_R - v_M} \text{ for } x \in [v_M, v_R], \text{ else } RH_w(x) = 0 \\ MH_w(x) = \neg LH_w(x) \otimes_m \neg RH_w(x) \end{gathered}$$

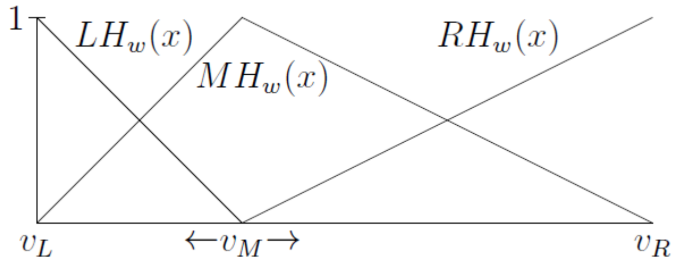


Figure 2: Functions representing the three horizons in general context $w = \langle v_L, v_M, v_R \rangle$.

The *horizons* are denoted as LH, RH, and MH when context w is standard context $\langle 0, 0.5, 1 \rangle$. Figure 2 represents an example of horizon functions.

Another important question to answer is how to mathematically model *linguistic hedges*. In [106] Zadeh proposed modeling of hedges by using certain operations that allow applying simple hedges for computing extensions of more complex linguistic expressions. For example, he proposed the function x^2 for a linguistic hedge *very*. This proposal was later questioned by G.Lakoff [107] where he pointed out that depending on the kind of linguistic hedge, the kernel of membership functions should be modified narrower or wider. Overall, linguistic hedges with a narrowing effect should have a steeper slope than linguistic hedges with a widening effect. As a result, we can model hedges mathematically as a class of non-decreasing functions $\nu_{a,b,c} : [0, 1] \to [0, 1]$, with three parameters $a, b, c \in [0, 1]$, such that $a < b < c$ and the functions, satisfy the three conditions below.

1. $\nu_{a,b,c}(x) = 0$ for all $x \leq a$

2. $\nu_{a,b,c}(b) = b$

3. $\nu_{a,b,c}(x) = 1$ for all $x \geq c$.

The functions $\nu_{a,b,c}$ allow for the horizon to be shifted (and/or deformed). Thus, by assigning specific a, b, c values in the function $\nu_{a,b,c}$, each *linguistic hedges* will be provided. A possible description of the functions $\nu_{a,b,c}$ is as follows:

$$\nu_{a,b,c}(x) = \begin{cases} 1, & c \leq x \\ 1 - \frac{(c-x)^2}{(c-b)(c-a)}, & b \leq x < c, \\ \frac{(x-a)^2}{(b-a)(c-a)}, & a \leq x < b, \\ 0, & x < a \end{cases}$$

Pure evaluative expressions are abstract in the sense that they do not describe the dimensions of concrete objects. As a result, considering standard context $w = \langle 0, 0.5, 1 \rangle$ is adequate. As a result, we can describe their intentions by their extensions.

$$\mathrm{Int}(\ \langle \text{ linguistic hedge } \rangle Sm)(x) = \nu_{a,b,c}(LH(x)),$$
$$\mathrm{Int}(\langle \text{ linguistic hedge } \rangle Me)(x) = \nu_{a,b,c}(MH(x)),$$
$$\mathrm{Int}(\langle \text{ linguistic hedge } \rangle Bi)(x) = \nu_{a,b,c}(RH(x)),$$

where, $x \in [0,1]$ and $\nu_{a,b,c}$ is the function assigned to a linguistic hedge. Remember that an evaluative prediction is formed as X is $\mathcal{A}$, with A being an evaluative expression of the form (14). Given the fact that a context w is a bijection, the extensions of X is $\mathcal{A}$ mentioned in the definition 4 can be written as follows:

$$\begin{aligned} \mathrm{Ext}_w(X \text{ is } \langle \text{ linguistic hedge } \rangle Sm)(x) &= \nu_{a,b,c}\left(LH\left(w^{(-1)}(x)\right)\right), \\ \mathrm{Ext}_w(X \text{ is } \langle \text{ linguistic hedge } \rangle Me)(x) &= \nu_{a,b,c}\left(MH\left(w^{(-1)}(x)\right)\right), \\ \mathrm{Ext}_w(X \text{ is } \langle \text{ inguistic hedge } \rangle Bi)(x) &= \nu_{a,b,c}\left(RH\left(w^{(-1)}(x)\right)\right), \end{aligned} \tag{17}$$

where $x \in w$, $w \in W$, and $w^{(-1)}$ is extended inverse that is described in definition 3.

5.4.2 Linguistic description

The evaluative predications can be used in a conditional clause in the form of fuzzy/linguistic IF-THEN rules. A finite set $\mathrm{LD} = \{\mathscr{R}_1, \ldots, \mathscr{R}_m\}$ of *fuzzy/linguistic IF-THEN rules* is called a *linguistic description* where each $\mathscr{R}_j$ is a particular conditional statement of natural language that describes a relation between variables and have the following form:

$$\mathscr{R}_j := \text{IF } X \text{ is } \mathscr{A}_j \text{ THEN } Y \text{ is } \mathscr{B}_j \tag{18}$$

where $\mathscr{A}_j$ and $\mathscr{B}_j$ are evaluative expressions and $j = 1, \ldots, m$. The linguistic predications "X is $\mathscr{A}_j$" and "Y is $\mathscr{B}_j$" are respectively called *antecedent* and *consequent* of the rule $\mathscr{R}_j$. Any linguistic description can be rewritten as a text consisting of conditional statements to explain some systems behaviour. In our case, it describes nonlinear relationships between different time series in a given data set. Linguistic descriptions are given by the experts or can be extracted from data sets. Novák introduced the principle of a *local perception*, which can be used to mine fuzzy/linguistic IF-THEN rules from a given dataset [108]. We will briefly explain this concept nevertheless, for more details please see [108, 98].

Let us consider a set of evaluative expressions $\{\mathscr{A}_1, \ldots, \mathscr{A}_m\}$ and a context w. Then a *local perception* (LPerc) is an evaluative expression assigned to each given value x in the context w, $x \in w$. More precisely,

$$\mathrm{LPerc}(x, w) \in \{\mathscr{A}_1, \ldots, \mathscr{A}_m\},$$

and is a sharpest (with respect to a special ordering) evaluative expression that characterizes the value x in the context w. For example, let us consider the context $w = \langle 0, 15, 40 \rangle$. Then $\mathrm{LPerc}(1.5, w) =$ *extremely small* or $\mathrm{LPerc}(1.5, w) =$ *very small* depending on the given set of evaluative expressions.

5.4.3 PbLD inference

In this section, we discuss an implicative inference mechanism named Perception based Logical Deduction (PbLD).

Let LD denote a linguistic description and w, w' represent contexts for the variables X and Y, accordingly. Assume that x_0 is a specified observation of variable X, $x_0 \in w$. The PbLD method assigns a proper fuzzy set $\mathscr{A}_i$ to x_0 and deduces a rule $\mathscr{R} \in \mathrm{LD}$ whose antecedent characterizes x_0 in the context w in the best possible way. As a result, a fuzzy set $\mathscr{B}_j$ is proposed as an appropriate consequent for the attribute Y occurring on the right side of IF-THEN rules. Using a defuzzification technique a value y_0, a regression/prediction of the attribute Y is selected that is the most proper value of Y specified by a consequent of $\mathscr{R}$. Formally,

$$r_{\mathrm{PbLD}} : \frac{\mathrm{LPerc}(x_0, w), \mathrm{LD}}{\mathscr{R}, y_0}.$$

Here are three examples of defuzzification operations, which are used in PbLD inference. Assume that $\left\{u_j^{\max} \mid j = 1, \ldots, r_{\max}\right\}$,then

- Mean of maxima (MOM) operation is defined as

$$\mathrm{MOM}(A) = \frac{1}{r_{\max}} \sum_{j=1}^{r_{\max}} u_j^{\max},$$

- First of maxima (FOM) operation is defined as

$$\mathrm{FOM}(A) = \min\left\{u_j^{\max} \mid j = 1, \ldots, r_{\max}\right\},$$

- Last of maxima (LOM) operation is defined as

$$\mathrm{LOM}(A) = \max\left\{u_j^{\max} \mid j = 1, \ldots, r_{\max}\right\},$$

In PbLD, there are three operations used in combination. The defuzzification operation DEE that is performed on the final fuzzy set F is described as follows:

$$\mathrm{DEE}(A) = \begin{cases} \mathrm{LOM}(A) & \text{if } A \text{ is non-increasing} \\ \mathrm{MOM}(A) & \text{if } A \text{ is increasing and decreasing} \\ \mathrm{FOM}(A) & \text{if } A \text{ is non-decreasing.} \end{cases}$$

The predicted/regressed value is obtained by defuzzification. PbLD inference was deployed in R package [109], as well as in a stand-alone program called LFL Controller [110]. See [98] for a more comprehensive explanation and variants.

5.5 Application of the F-transform and FNL to the analysis of time series

In this thesis, we will restrict our consideration to financial data (e.g., exchange rates, gold prices) where the influence of the seasonality is weak, and therefore can be ignored. Particularly, we deal with a time series X that can be additively decomposed into a trend-cycle (deterministic component) and an irregular fluctuation

$$X(t) = TC(t) + R(t), \quad t \in \mathbb{T}, \tag{19}$$

where TC is a smooth function characterized by a F-transform and R is a realization of a stationary random process ζ with zero mean and finite variance.

5.5.1 Analysis of the time series

The application of the F-transform to the time series analysis is based on the following result (cf. [16, 111]). Let us now assume (without loss of generality) that the time series (1) contains periodic subcomponents with frequencies $\lambda_1 < \cdots < \lambda_r$. These frequencies correspond to periodicities

$$T_1 > \cdots > T_r, \tag{20}$$

respectively (via the equality $T = 2\pi/\lambda$).

Theorem 5.5 *Let $\{X(t) \mid t \in \mathbb{T}\}$ be a realization of the time series (1). Let us assume that all the subcomponents with frequencies λ lower than λ_q are contained in the trend-cycle TC. If we construct a fuzzy partition $\mathcal{A}_h$ over the set of equidistant nodes (2) with the distance $h = d\,T_q$ where $d \in \mathbb{N}$ and T_q is a periodicity corresponding to λ_q then the corresponding inverse F-transform $\hat{X}$ of $X(t)$ gives the following estimation of the trend-cycle:*

$$|\hat{X}(t) - TC(t)| \leq 2\omega(h, TC) + D \tag{21}$$

for $t \in \mathbb{T}$, where D for $d \geq 2$ is a certain small *number and $\omega(h, TC)$ is a modulus of continuity of TC w.r.t. h.*

The precise form of D and the detailed proof of this theorem can be found in [16, 112]. The theorem holds both for F^0 as well as for F^1-transform. It implies that the F-transform makes it possible to filter out frequencies higher than a given threshold and also reduces the noise R.

Theorem 5.5 tells us how the distance between nodes of the fuzzy partition should be set to obtain a good estimation of the trend-cycle. Namely, we put it equal to $h = d\,T_q$ where d is in practice equal to 1 or 2. The fuzzy partition is constructed over the interval $[0, p]$ of real numbers, i.e., it is not constructed over the discrete set of natural numbers $\mathbb{T}$. Consequently, we obtain $\hat{X} \approx TC$ which means that we can estimate the trend-cycle (TC) with high fidelity. Of course, the estimation depends on the course of TC and it is the better, the smaller is the modulus of continuity $\omega(h, TC)$. This requirement is usually fulfilled, because the trend-cycle characterizes the general behaviour of the time series which should not bee too volatile.

The periodicity T_q can be found using the well known periodogram (-see [113, 114]). Then, by setting a proper fuzzy partition due to Theorem 5.5, we first compute the F-transform of $X(t)$ (either zero or first degree)

$$\mathbf{F}[X] = (F_0[X](t), \ldots, F_n[X](t))$$

and then compute the inverse $\hat{X}$, which approximates the trend-cycle TC. Hence, in the sequel for a given time series X we will work with estimation $\widetilde{TC}$ of its trend-cycle given by

$$\widetilde{TC} = \hat{X}. \tag{22}$$

The edge components $F_0[X], F_n[X]$ are distorted because of using only half of the corresponding basic functions. This problem can be solved in two ways: either we confine only to the complete components $F_k[X]$ for $k = 1, \ldots, n-1$, or we artificially prolong $\mathbb{T}$ to the left and right by h and extrapolate the corresponding values $X(t_{-i}) = X(t_i)$ for $i = 1, \ldots, h$, and similarly for the right side of $\mathbb{T}$.

5.5.2 Forecasting the time series

Let the components $F_0[X], \ldots, F_n[X]$ be computed from the time series $\{X(t) \mid t \in \mathbb{T}\}$. Our goal is to forecast future values

$$\{X(t) \mid t \in \{p+1, \ldots, p+\ell\}\} \tag{23}$$

where ℓ is a forecasting horizon. Based on the decomposition model (1) we forecast separately trend-cycle (TC) and the seasonal component (S). The resulting forecast is obtained by summing both components.

Forecasting trend or trend-cycle By the combination of the F-transform and FNL, we can forecast the future development of the trend-cycle. For the user, it can also be interesting to know just trend Tr and its forecast. This component is computed separately and is not used in the forecast of the whole time series.

To forecast the components $F_{n+1}[X], \ldots$ corresponding to the future time moments $p+1, \ldots, p+\ell$ (cf. (23)), the perception-based logical deduction can be applied to a linguistic description consisting of the fuzzy/linguistic IF-THEN rules generated by the linguistic learning algorithm. The antecedent variables are the components $F_i[X]$ and also their first- and second-order differences:

$$\Delta F_i[X] = F_i[X] - F_{i-1}[X], \qquad i = 2, \ldots, n-1 \tag{24}$$
$$\Delta^2 F_i[X] = \Delta F_i[X] - \Delta F_{i-1}[X], \; i = 3, \ldots, n-1. \tag{25}$$

The fuzzy/linguistic IF-THEN rules take the form:

$$\text{IF } X_{i-1} \text{ is } \mathcal{A}_{i-1} \text{ AND } X_{i-1} \text{ is } \mathcal{A}_{i-1} \text{ AND } X_i \text{ is } \mathcal{A}_i \text{ THEN } X_{i+1} \text{ is } \mathcal{B}_{i+1} \tag{26}$$

where X_i can be any of the components $F_i[X]$ or their differences (24), (25). The rules (26) characterize basic dynamics of the time series because they describe logical dependencies among changes in the trend-cycle (or trend). The linguistic description consisting of the rules (26) helps to understand functionalities and motive factors determining the changes in a process yielding the times series in question.

On the basis of the forecasted F-transform components we can compute the forecasted trend-cycle of the time series, where the latter consists of the values of the inverse F-transform:

$$\{\hat{X}(p+1), \ldots \hat{X}(p+\ell)\}.$$

6 The suggested similarity methods

As already mentioned, there are many kinds of similarity indexes introduced for time series. Most of them are based on the distance between the *values* of time series (cf., e.g., [2, 115]). The problem is that all such indexes are necessarily distorted by random noise. Consequently, the real shape of the time series is hidden. A solution to these difficulties can be given by the F-transform. In this section, we demonstrate two new methods for measuring relations among time series in the form of similarity in their behaviour. The basis of these methods is in the application of the F-transform and custom metrics. The time series can then be paired together according to the similarity of the adjoint time series consisting of the local trends. We demonstrate the application of the suggested methods to real financial time series, and two artificial ones. Experimental results verify the capability of these suggested methods to measure similarity between time series.

6.1 Fuzzy equality

Let $\langle [0,1], \vee, \wedge, \otimes, \rightarrow, 0, 1\rangle$ be an algebra of truth values where $\vee = \max$, $\wedge = \min$, $a \otimes b = \max\{0, a+b-1\}$, $a \rightarrow b = \min\{1, 1-a+b\}$, $a, b \in [0,1]$. This algebra is called *Łukasiewicz standard MV-algebra*[§)].

Definition 6 *Let* $\doteq : U \times U \longrightarrow [0,1]$ *be a binary fuzzy relation.*

(i) It is reflexive *if* $(u \doteq u) = 1$ *for all* $u \in U$.

(ii) It is symmetric *if* $(u \doteq v) = (v \doteq u)$ *for all* $u, v \in U$.

(iii) It is transitive *if* $(u \doteq v) \otimes (v \doteq w) \leq (u \doteq w)$ *for all* $u, v, w \in U$.

(iv) It is separated *if for all* $u, v \in U$

$$(u \doteq v) = 1 \text{ iff } u = v.$$

Definition 7

(i) A binary fuzzy relation $\doteq$ *on* U *is a* fuzzy symmetry *if it is reflexive and transitive.*

(ii) A fuzzy symmetry is a fuzzy equality *if it is also transitive.*

§) Of course, there are also other algebras used as algebras of truth values for fuzzy set theory and fuzzy logic. The Łukasiewicz algebra, however, has a prominent position because of its good properties in many respects and, therefore, we confine our theory to it only.

6.2 Zero degree similarity method

Let us consider realizations of two time series $\{F_i(t) \mid t = 1, \dots, n\}$, $i = 1, 2$ and let $S : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow [0, 1]$ be a fuzzy relation defined by

$$S(F_1, F_2) = 1 - \frac{1}{n} \sum_{t=1}^{n} \frac{|F_1(t) - F_2(t)|}{|F_1(t)| + |F_2(t)|}. \tag{27}$$

It is easy to show that $S(F_1, F_2) \in [0, 1]$.

Theorem 6.1 *The fuzzy relation S given in (27) is a separated fuzzy symmetry. Let F_1, F_2, F_3 be realizations of time series of the length n. If $|F_2(t)| \leq \min\{|F_1(t)|, |F_3(t)|\}$ for all $t = 1, \dots, n$ then S is a fuzzy equality w.r.t. Łukasiewicz t-norm $\otimes$.*

Proof: (a) The reflexivity $S(F_1, F_1) = 1$ is immediate. The symmetry follows from the properties of absolute value.

(b) Separateness: If $F_1 = F_2$ then $S(F_1, F_2) = 1$ by reflexivity. Conversely, let $S(F_1, F_2) = 1$. Then

$$\frac{|F_1(t) - F_2(t)|}{|F_1(t)| + |F_2(t)|} = 0$$

for all t, which holds only if $F_1 = F_2$.

(c) The transitivity requires $S(F_1, F_2) \otimes S(F_2, F_3) \leq S(F_1, F_3)$. This holds if

$$\frac{|F_1(t) - F_3(t)|}{|F_1(t)| + |F_3(t)|} \leq \frac{|F_1(t) - F_2(t)|}{|F_1(t)| + |F_2(t)|} + \frac{|F_2(t) - F_3(t)|}{|F_2(t)| + |F_3(t)|},$$

for $t = 1, \dots, n$. This inequality is fulfilled if both $|F_2(t)| \leq |F_1(t)|$ as well as $|F_2(t)| \leq |F_3(t)|$ hold for all $t = 1, \dots, n$.

Definition 8 Let $X = \{X(t) \mid t = 1, \dots, n\}$ and $Y = \{Y(t) \mid t = 1, \dots, n\}$ be two time series of the length n and $\widetilde{TC}_X$ and $\widetilde{TC}_Y$ be estimations of trend-cycles of X and Y respectively[¶)], calculated using equation (22) for a suitable fuzzy partition $\mathcal{A}_h$. Then we define the similarity between these two time series as follows:

$$S(\widetilde{TC}_X(t) - \mathbf{E}(\widetilde{TC}_X), \widetilde{TC}_Y(t) - \mathbf{E}(\widetilde{TC}_Y)) =$$
$$1 - \frac{1}{n} \sum_{t=1}^{n} \frac{|\,\widetilde{TC}_X(t) - \mathbf{E}(\widetilde{TC}_X) - (\widetilde{TC}_Y(t) - \mathbf{E}(\widetilde{TC}_Y))|}{|\,\widetilde{TC}_X(t) - \mathbf{E}(\widetilde{TC}_X)| + |\,\widetilde{TC}_Y(t) - \mathbf{E}(\widetilde{TC}_Y)|}, \tag{28}$$

where $\mathbf{E}(\widetilde{TC}_X)$ and $\mathbf{E}(\widetilde{TC}_Y)$ are mean values (averages) of $\widetilde{TC}_X$ and $\widetilde{TC}_Y$, respectively.

¶) It is necessary to emphasize that we can work with estimations $\widetilde{TC}_X$ and $\widetilde{TC}_Y$ of the trend-cycle only, because we do not know the real ones.

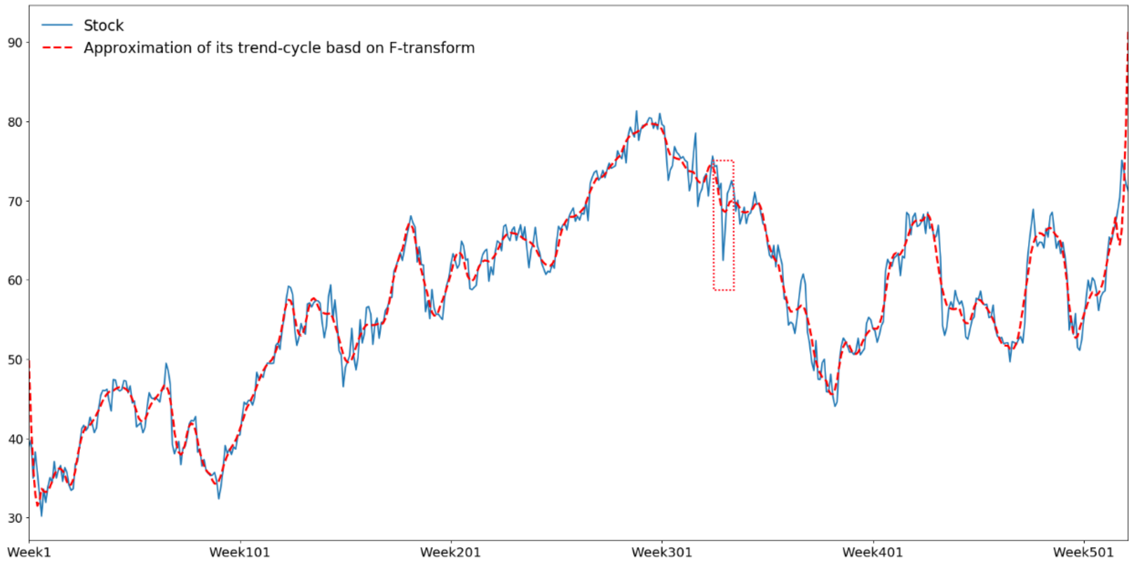


Figure 3: A Stock and its TC approximation based on F-transform

It follows from Theorem 6.1 that the similarity (28) is a fuzzy symmetry, or sometimes even fuzzy equality. For simplicity, in the sequel we will write (28) simply as $S(X, Y)$.

Remark 1 *If we take X and Y as simple linear functions $X = \{k_1 t + q_1 \mid t = 1, \dots, n\}$, $Y = \{k_2 t + q_2 \mid t = 1, \dots, n\}$ then, by simple computation, we obtain that the similarity $S(X, Y) = 1 - \frac{|k_1 - k_2|}{|k_1| + |k_2|}$. Hence, if these lines are (almost) parallel then $S(X, Y) \approx 1$.*

Stock price, can be seen as a time series $X = \{X(t) \mid t = 1, \dots, n\}$ where $X(t)$ is a closing price at time $t \in \{t = 1, \dots, n\}$. For instance, let us consider closing price of a stock from Nasdaq INC[‖)], from 05.10.2008 to 30.09.2018 (522 weeks). In order to estimate its local trend-cycle, we first build a uniform fuzzy partition $\mathcal{A}_h$ such that the length of each basic function $A_2, \dots, A_m \in \mathcal{A}_h$ is equal to a proper time slot. In our case, by setting the length $h \in \{2, 3\}$, we obtain the approximation of the trend-cycle for one month. In other terms, the monthly behaviour of this stock is our concern here. Figure 3 depicts the mentioned weekly stock and the fuzzy approximation of its local trend-cycle. The first and the last components of F-transform are subject to a big error (because the corresponding basic functions A_1 and A_m are incomplete). Regardless of this, it is clear that the F-transform approximates the local trend-cycles of the stock successfully. As we mentioned before, stock markets react to many exogenous factors; thus, the presence of outliers is unavoidable. A red square in Figure 3 shows one of these outliers for the mentioned stock. It can be seen that the F-transform has successfully wiped out the outlier while preserving the core behaviour of the stock.

The similarity from definition 8 can be used for measuring similarity for any number of stocks based on their local behaviour.

[‖)] https://www.nasdaq.com/Second footnote

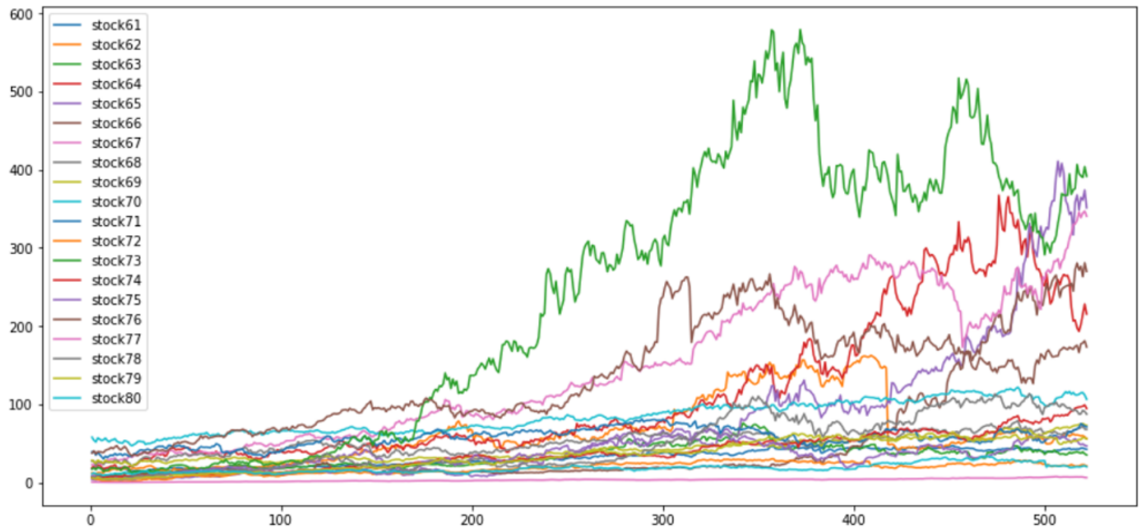


Figure 4: Depiction of 20 stocks from the dataset for 522 weeks

In the next section, we will demonstrate how our suggested method works with a relatively large data set of stock prices in conjunction with its comparison to standard the Euclidean distance. The goal is to demonstrate the performance of the method in comparison to one of the most known similarity measurements. Further, we demonstrate how our method allows us to assess the lead-lag relation among returns of different assets in complementary to statistical correlation analysis.

6.3 Demonstration

Our first data set consists of a closing price of 92 stocks over 522 weeks obtained from Nasdaq INC[*)]. An example of twenty stocks from the mentioned data set is depicted in figure 4, where the x-axis and y-axis represent price values in dollars and number of weeks, respectively. From this figure, it is clear that any decision about the similarity between time series is impossible. Therefore it seems necessary to consider similarity between time series.

6.4 Evaluation

One possible way to evaluate the competency of any new similarity measurement (distance measurement), is to apply it to data clustering. The quality of clustering based on the new and current similarities can validate the competency of the suggested method [116, 117]. Therefore, we will below apply clustering of time series and compare the behaviour of our similarity with the Euclidean one. However, let us emphasize that time series clustering is not the primary goal of this research since our focus is on discovering the most similar pairs of stocks available in the database. As we mentioned before, the Euclidean distance is an accurate, robust, simple, and efficient way to measure the similarity between two time series and,

*) https://www.nasdaq.com/Second footnote

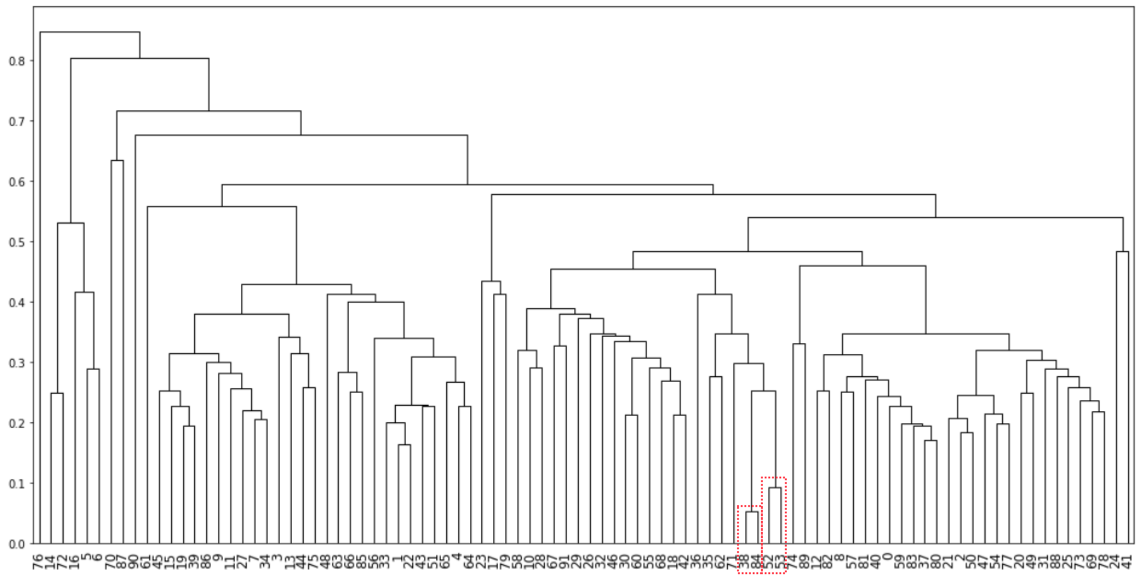


Figure 5: Hierarchical clustering based on the suggested method

surprisingly, can outperform most of the more complex approaches (-see [7], [118]). Therefore we will compare our method with the Euclidean distance by means of the quality of hierarchical clustering on a dataset. Hierarchical clustering is a method of cluster analysis that attempts at building a hierarchy of similar groups in data [119]. In this case, one problem to consider is the optimal number of clusters in a dataset. Overall, none of the methods for determining the optimal numbers of clusters is flawless, and none of the suggested similarities is fully satisfactory. Hierarchical clustering does not reveal an adequate number of clusters and estimation of the proper number of clusters is rather intuitive. Hence, there is a fair amount of subjectivity in the determination of separate clusters. Figures 5 and 6, demonstrate the dendrogram of hierarchical clustering of the 92 stocks based on the suggested and Euclidean similarity, respectively. The proper number of clusters for both similarities is equal to six. In these figures, the 92 stocks are represented in the x-axis, and their distances are depicted on the y-axis accordingly. Since the stocks are from various industries, they have different scales, and in the case of the clustering with the Euclidean distance, we will eliminate the different scaling by normalizing the data. Nevertheless, this step is not demanded by the suggested method since the scale does not influence it.

Red dashed squares in 5 and 6 represent the most similar stock pairs, determined according to each method. Interestingly, both methods selected the same stock pairs; (38 and 84) and (52 and 53) as the most similar stocks. However, the suggested method primarily determines the stock pair (38 and 84) as the most similar stocks, following by the stock pair (52 and 53) while the Euclidean method suggests otherwise. Figure 7 and 8 show the behaviour of these stock pairs.

To measure the quality of clustering, we apply the Davies-Bouldin index, which is usually used in clustering. This measure evaluates intra-cluster similarity and inter-cluster differences [120]. Therefore, it can be a proper metric for clustering

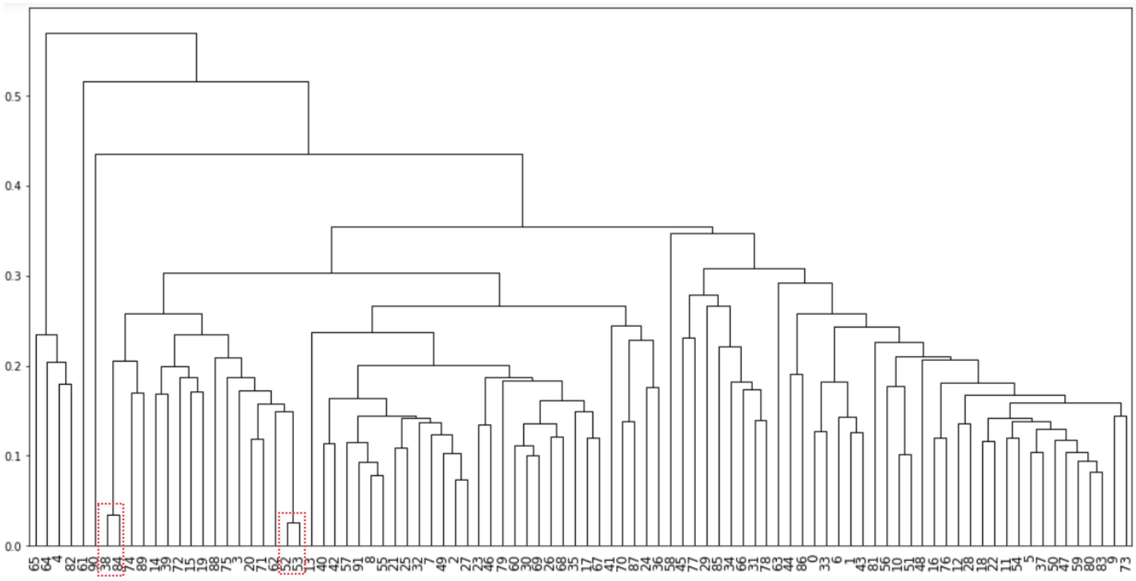


Figure 6: Hierarchical clustering based on the Eucliden method

Figure 7: The stock pair (38 and 84)

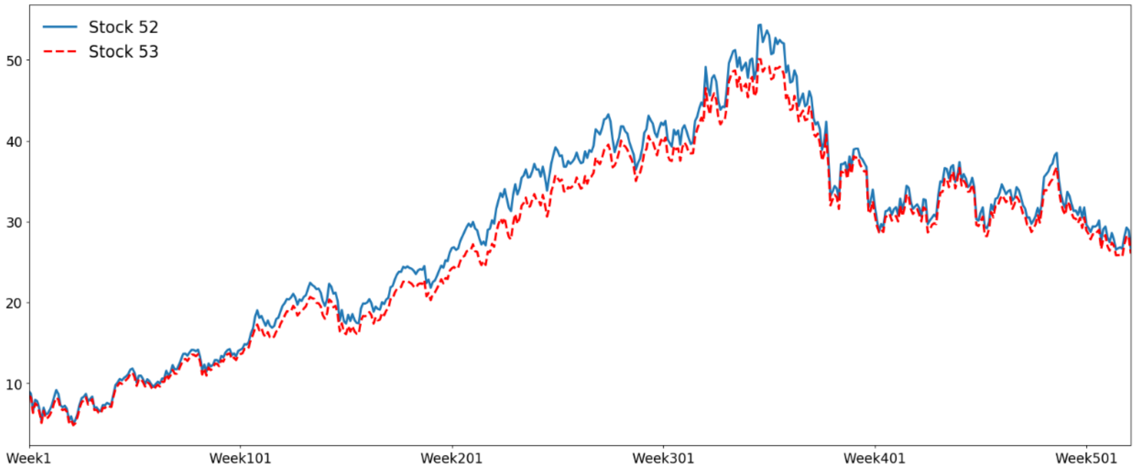


Figure 8: The stock pair (52 and 53)

Table 4: The Davies-Bouldin index for clustering based on the proposed method and Euclidean method

Method	6 **clusters**	8 **clusters**	10 **clusters**
The suggested method	0.61	0.64	0.72
The Euclidean method	0.71	0.85	0.82

evaluation.

Table 4 demonstrates the Davies-Bouldin index for a different number of clusters based on the both similarities. Since the lower score indicates better quality of clustering, the results reveal that not only is our method reasonably comparable to the Euclidean method, but also provides more efficient clustering for these examples.

Furthermore, as we mentioned before, stock markets are prone to exogenous factors such as bad or good news (-see [121]). If a method pairs two stocks as similar, one can expect that after the occurrence of an outlier(s), the method would still evaluate these stocks alike. Hence, we will compare the performance of our method, and the Euclidean distance metric for the stocks containing outliers. Recall from the previous section that based on both methods, stocks 52 and 53 are very similar to each other since their distance is minimal. Therefore, first, we will add some random artificial outliers to the stock 52, but we do not alter the stock 53 as shown in Figure 9. Subsequently, we apply both methods to re-evaluate the similarity between these stocks.

Table 5 demonstrates the results. It is apparent, after including artificial outliers, that the Euclidean distance has a dramatic jump (around 1800% increase). At the same time, the purposed method shows a minimal increase in distance (33%), which means that the suggested method is much less sensitive to the presence of outliers. Considering that the suggested method is based on the F-transform, it evaluates the similarity between the stocks concerning their local trend-cycles; therefore, it does not have the drawbacks of raw-data based approaches such as the Euclidean

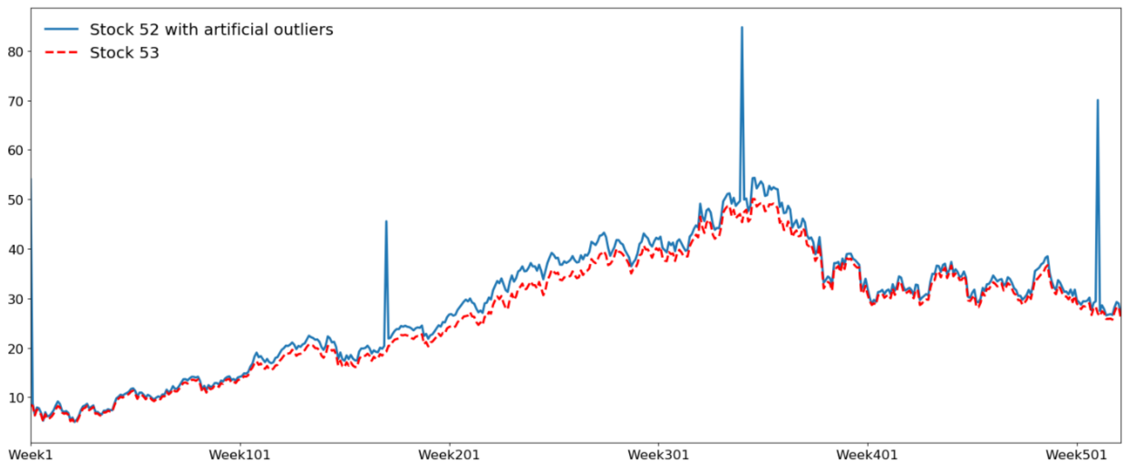


Figure 9: The stock pair (52 and 53) containing artificial outliers

Table 5: The distance between stocks 52 and 53, before and after outliers

Method	**Distance before outliers**	**Distance after outliers**
The suggested method	0.09	0.12
The Euclidean method	0.17	3.33

distance. The latter methods are sensitive to noisy data [122]. One advantage of the Euclidean method is its simplicity; however, the suggested method is also relatively simple since it has only one parameter to set (the length of the basic functions). Moreover, experts are able to adjust the suggested similarity measure, according to their time slot of interest.

6.5 Similarity at shifted-lag or lead-lag relation

The examples we provided earlier demonstrate the applicability and strength of the suggested method in finding similar behaviour between stocks at lag zero. However, there exist situations that two stocks might not be significantly similar at lag zero, but they are more similar in shifted-lag(s). A condition where one (leading) variable is cross-correlated with the values of another (lagging) variable at other times is characterized as a lead-lag effect. The existence of the lead-lag effect between markets and its causes has been authenticated by many researchers [123, 124, 125, 126]. Generally, in practice, investors are interested in the relation between the return of the market and not actual price values. Let N be the number of assets, $P_i(t)$ be the price at time t of asset i, 1≤i≤N, then the return of an asset $R_i(t)$, is calculated as:

$$R_i(t) = \frac{P_i(t) - P_i(t-1)}{P_i(t)}.$$

Lo and MacKinlay were among the first pioneers who showed how the return of small firms correlate with past returns of big firms [127], and more recently, Kewei

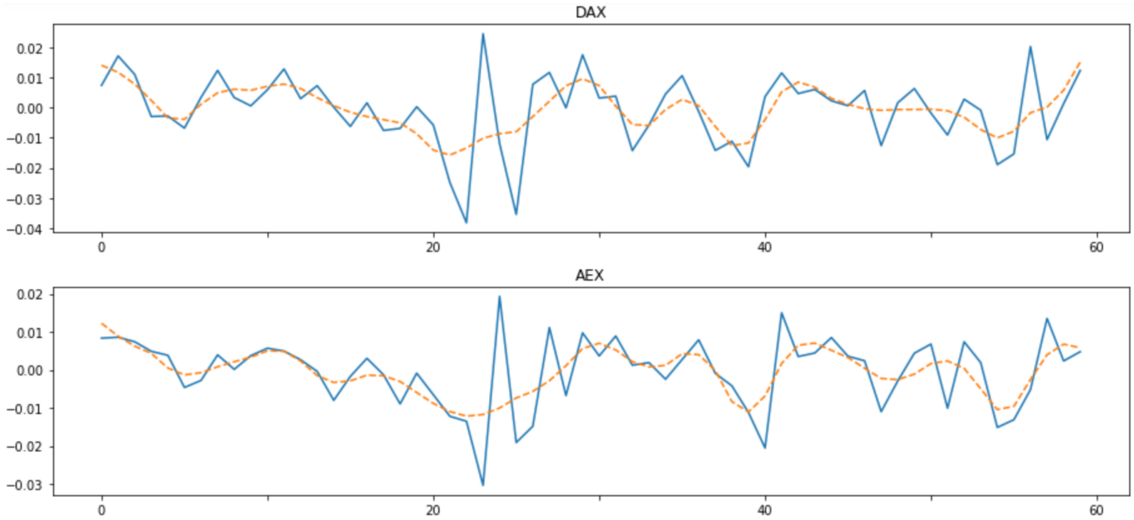


Figure 10: Indices and their trend

Hou argues that there is strong intra-industry lead-lag effect [128]. The conventional method to evaluate this lead-lag relation among international stock indices is by cross-correlation. In order to assess the lead-lag relation with the suggested method, we will move the stock returns against each other in different lags. Since the suggested method evaluates the strength of similarity between stocks, therefore if the similarity degree in shifted-lag is considerably higher, we can assume that there is a lead-lag relation between the return of stocks. To illustrate the method in practice, here we demonstrate the relation between the daily return of two international stock indices, Germany (DAX) and the Netherlands (AEX). We obtained their daily closing prices from yahoo finance[**)] from 03/01/2018 to 29/03/2018 and calculate their return for that period.

The behaviour of these daily returns, as well as their trend estimation, is represented in figure 10. The blue line demonstrates the data, and the dashed orange line is their estimation by F-transform. Note that unlike the prices in the previous example, here, we do not seek exact estimation for daily returns.

Data shows that lag one is a proper choice for shifting, meaning that we measure the similarity among the returns with one shift. Table 6 demonstrates the degree of their similarity at lag zero and lag one. These results suggest that the return of AEX follows a similar behaviour as DAX after one day. The suggested similarity measure shows that the similarity between Dax and AEX(-1) is higher than their association at lag zero. Seemingly, Cross-correlation confirms this conclusion as well.

As shown in Figure 11, by shifting the AEX for one lag, its similarity to DAX increases. Therefore, arguably for this period, Dax has a leading effect on AEX. Note that this relationship should not be considered as a causal relation.

However, it is possible to examine if this lead-lag relation founded by the suggested method can be causal. By causal relation, we mean the so-called Granger

[**)] www.finance.yahoo.com

Table 6: Cross-similarity and cross-correlation between daily return of DAX and AEX

Returns	i	Similarity	Correlation
AEX, DAX(i)	0	0.46	0.35
AEX, DAX(i)	-1	0.60	0.70

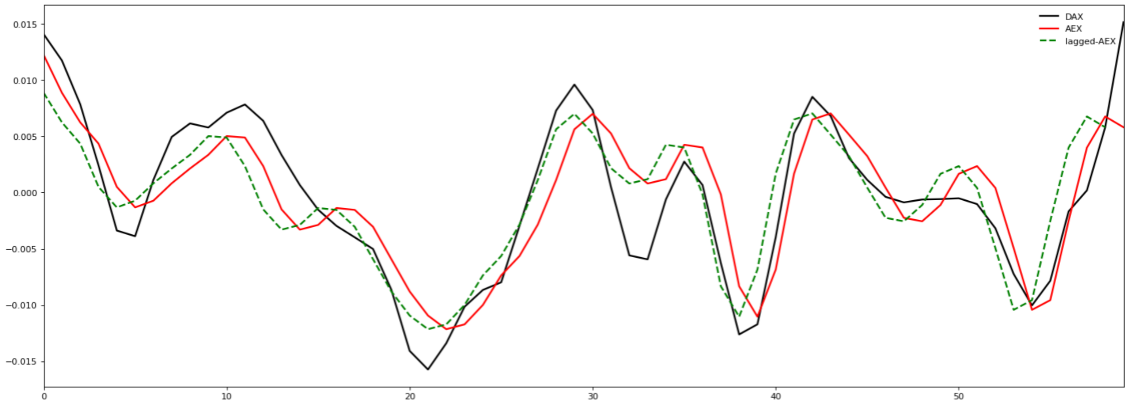


Figure 11: DAX, AEX, shifted-AEX

causality [129]. This concept is defined in terms of the predictability of a variable from its own past or the past of another variable. A time series X is said to Granger-cause a time series Y if the available information apart from X provides statistically significant information about future values of Y.

In practice, to model the causality between two variables, it is imperative to determine the direction of the causality and its lag. To measure the causality, in [129], Granger proposed to compute the causal lag and causal strength (concerning two distinguish directions) based on the coherence and the phase functions defined with the help of the cross-spectrum of two stationary processes [130].

Granger originally proposed a test based on comparing the mean square error of the forecasts of a variable with and without using the past of another variable. This work is then generalized in [131], where he assumes that, at time t, the value $Y(t+1)$ is a random variable that can be characterized by a probability statement of the form $\text{Prob}(Y(t+1) \in A)$, for some given set A. Then, a general Granger causality definition is the following.

Definition 9 A time series X is said to cause Y if

$$\text{Prob}(Y(t+1) \in A \mid \mathcal{F}_t) \neq \text{Prob}(Y(t+1) \in A \mid \mathcal{F}_{-X}(t)),$$

where A is a universe in which $X(t)$ and $Y(t)$ are measured at specific time points $t \in \{1, \ldots, t\}$. Furthermore, $\mathcal{F}(t)$ represents information available at time t in the entire universe, and $\mathcal{F}_{-X}(t)$ is this information after X being excluded.

Hence, we tested the Granger causality between the pair (DAX, AEX) at lag 1. Table 7 depicts the results. Here, we cannot reject the hypothesis that Dax does not

Table 7: Granger causality between the pair DAX-AEX provided by the software EViews10

Null hypothesis	Observations	F-statistic	Prob.
DAX does not Granger-cause AEX	59	68.5058	3.E-11
AEX does not Granger-cause DAX		1.01926	0.3170

Granger-cause AEX, but we do reject the hypothesis that AEX does not Granger-cause DAX. Therefore it appears that Granger causality runs one-way from DAX to AEX at lag 1, and not the other way. This finding can be used later on for improving the prediction of AEX based on the past values of DAX. Thereby, the lead-lag relation that we found earlier, can be regarded as a Granger-causal relation.

6.6 Higher degree similarity method

Let us consider two time series X, Y with the same time domain $\mathbb{T}$. Since values of the time series can fall in very different ranges, we first normalize both time series to make their values comparable. The normalization will be done w.r.t. maximal values $\overline{X} = \max\{|X(t)| \mid t \in \mathbb{T}\}$, $\overline{Y} = \max\{|Y(t)| \mid t \in \mathbb{T}\}$. Then we put

$$X^N = \left\{ \frac{X(t)}{\overline{X}} \;\middle|\; t \in \mathbb{T} \right\}, \tag{29}$$

$$Y^N = \left\{ \frac{Y(t)}{\overline{Y}} \;\middle|\; t \in \mathbb{T} \right\}. \tag{30}$$

Let us choose two numbers $h_0, h_1 > 0$, compute $n_0 = \frac{|\mathbb{T}|}{h_0}$ and $n_1 = \frac{|\mathbb{T}|}{h_1}$ and form two h_0, h_1-uniform fuzzy partitions $\mathcal{A}$, $\mathcal{B}$ of the time domain $\mathbb{T}$.

Let us compute components of zero and first-degree F-transforms of the time series (29) and (30):

$$\mathbf{F}^0[X^N] = (F_1^0[X^N](t), \dots, F_{n_0-1}^0[X^N](t)),\ \mathbf{F}^1[X^N] = (F_1^1[X^N](t), \dots, F_{n_1-1}^1[X^N](t)) \tag{31}$$

$$\mathbf{F}^0[X^N] = (F_1^0[Y^N](t), \dots, F_{n_0-1}^0[Y^N](t)),\ \mathbf{F}^1[X^N] = (F_1^1[Y^N](t), \dots, F_{n_1-1}^1[Y^N](t)) \tag{32}$$

where the zero-degree components are computed on the basis of the fuzzy partition $\mathcal{A}$, and the first-degree ones on the basis of $\mathcal{B}$. Note that in the direct F-transforms above we omitted the first and the last components. For further processing, we need only the coefficients β_k^0, $k = 1, \dots, n_0 - 1$, and β_k^1, $k = 1, \dots, n_1 - 1$ defined in (7), (8).

Based on that, we form to each time series X^N, Y^N two new reduced time series, namely a time series of values and that of tangents:

$$\boldsymbol{\beta}_X^0 = \{\beta_{X,k}^0 \mid k = 1, \dots, n_0 - 1\}, \tag{33}$$

$$\boldsymbol{\beta}_X^1 = \{\beta_{X,k}^1 \mid k = 1, \dots, n_1 - 1\}, \tag{34}$$

$$\boldsymbol{\beta}_Y^0 = \{\beta_{Y,k}^0 \mid k = 1, \dots, n_0 - 1\}, \tag{35}$$

$$\boldsymbol{\beta}_Y^1 = \{\beta_{Y,k}^1 \mid k = 1, \dots, n_1 - 1\}. \tag{36}$$

Hence, $\boldsymbol{\beta}_X^0$ and $\boldsymbol{\beta}_Y^0$ are time series of average values of the respective time series X^N, Y^N over the imprecisely specified areas $A_k \in \mathcal{A}$, $k = 1, \dots, n-1$, and $\boldsymbol{\beta}_X^1$ and $\boldsymbol{\beta}_Y^1$ are time series of average values of tangents of the time series X^N, Y^N over the imprecisely specified areas $B_k \in \mathcal{B}$, $k = 1, \dots, n-1$.

Definition 10 Let $X = \{X(t) \mid t = 1, \dots, n\}$ and $Y = \{Y(t) \mid t = 1, \dots, n\}$ be two time series of the length n, then the *index of similarity* of two time series is the number

$$S(X,Y) = \max\left\{0, 1 - \frac{\kappa_0}{n_0 - 1}\sum_{k=1}^{n_0-1} |\beta_{X,k}^0 - \beta_{Y,k}^0| + \frac{\kappa_1}{n_1 - 1}\sum_{k=1}^{n_1-1} \frac{|\beta_{X,k}^1 - \beta_{Y,k}^1|}{\varphi}\right\} \tag{37}$$

where φ is a common normalization factor assuring that both $|\beta_{X,k}^1|, |\beta_{Y,k}^1| \leq 1$ for all $k = 1, \dots, n_1 - 1$ and κ_0, κ_1 are sensitivity constants.

The suggested similarity index thus considers not only the distances between average values of time series but also the distances between average values of tangents in the same areas. The constants κ_0, κ_1 increase or decrease the sensitivity of values and slopes of the compared time series. Clearly, if $\kappa_0 > 1$ then $S(X,Y)$ is more sensitive to differences between the corresponding values of X, Y, while $\kappa_1 > 1$ does the same for their slopes.

The normalization factor φ can be specified, e.g., as follows. Let X, Y, Z be time series (1) and put

$$\overline{\beta_X^1} = \max\{|\beta_{X,k}^1| \mid k = 1, \dots, n_1 - 1\},$$
$$\overline{\beta_Y^1} = \max\{|\beta_{Y,k}^1| \mid k = 1, \dots, n_1 - 1\},$$
$$\overline{\beta_Z^1} = \max\{|\beta_{Z,k}^1| \mid k = 1, \dots, n_1 - 1\}.$$

Then

$$\varphi = \max\{\overline{\beta_X^1}, \overline{\beta_Y^1}\} \tag{38}$$

is a normalization factor common for X, Y and

$$\varphi = \max\{\overline{\beta_X^1}, \overline{\beta_Y^1}, \overline{\beta_Z^1}\} \tag{39}$$

is a normalization factor common for X, Y, Z.

The following is immediate.

Lemma 6.2

$$S(X,Y) \in [0,1].$$

Theorem 6.3 *Let X, Y be two time series.*

(a) *The similarity index $S(X, Y)$ is a separated fuzzy symmetry.*

(b) *If $S(X, Y) \neq 0$ then the transitivity property $S(X, Z) \otimes S(Z, Y) \leq S(X, Y)$ holds for any time series Z.*

Proof: (a) If $X = Y$ then, obviously, $S(X, Y) = 1$. The symmetry follows immediately from the properties of absolute value.

(b) Let φ be common normalization factor for X, Y, Z. After rewriting we obtain

$$\max\Bigg\{0,\ 1 - \frac{\kappa_0}{n_0 - 1}\sum_{k=1}^{n_0-1}|\beta^0_{X,k} - \beta^0_{Y,k}| + \frac{\kappa_1}{n_1 - 1}\sum_{k=1}^{n_1-1}\frac{|\beta^1_{X,k} - \beta^1_{Y,k}|}{\varphi} +$$
$$1 - \frac{\kappa_0}{n_0 - 1}\sum_{k=1}^{n_0-1}|\beta^0_{Y,k} - \beta^0_{Z,k}| + \frac{\kappa_1}{n_1 - 1}\sum_{k=1}^{n_1-1}\frac{|\beta^1_{Y,k} - \beta^1_{Z,k}|}{\varphi} - 1\Bigg\} \leq$$
$$1 - \frac{\kappa_0}{n_0 - 1}\sum_{k=1}^{n_0-1}|\beta^0_{X,k} - \beta^0_{Z,k}| + \frac{\kappa_1}{n_1 - 1}\sum_{k=1}^{n_1-1}\frac{|\beta^1_{X,k} - \beta^1_{Z,k}|}{\varphi}.$$

If the left-hand side is equal to 0 then the inequality is trivially fulfilled. By the assumption, we have to verify that

$$\sum_{k=1}^{n_0-1}|\beta^0_{X,k} - \beta^0_{Z,k}| \leq \sum_{k=1}^{n_0-1}|\beta^0_{X,k} - \beta^0_{Y,k}| + \sum_{k=1}^{n_0-1}|\beta^0_{Y,k} - \beta^0_{Z,k}|,$$
$$\sum_{k=1}^{n_1-1}\frac{|\beta^1_{X,k} - \beta^1_{Z,k}|}{\varphi} \leq \sum_{k=1}^{n_1-1}\frac{|\beta^1_{X,k} - \beta^1_{Y,k}|}{\varphi} + \sum_{k=1}^{n_1-1}\frac{|\beta^1_{Y,k} - \beta^1_{Z,k}|}{\varphi}$$

which holds using the triangular inequality and the properties of ordered groups.

Finally, let $S(X, Y) = 1$. Then it follows from (37) that $\beta^0_{X,k} = \beta^0_{Y,k}$ for all $k = 1, \ldots, n_0 - 1$, as well as $\beta^1_{X,k} = \beta^1_{Y,k}$ for all $k = 1, \ldots, n_1 - 1$. Since all the fuzzy sets $A_1, \ldots, A_{n_0-1}$ as well as $B_1, \ldots, B_{n_1-1}$ cover the whole time domain $\mathbb{T}$, we conclude that $X = Y$.

Theorem 6.4 *Let $X \equiv q_1$ and $X \equiv q_2$ be two constant time series. Then*

$$S(X, Y) = 1.$$

Proof: It follows from (29), (30) that $X^N = Y^N \equiv 1$. Then $\beta^0_{X,k} = \beta^0_{Y,k} = 1$ by Theorem 5.1(d), and $\beta^1_{X,k} = \beta^1_{Y,k} = 0$ by the properties of tangent.

It follows from this proposition that if both time series X, Y are constant then they are fully similar.

6.7 Demonstration

In this section, we will demonstrate the new similarity index on real data. In all cases, we applied F^0-transform with $h = 3$ and the sensitivity constant $\kappa_0 = 2.5$,

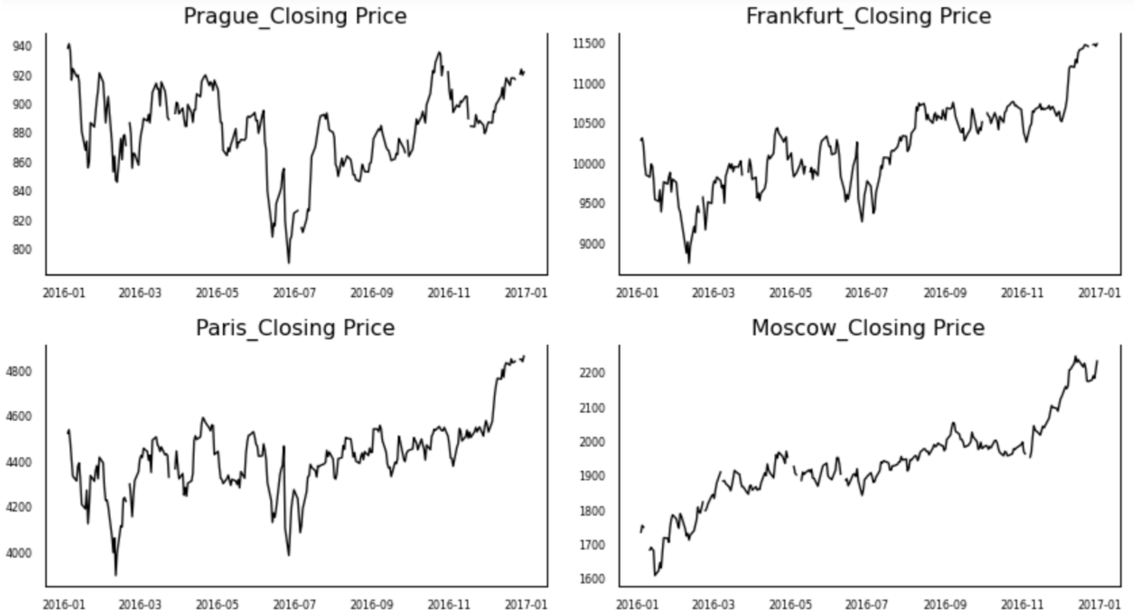


Figure 12: Closing price of four Stocks

and F^1-transform with $h = 5$ and the sensitivity constant $\kappa_2 = 2$. We chose four real financial time series containing daily closing prices of stock markets for 2016. Several interesting events, such as falling the oil price and the Brexit announcement and voting, heavily affected the stocks market in 2016 and the interrelations. Therefore it is valuable to investigate the underlying relations. Figure 12 demonstrates the daily closing price of these four markets. Due to different national holidays, each market contains some missing values which was omitted for similarity measurement.

Using the suggested method, we evaluate the pairwise similarities among the stocks and indexes are as follows:

$$\begin{aligned} S(\text{Prague}, \text{Moscow}) &= 0.77, \\ S(\text{Prague}, \text{Paris}) &= 0.9, \\ S(\text{Prague}, \text{Frankfurt}) &= 0.84, \\ S(\text{Paris}, \text{Frankfurt}) &= 0.93. \end{aligned}$$

The results reveal several interesting relations. The most similar stock to Prague is the Paris market, while Moscow has the lowest similarity. Another exciting relation is about the Frankfurt market. The pair (Frankfurt-Paris) has a higher similarity compare to (Frankfurt-Prague). These relations are also visible in Figure 12. In Figures 13–16, we demonstrate the behaviour of the normalized prices for the mentioned similarity Indexes.

Adding massive noise to the Moscow market and inverting the Prague market's price values, we have constructed two artificial time series. The purpose is to investigate the similarity index changes between Prague and these two artificial time series. Figures 17–18 demonstrate the comparison of the mentioned time series and

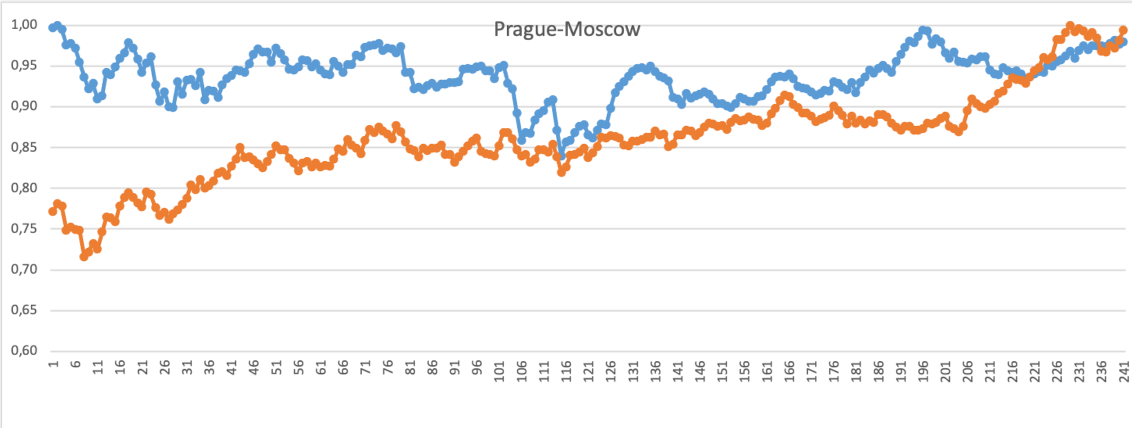


Figure 13: Comparison of graphs of normalized prices for Prague and Moscow; $S(\text{Prague}, \text{Moscow}) = 0.77$.

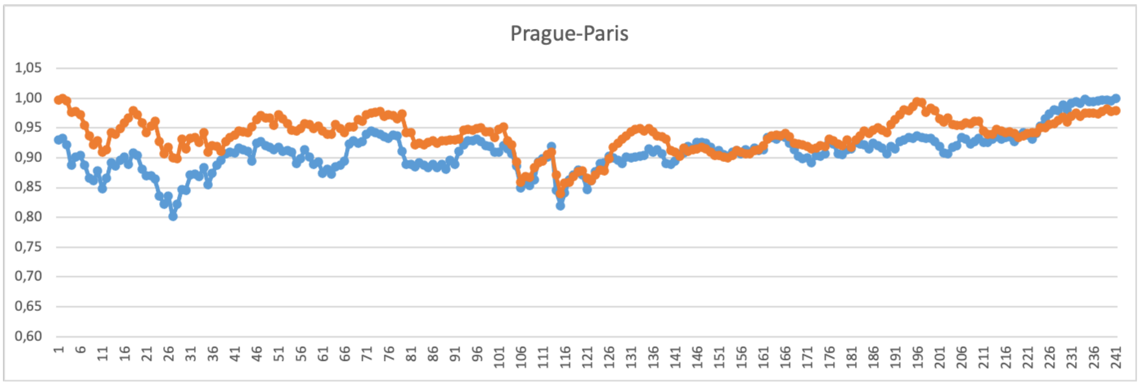


Figure 14: Comparison of graphs of normalized prices for Prague and Paris; $S(\text{Prague}, \text{Paris}) = 0.9$.

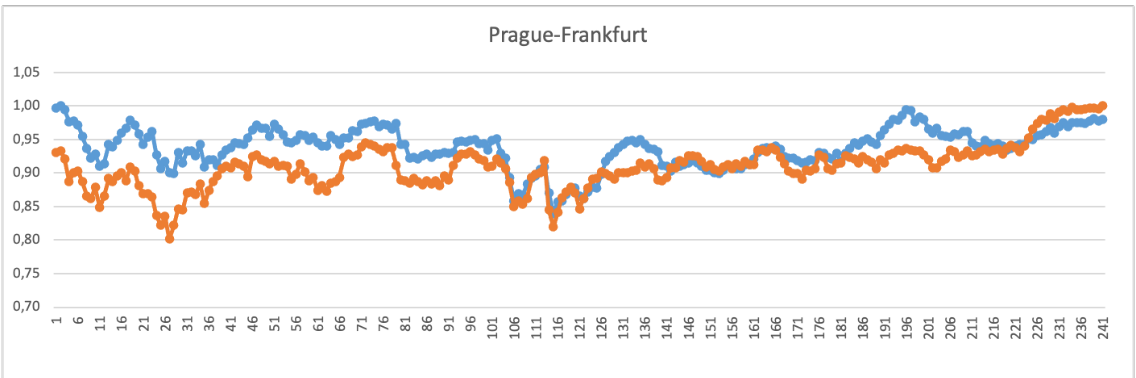


Figure 15: Comparison of graphs of normalized prices for Frankfurt and Prague; $S(\text{Prague}, \text{Frankfurt}) = 0.84$.

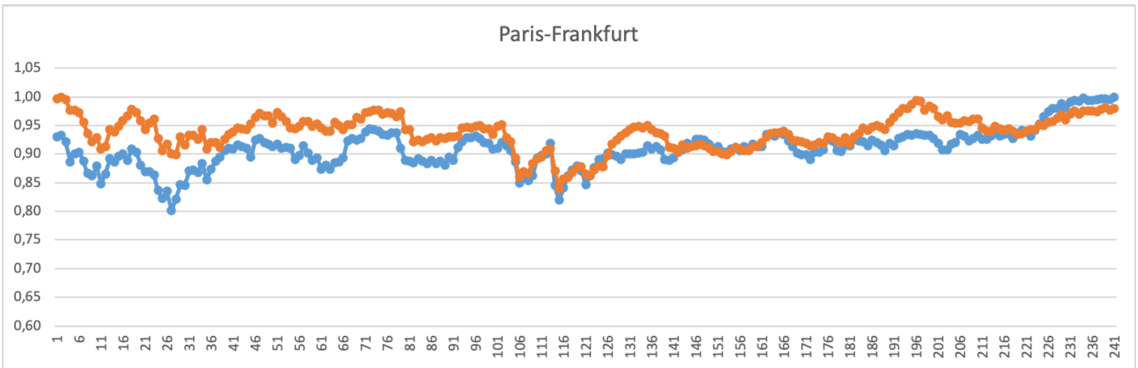


Figure 16: Comparison of graphs of normalized prices for Frankfurt and Paris; $S(\text{Paris}, \text{Frankfurt}) = 0.93$.

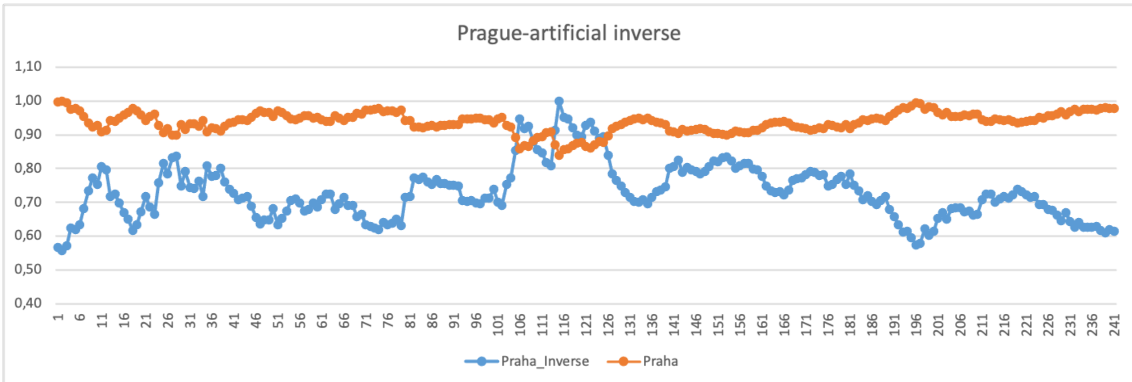


Figure 17: Comparison of graphs of Prague and its artificial inverse stock prices; $S(\text{Prague}, \text{Prague-artificial inverse}) = 0.34$.

the similarity indexes are as follows:

$$S(\text{Prague}, \text{Moscow-distorted}) = 0,$$
$$S(\text{Prague}, \text{Prague-artificial inverse}) = 0.34.$$

The similarity index dropped dramatically but reasonably. As one expects, there is zero similarity between Prague and distorted Moscow. However, similarity between Prague and its inverted version still remained non-zero which is caused especially by found similarities between their corresponding slopes.

We introduced two new methods for measuring similar behaviour among financial time series in this section. In the following section, we will build on these methods and apply them to prediction.

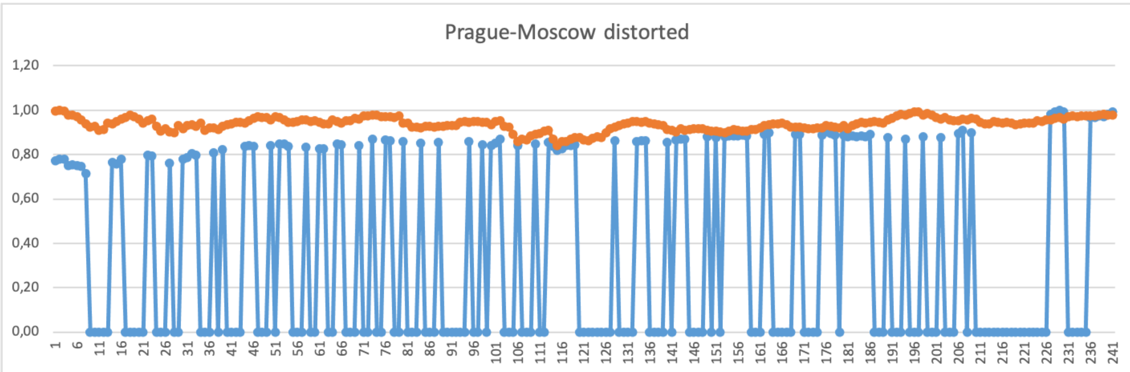


Figure 18: Comparison of graphs of Prague and artificially distorted Moscow stock prices; $S(\text{Prague}, \text{Moscow-distorted}) = 0$.

7 Suggested workflow for financial time series analysis and prediction

In this section, we introduce a workflow that prepares financial time series, assess their similarity and provide a predictive linguistic description in natural language. Figure 19 demonstrate the mentioned workflow is automated and is available in a Github repository [††)]. 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 [132].

7.1 Data preparation

This stage is accountable for manipulating, modeling, and cleaning time series. In order to do that, the following tasks are performed:

- **length Leveling:** possibly, some of the time series might miss some data points due to many reasons such as system errors, human errors, various vacation days, or unexpected closing days (e.g., Covid-19 pandemic). For some methods such as Dynamic Time Warping (DTW), this task is not necessary, but for the purposed method, it is essential that time series have the same length size; therefore we implemented and examined the following fixing methods.
  - **Uniform scaling:** is a common technique for equalizing the length of time series [133]. It resizes one time series based on the length of the other one. Here we apply time join, which is one of the most popular techniques. If one time series misses a time point(s) while the other does not, we will eliminate the extra time point(s). It is also possible to uniform the scales of time series utilizing F-transform. To achieve that, we can apply the F^0-transform with an equal number of components to all time series.
  - **Padding:** is another technique to fix the problem of varying length between time series. There are different types of padding techniques, but the most common one is to concatenate the shorter time series with zero or another to duplicate value at the beginning of the shorter time series. While the missing points are not too many, padding does not affect the shape of the time series. In our implementation, we apply the duplication.
- **Re sampling:** sometimes time series have different time horizons; daily, weekly, monthly, or annually. For the suggested method, it is essential that the time series of interest have the same duration. Therefore, in this phase, for instance, a daily time series can be converted to a weekly one.
- **Delayed time join:** is applied to find the similarity in shifted-lags. In order to do that, time series are pushed t time pints backward (in the case of the

[††)] https://github.com/Soheyla Mirshahi/ FuzzyMultipleTimeseriesAnalysis

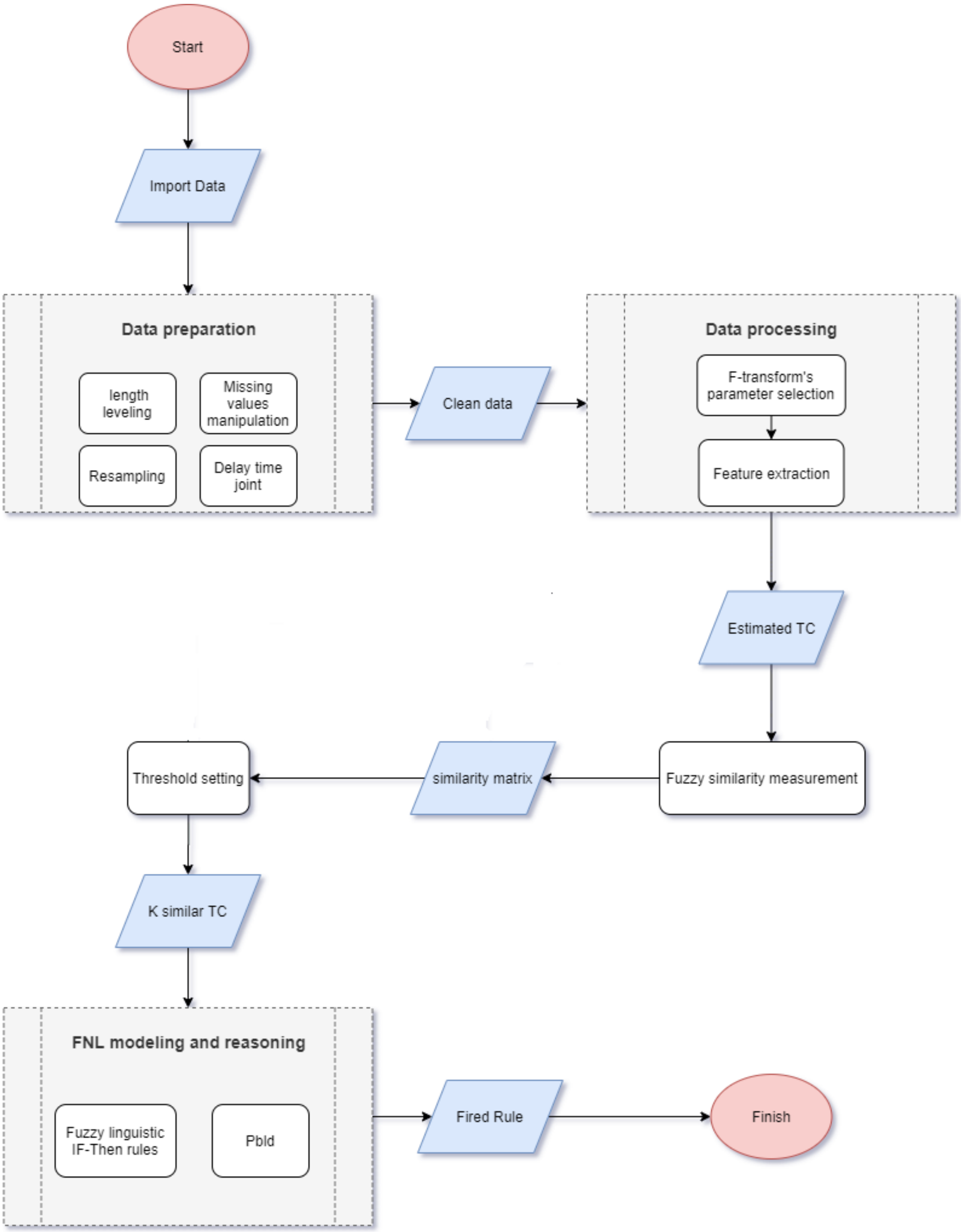


Figure 19: Workflow of the suggested method

stock market, the t is usually small and equal to one). This task is usually used to identify if movement in one time series leads to another one with some delay.

7.2 Data Processing

To process the data, we apply F-transform, which was described in section [5.3.1]. Now, a natural question is raised, how the bandwidth h and the degree m should be chosen to obtain a reliable estimation of the trend-cycle. We have suggested some practical guidance on how to do it:

- *Choosing of the degree m:* this issue depends mainly on the behaviour of the trend-cycle (recall that the trend-cycle is a pseudo-polynomial of the degree m), which can be evaluated by observing the time series data. The more the time series changes its course (or has higher volatility), the higher the degree m should be chosen. It seems that m can be a big positive integer to ensure that the trend-cycle is a pseudo-polynomial of degree m. However, by the fact that the lower degree m, the better reduction of irregular fluctuation (i.e., the smaller values of $|\hat{R}_h^m(t)|$, $t \in \mathbb{R}$) (-see [134, 135, 136,]), we have to choose m as small as possible. A rule of thumb says that if the observed trend-cycle is a nearly linear function then we should choose $m = 0$ or $m = 1$; otherwise, choose $m = 2$ or $m = 3$.

- *Choosing of the bandwidth h:* let $\tilde{R}_h(t) = X(t) - \hat{X}_h^m(t)$ be the residuum of the time series after subtracting the inverse fuzzy transform. The bandwidth should be set to a value h_0, which assures that $\tilde{R}_{h_0}(t)$ satisfies the assumptions of the irregular fluctuation. Namely, it is a realization of a stationary process with zero mean, and its sample autocorrelation functions $\hat{\rho}_{h_0}$. The latter condition can be described as follows:

  $$\sum_{k=0}^{N_{max}} |\hat{\rho}_{h_0}(k)| = \min \left\{ \sum_{k=0}^{N_{max}} |\hat{\rho}_h(k)| \mid h = 1, 2, \ldots, H_{max} \right\},$$

  where N_{max} and H_{max} are two positive integers chosen by the users and $\hat{\rho}_h(k) = \frac{\hat{\gamma}_h(k)}{\hat{\gamma}_h(0)}$ with

  $$\hat{\gamma}_h(k) = \frac{1}{T} \sum_{t=0}^{T-k} \left[\tilde{R}_h(t+k) - M_h\right] \cdot \left[\tilde{R}_h(t) - M_h\right],$$

  $$M_h = \frac{1}{T} \sum_{t=0}^{T} \tilde{R}_h(t),$$

  where T is the dimension of the data. Note that there are several techniques to test the stationarity property of $\tilde{R}_{h_0}(t)$, for example, Augmented Dickey-Fuller test, Kwiatkowski-Phillips-Schmidt-Shin test, and other ones.

More examples about its applicability in financial markets are available in our paper [137].

7.3 Fuzzy similarity assessment

After data preparation and data processing, it is possible to evaluate similarity among the time series of a given data set. In section [6] we described two similarity measures in detail. Moreover, we demonstrated their capability in measuring dependency among financial time series. As we mentioned earlier, the proposed methods are under the assumption that a time series can be additively decomposed into a trend-cycle and an irregular fluctuation. In the suggested methods, first, we assign to each time series an adjoint one that consists of a sequence of trend-cycle of a time series estimated using fuzzy transform. Then we measure the distance between local trend-cycles. The output of this stage is a heat map containing similarity coefficients. At this point, there are two important questions to be answered.

- Similarity degree threshold γ: as we mentioned before, the similarity degree $S(F_1, F_2) \in [0, 1]$. Therefore, it is important to establish a γ criterion as a threshold for determining if two time series are sufficiently similar to warrant further investigation. We set the $\gamma =$ 0.7. As a result, if $S(F_1, F_2) >$ 0.7, we would consult experts to evaluate whether the detected dependency is legitimate.
- Number of similar time series K: there might exist the number of financial time series with $S(F_1, F_2) > \gamma$ is large. Considering all of the time series in the antecedent part of the rule, add complexity to the rule base and reduce the readability.

The output of this phase is time series with dependency relation where the leading and the lagging time series are in antecedent and consequent of the rules, respectively.

7.4 FNL modeling and reasoning

This phase is responsible for evaluating the relationship among time series in a natural language based on the theory of evaluative linguistics description that we mentioned in section [5.4]. To our knowledge, most methods such as vector autoregressive, vector autoregressive integrated moving averages, or multivariate GARCH, define the relationship among time series with a mathematical or statistical formula. These methods are beneficial and precise; however, they are not transparent. The suggested method provides a full interpretable description of the relationships among multiple time series. Using fuzzy evaluative linguistic description theory, we model the relation between our time series by a set of fuzzy linguistic IF-THEN rules. Assume that as a result of the previous step, we discovered that financial time series $F_{1(t)}$ is dependant on its own lag $F_{1(t-m)}$ and lag of another time series $F_{2(t-m)}$. Note that in practice, the m is usually equal to 1, meaning that the stock market can have a dependency on the previous day's value and not days before. The reason for that is the efficient market theory mentioned in section [1]. In a perfectly efficient market, every trader/investor has access to the same information and affects the price immediately. As a result, lead-lag relations between stocks should

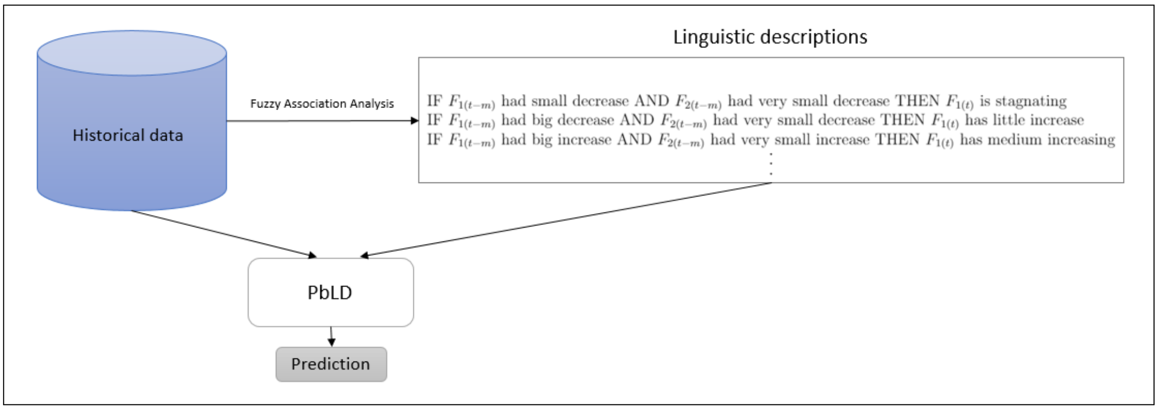


Figure 20: Diagram representing the process

not be feasible. However, as we mentioned earlier, many reasons, such as a global pandemic, change in a country's policies and rules, a shift in market structures, etc., might form a new lead-lag relation between two stocks or a previous dependency disappeared. Therefore, it is important to evaluate the similarity in a longer period of time to ensure the lead-lag relation continues. Clearly, if this lead-lag relationship actually exists, it is very beneficial for participants in the market. Using the historical data, we can extract the linguistic description (rule base) for this lead-lag relation. This linguistic description can be formed using various algorithms; this step is called fuzzy association analysis. After acquiring a new input, we use the mined Linguistic description to predict the future value of the $F_{1(t)}$ using PbLD. Figure 20 shows a schematic of the FNL modeling and reasoning process. For more detailed information see [138].

8 Illustration

In this section, we present an extensive experiment. As we mentioned in the introduction section, there are two main goals in multivariate time series analysis. The primary goal is to find proper relations among time series and afterwards utilize them to adjust the forecast. Hence, in section [6] we provided several examples demonstrating the applicability and strength of the suggested similarity methods in finding similar behaviour between stocks at lag zero and shifted lag(s). The second goal of multivariate time series analysis is to employ the finding relation for multivariate prediction; therefore, we will conduct an experiment on a more extensive data set and provide a predictive linguistic description of the multivariate relations.

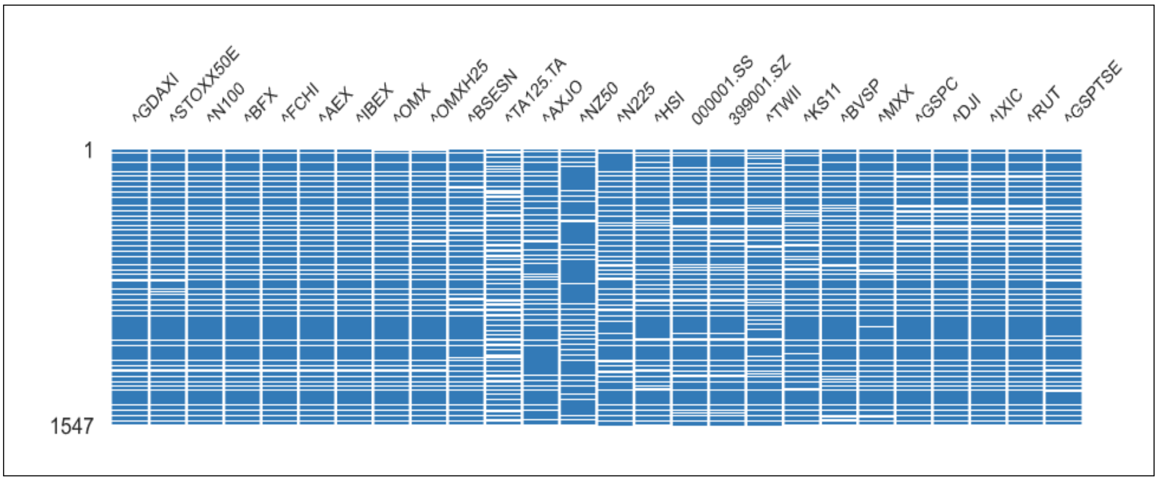
Our dataset contains the daily closing price of 26 global stock indexes (indices) from January 2016 to December 2020 collected from yahoo finance [‡‡)]. A stock index is an indicative index for a set of stocks. This index is calculated using the prices of specific stocks, and its movement can represent the aggregate performance of the stocks in the index. To study various stock markets, geographically, politically, and with different trading volumes, we select these indices from six different world regions, named the US and Canada, Latin America, East Asia, Oceania, south and west Asia, and Europe. Appendix A contains the names of the indices used and the related countries.

8.1 Data preparation

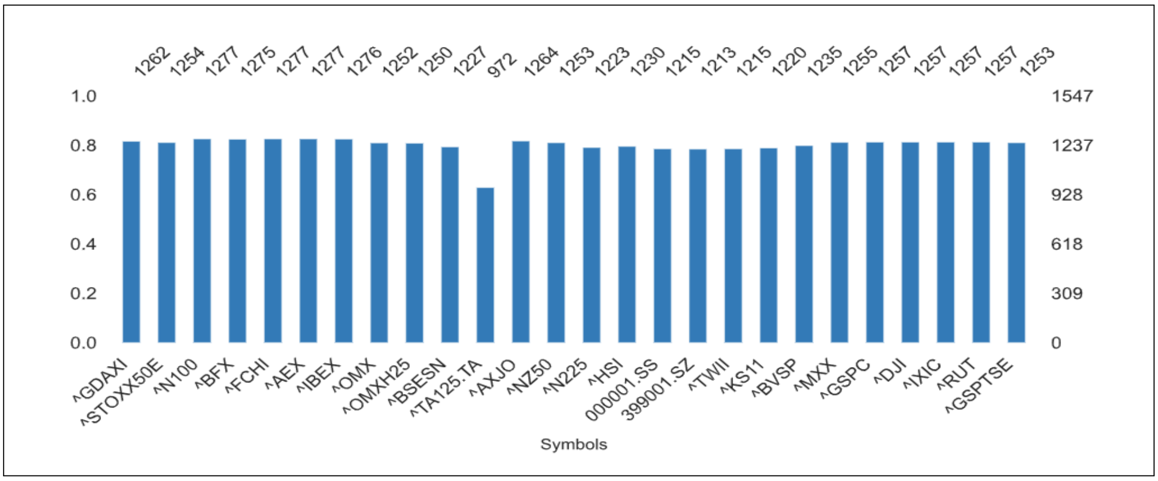
The first step of the suggested workflow is data preparation. Thus we firstly need to investigate the quality of data regarding missing values. Figure 21 shows the matrix and bar chart for the missing values where the x-axis and y-axis represent the market symbols and length of each time series, respectively. From this figure, it is clear that a significant number of values are missing. The reason is that 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 use an approach suggested in these papers [139, 140] to define a condition for omitting the data. Therefore, we remove the days were less than 60% of stock markets operate. To fill the missing values for markets that did not operate in filtered days, we repeat its previous day's value. This approach primarily affected the index of Israel deeply, which has different weekends. For these countries, many working days were removed. Even after eliminating data and filling missing values, it appears that five markets have a shorter length. Using the padding technique, we level all 26 time series. Figure 22 shows the clean data set without missing value.

In practice, investors are interested in the profitability of their investment or the daily return of stock markets; thus, we continue our analysis using daily returns as we already demonstrated in section [6]. Note that price returns reflect market regular profitability; therefore, negative and positive returns mean that the trader profited

[‡‡)] www.yahoo.com



a) Matrix of missing values



b) Bar chart of the length of each variable

Figure 21: Overview of the data set

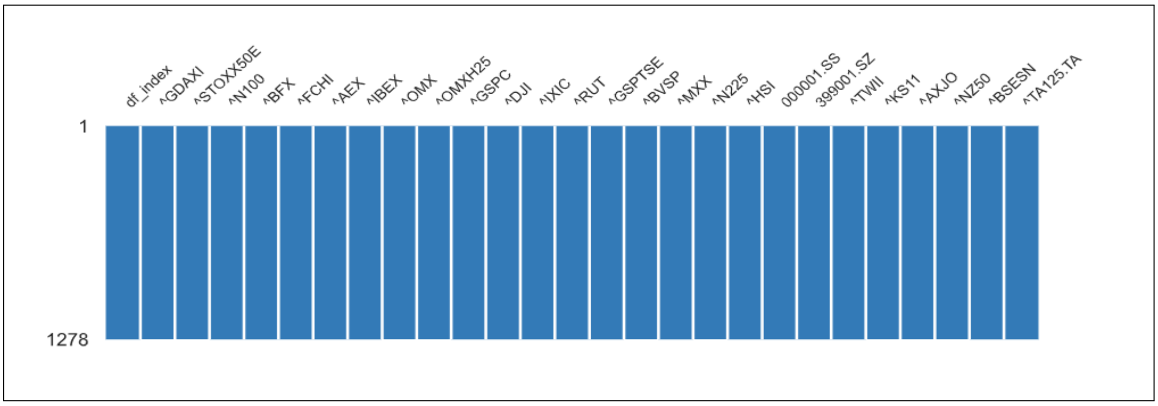


Figure 22: Clean dataset

or lost. Traders usually prefer assets that do not behave similarly. This strategy is known as diversification, and it is used to lower trading risk. As a diversification strategy, some investors choose assets from various markets and industries. However, in reality, it might happen despite the effort; these assets might behave similarly, and in a bullish market (going down), the portfolio value will drop significantly, leading to a considerable loss. On the other hand, if the trader identifies a lead-lag relation outside their portfolio, it can be an excellent indicator for trading the portfolio; hence, there is no doubt about the importance of the suggested method. As we demonstrated in section [6.5], to find the lead-lag relation between international stock indexes, we add one-day lagged values of stocks to the data set. The lagged values are considered as different variables. Thus our database forms an extensive database containing both same-day and previous-day return values in a total of 52 time series.

8.1.1 Data processing

As we explain in the previous section, stock markets are susceptible to news; thus, any shock immediately affects the market. We have shown before that F-transform reduces the outliers, which is very important in analyzing the financial market. Therefore, we apply F-transform to both stock data and their lagged values, so the data is ready for similarity assessment, which is the next step.

8.2 Fuzzy similarity measurement

Using the similarity assessment method described in section [6], we evaluate the similarity among original data (same day) and their lagged values (previous day). Figure 23 demonstrates the results as an enlarged similarity matrix where the indices are arranged from 0 to 25 and their lagged indices from 26 to 52. The color represents the degree of similarity; thus, darker shades represent higher similarity. Moreover, the heat map is divided into four sections. In sections 1 and 3, the same-day similarity is defined, while sections 2 and 4 represent the previous-day and same-day similarity. To show the fuzzy similarity for each region, Figure 24

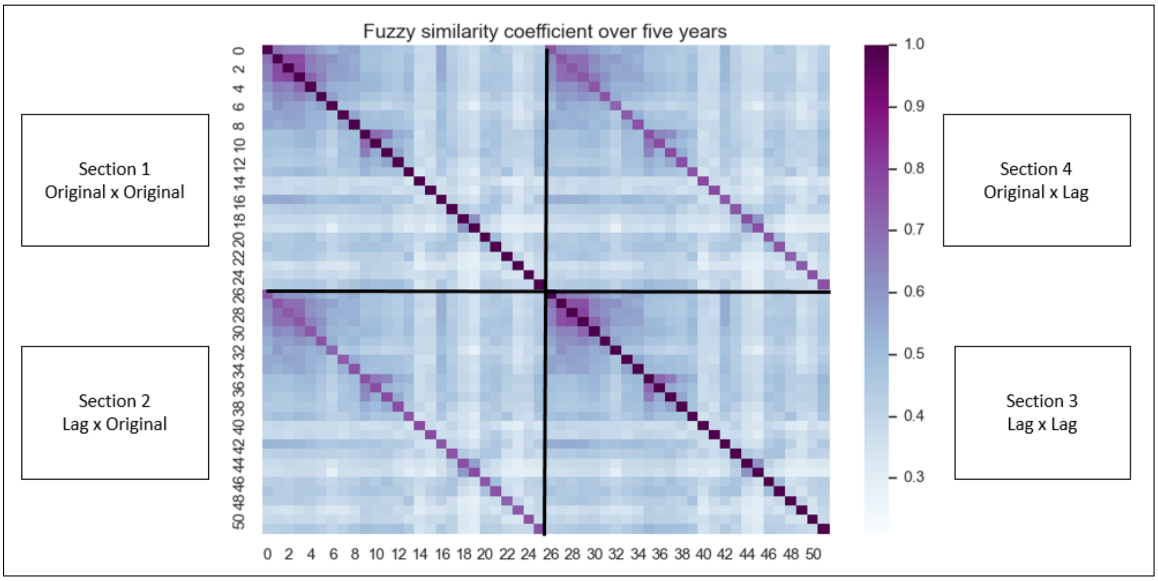


Figure 23: Heat map of the enlarged similarity matrix of both original and lagged indices.

is prepared. 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. Primer analysis shows that the whole European region, the American/Canadian region, and a small section of East Asia form relatively interesting similarity clusters. We will further investigate the relationship between the European region since this region has a bigger cluster to gain more insight. Figure 25 demonstrate the five-year similarity for the European region. This result is already valuable since section 1 contains information about the markets that behave similarly in Europe. Thus it can be used as another indicator for portfolio diversification. Moreover, section 2 manifests the potential similarity with one day delay, which can be used as a guidance for prediction. There is another interesting piece of information in figure 25. Our finding shows that France and Euronext 100 have high similarity for both same day and previous day.

As we stated earlier, the stock market is very dynamic; thus, the dependency relation between two stocks might change. One critical question to answer is whether 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 26 shows the similarity between European markets for each year. As influenced by several political and economic factors, markets behave differently; thus, relations change. For instance, in general, the European markets behave more similarly in 2018. One possible reason might be the US-China trade war started from the beginning of January 2018, when the US began to impose sanctions on Chinese companies, affecting the European economy as well. Note that the overall similarity in the European region decreased in 2020. Early this year, the worldwide pandemic highly affected all the markets globally. Stock markets initially hit harshly; however, after a certain period of time, each country initiated unique rules and policies such as stimulus packages to run the economy while keeping the

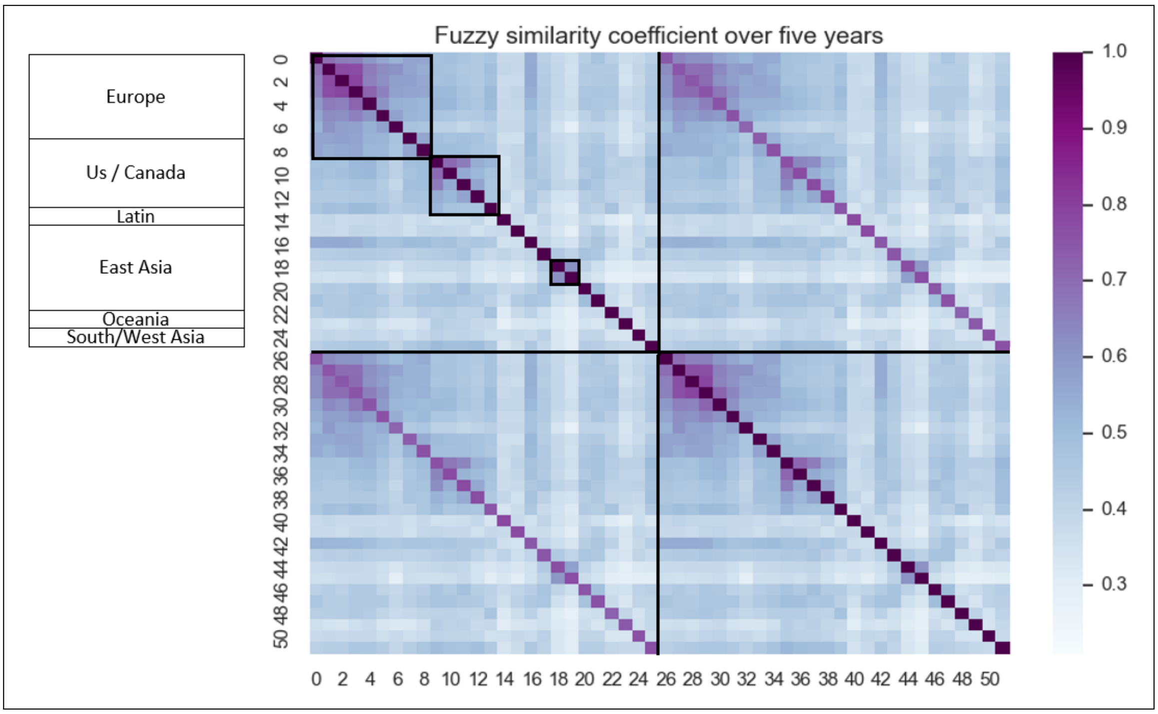


Figure 24: Heatmap of fuzzy similarity based on estimated trend-cycle for different regions.

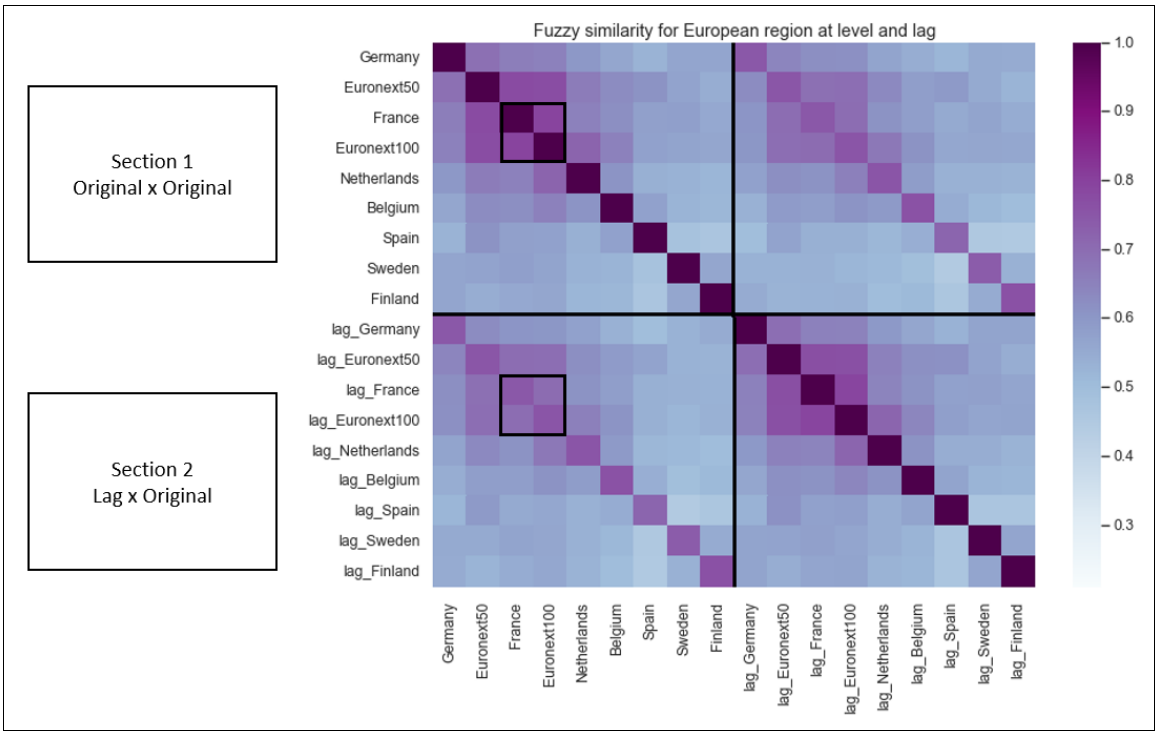


Figure 25: Heatmap of fuzzy similarity between European indices at level and lag from 2016 to 2020

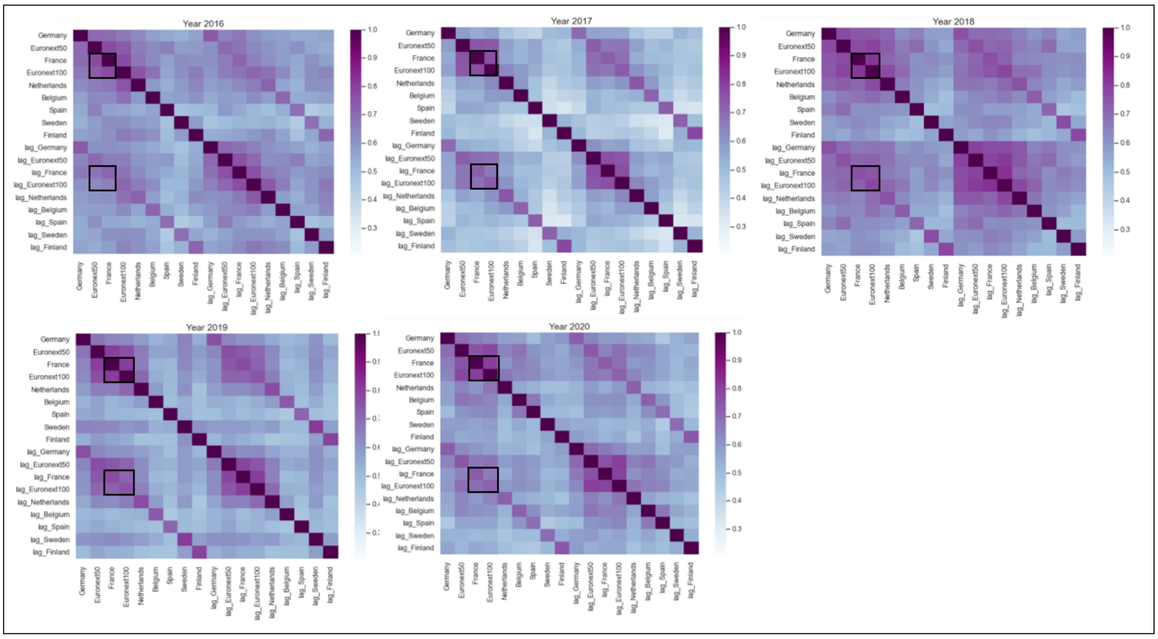


Figure 26: Heatmap of fuzzy similarity in lagged Euronext100 and France from 2016 to 2020

population safe. As a result, the effect of these policies was unconventional, leading to distinct behaviour for each market. Nevertheless, our finding shows that although the relation between France and the Euronext 100 varies over the five years, their pairwise similarity remains high. Since we are interested in the lead-lag relation, the similarity degree between lagged values of Euronext 100 and current values of France is presented in the barchart 27. The similarity varies between 65% to 72%, which is relatively high for stock markets. For the next step, we select the data from 2018 to 2020, which has the highest similarity and is more recent.

8.2.1 FNL modeling and forecasting

In the previous stage, we provided a heatmap showing the markets that behave similarly for the same day and the previous day, which can be used for portfolio diversification and provide some hints for prediction. This section provides a linguistic describing of these relationships in natural language, which can also be necessary for investors.

Each investor has an individual perspective toward evaluating the performance of the stocks. Undoubtedly, every individual has a different tolerance for risk; thus, every investor has a different standard for assessing stock performance, diversification plans, and investment strategies. A conservative investor may assess an average return of 3% as high, while another investor considers it a low return. As a result, the proper context for a stocks performance depends on each investor and their risk tolerance. Therefore, we asked three experts to define a customed context for our targeted time series at this stage. This is a unique feature of the proposed approach empowering investors to customize the method according to their strategy.

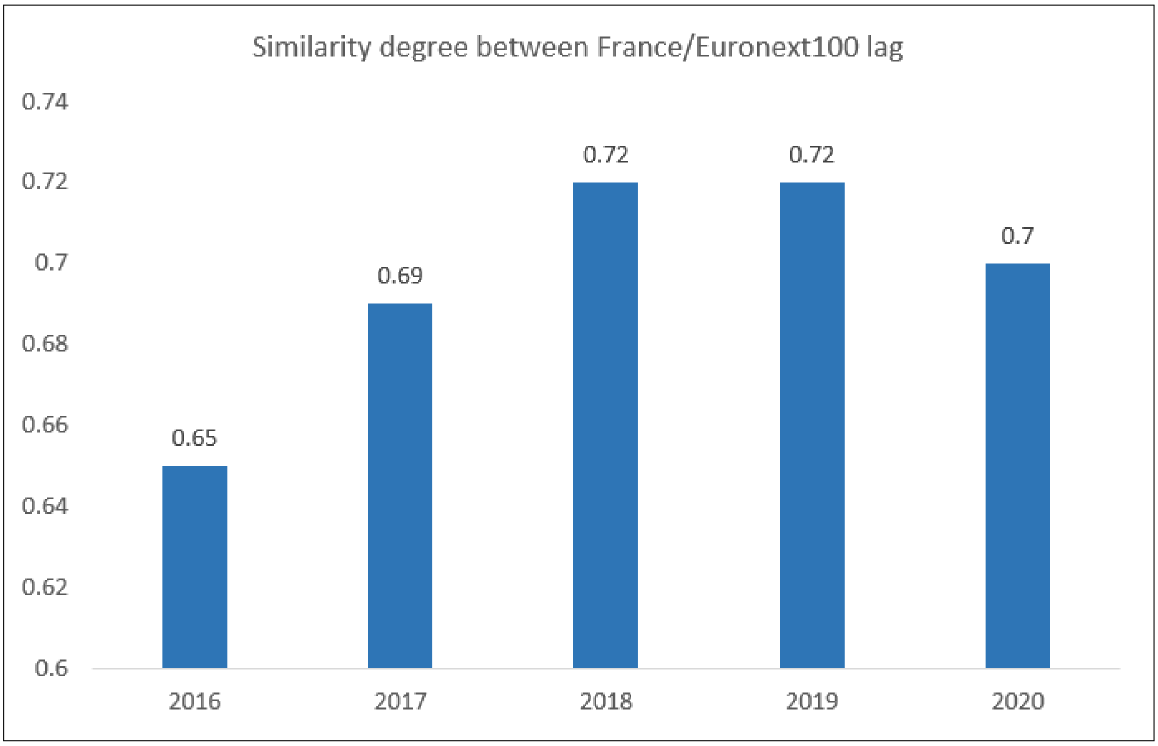


Figure 27: Similarity degree for France and lag-Euronext100 from 2016 to 2020

Using this context, we then create the Fuzzy linguistic IF-THEN rules to define the relationship between France and Euronext 100. As we showed previously, France is dependant on the lag values of Euronext and itself. Thus the general form of the rule base is as follows:

$$\mathscr{R}_j := \text{IF lag-Euronext100 is } \mathscr{A}_i \text{AND lag-France is } \mathscr{A}_j \text{ THEN } France \text{ is } \mathscr{B}_j. \quad (40)$$

Our goal is to find predictive patterns among data. Using the method outlined in the previous section [5.4], we have obtained a fuzzy linguistic description with over 300 rules. This fuzzy rule base is fairly broad, with some redundant rules. We ask the experts to verify and choose the rules and reduce the rule base. According to the findings, all rules can be classified into three groups. The first group suggests that a big increase in previous days results in a big boost in the French market the following day. The second one indicates that a small decrease in previous days will continue the following day, and the last group denotes that when markets stagnate, they will remain so for another day. The example of rules from the linguistic definition is shown in table 8.

We employ PbLD as the inference mechanism to estimate a proper prediction from the reduced linguistic description. Therefore, for new values for *lag-Euronext 100* and *lag-France*, PbLD generates the appropriate rule and the proper value for *France*. Note that our rule base has been adjusted by experts where rules can be categorized within three groups. When the new inputs fall into these categories, we can achieve fully transparent results using PbLD. Because of the peculiar characteristics of financial markets, accuracy in forecasting is not our primary objective. Instead,

lag-Euronext100	&	**lag-France**	–>	**France**
ex bi	&	ex bi	–>	very bi
ro bi	&	bi	–>	ro bi
ro neg sm	&	neg sm	–>	ro neg.sm
neg sm	&	ve neg sm	–>	neg.sm
ro ze	&	ze	–>	ze
ex neg sm	&	ze	–>	ro ze

Table 8: An example of linguistic description describing relation between Euronext100 and France

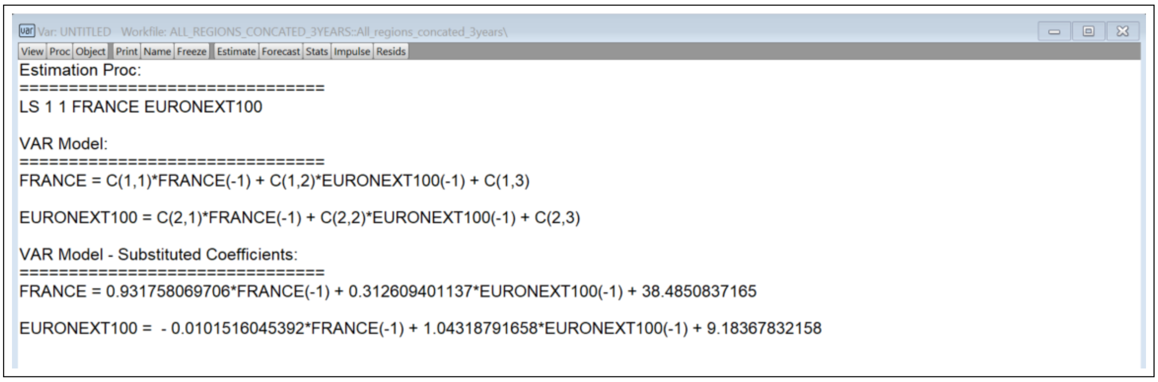


Figure 28: Vector autoregressive model for the data set

our priority is to provide a data-driven rule base that is completely transparent and understandable to investors/traders and can be tailored based on their expertise and strategies. To the best of our knowledge, we are the first to provide a fuzzy linguistic rule base for multivariate financial time series analysis that is completely transparent and personalized based on the expertise of the investors/traders; therefore, a comparison is not legitimate. Nevertheless, we also represent the dependency between France and Euronext 100 market indices using the vector autoregressive model in Eviews software. Figure 28 demonstrate the results. Both indices (France and Euronext 100) can be functions of each other here. However, we can see that France is more dependent on the Euronext 100 than vice versa. This finding supports our finding of their relationship. It should be noted that this is a typical vector autoregressive model representation; the variable dependency is described in a system of classical mathematical equations. One of the key drawbacks of the vector autoregressive models is that the coefficients are unclear and cannot be explained by experts. In the following section, we go through the discrepancies between the proposed method and other models in greater depth.

9 Discussion

Many analysts, academics, and investors have long been interested in stock market forecasting. Because of the number of variables involved, predicting stock prices is a complex problem. According to the efficient market theory, stock prices are simply a random walk. They result from the related information about a company finance and reasonable expectations of the investors. Since newly disclosed information regarding a company prospects is almost immediately reflected in the current stock prices, attempting to forecast them is a fool's game. However, markets participant are not always acting rationally and do not have logical expectations. Sentiment plays an essential role in the behaviour of the markets resulting in a short-term disconnection between available information and market price. This phenomenon was so much visible and influential in 2020. The pandemic transformed 2020 into a year of extraordinary events, leading to the stock markets rapid collapse and then record-breaking recovery. During the pandemic, many businesses have been struggling for survival, the unemployment rate has been high, and economic indicators have shown a very sluggish growth. Nonetheless, the market has risen analogically, fueled by hopes of a near-future time when vaccinations are widely spread, and the economy completely reopens. Many events made 2020 a unique year. Therefore relying only on historical data for stock prediction and not including the expertise of the investors looks pretty unrealistic.

As we mentioned earlier, to improve their chances of earning higher returns, investors must understand the nature of individual stocks and the factors that influence stock prices. Predictions based solely on the stock history itself would be easy, but the result may be misleading as particularly in financial markets, other factors could also be relevant in understanding stock movement. As a result, multivariate models are needed in the context of financial markets. Vector autoregressive and its derivatives are among the most influential models for multivariate financial time series analysis and forecasting. There is no doubt that these models are very efficient. However, they have been criticized for being unconcerned with theory; that is, they are not based on any economic theory that imposes a theoretical framework on the equations. Any variable in the method is supposed to affect any other variable, making direct interpretation of the estimated coefficients difficult. Despite this, vector autoregressive models are helpful in a variety of situations such as forecasting a set of similar variables without clear interpretation; checking whether one variable is useful in forecasting another (based on Granger causality tests), or impulse response analysis; to examine the response of one variable to a sudden but temporary shift in another variable [35]. Vector autoregressive models impose unidirectional relationships. In these relationships, the predictor variables influence the follower variables but not vice versa. This has been noted as one of the limits of vector autoregressive models, which also applies to our proposed method. Nevertheless, the suggested method overcomes the second limitation of vector autoregressive models, namely the difficulty in the model interpretation. Our suggested method is completely interpretable, easy to understand, and explained by experts.

Another prevalent and state-of-the-art model for financial time series analysis and forecasting are neural networks. They have several advantages when applied for financial time series prediction. They are powerful in detecting nonlinear patterns within the data, which is very important for stock markets. They have been reported to have higher prediction accuracy compared to other classical methods. However, this high precision is often exacerbated by overfitting, which raises the risk of investment because it can confuse investors. Secondly, they are computationally costly and subject to parameter selection. Finally, in the case of stock market predictions, investors are confronted with a black box that brings uncertainty to an already unpredictable market. In practice, traders/investors already struggle with a high degree of uncertainty and risk. As a result, although neural networks function well in theory, they still need to be improved to reach to their highest potential in stock market analysis. The proposed approach has some limitations as well.

We have already mentioned one drawback, which is similar to vector autoregressive models. Furthermore, it is fair to say that the presented method is more beneficial for detecting the *predictive relationship* among multiple stocks than for the value prediction itself. Thus, in the case of forecasting, the model uniquely predicts one step if the input falls within the linguistic description. On the other hand, 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 investors think and reason in practice and is highly interpretable. The proposed approach is resistant to outliers. Since financial markets are susceptible to many outliers, this feature is critical. Furthermore, the proposed method can be tailored to each individual investor based on their specific strategies. Finally, it allows investors to express their knowledge and expertise using natural language, which is then incorporated in the linguistic description. This is yet another intuitive and beneficial feature.

10 Conclusion

This thesis began with a study of the literature on financial time series analysis. Several approaches were thoroughly described, and the literature was summarized in a table. Later, we presented the theoretical theories required for the proposed method and explained them in-depth to provide a better understanding of the suggested approach. Following that, we developed and mathematically proved two similarity methods for assessing stock similarity. In addition, several detailed studies were carried out to show the efficacy of the proposed approach in practice and to compare it to other approaches. We later proposed a systematic methodology flowchart for multivariate financial time series analysis and prediction. Python (a high-level programming language and the most common language for data analysis at the moment) is used to automate this flowchart fully. The code for the proposed model has been uploaded and made available on Github, one of the most authoritative code-sharing repositories, to ensure its usability. This enables other data analysts to refine and adapt the data to their specific requirements rapidly. Finally, we conducted another in-depth experiment with comprehensive data collection from around the world. Over a five-year period, the relationship between different stocks was studied. A predictive rule based on that study is developed and validated by experts. In the final section, we discussed the benefits and drawbacks of the proposed method as well as other similar methods. This thesis contributed to both theoretical and applied research in fuzzy methods for financial time series analysis.

References

[1] W. I. D. Mining, Data mining: Concepts and techniques, Morgan Kaufinann.

[2] E. Keogh, S. Kasetty, On the need for time series data mining benchmarks: a survey and empirical demonstration, Data Mining and knowledge discovery 7 (4) (2003) 349–371.

[3] T.-c. Fu, A review on time series data mining, Engineering Applications of Artificial Intelligence 24 (1) (2011) 164–181.

[4] T. W. Liao, Clustering of time series data—a survey, Pattern recognition 38 (11) (2005) 1857–1874.

[5] J. Han, J. Pei, M. Kamber, Data mining: concepts and techniques, Elsevier, 2011.

[6] Q. Yang, X. Wu, 10 challenging problems in data mining research, International Journal of Information Technology & Decision Making 5 (04) (2006) 597–604.

[7] J. Serra, J. L. Arcos, An empirical evaluation of similarity measures for time series classification, Knowledge-Based Systems 67 (2014) 305–314.

[8] I. Perfilieva, Fuzzy transforms: theory and applications, Fuzzy Sets and Systems 157 (2006) 993–1023.

[9] I. Perfilieva, F-transforms — Where fuzzy and classical mathematics meet, Applied and computational mathematics 10 (2011) 122–132.

[10] V. Novák, I. Perfilieva, M. Holčapek, V. Kreinovich, Filtering out high frequencies in time series using f-transform, Information Sciences 274 (2014) 192–209.

[11] 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.

[12] 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.

[13] M. Štěpnička, A. Dvŏrák, V. Pavliska, A linguistic approach to time series modeling with the help of F–transform, Fuzzy Sets and Systems 180 (2011) 164–184.

[14] V. Novák, I. Perfilieva, V. Krejnovich, F-transform in the analysis of periodic signals, in: M. Inuiguchi, Y. Kusunoki, M. Seki (Eds.), Proc. 15^{th} Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty, Osaka University, Osaka, Japan, 2012, pp. 150–158.

[15] 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.

[16] V. Novák, I. Perfilieva, M. Holčapek, V. Kreinovich, Filtering out high frequencies in time series using F-transform, Information Sciences 274 (2014) 192–209.

[17] 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.

[18] V. Novák, I. Perfilieva, A. Dvorak, Insight into Fuzzy Modeling, John Wiley & Sons, 2016.

[19] 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.

[20] V. Novák, I. Perfilieva, A. Romanov, N. Yarushkina, Time series grouping and trend forecast using F^1-transform and fuzzy natural logic, in: R. Marco se Moraes, E. E. Kerre, L. dos Santos Machado, J. Lu (Eds.), Decision Making and Soft Computing, World Scientific, 2014, pp. 143–148.

[21] P. Rusnok, Linguistic descriptions as a modeling tool for multivariate time series, in: 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT-15), Atlantis Press, 2015.

[22] A. Nanopoulos, R. Alcock, Y. Manolopoulos, Feature-based classification of time-series data, International Journal of Computer Research 10 (3) (2001) 49–61.

[23] E. F. Fama, The behavior of stock-market prices, The journal of Business 38 (1) (1965) 34–105.

[24] F. E. Tay, L. Cao, Application of support vector machines in financial time series forecasting, omega 29 (4) (2001) 309–317.

[25] 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.

[26] X. Zhong, D. Enke, Forecasting daily stock market return using dimensionality reduction, Expert Systems with Applications 67 (2017) 126–139.

[27] 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.

[28] P. C. S. Bezerra, P. H. M. Albuquerque, Volatility forecasting via svr–garch with mixture of gaussian kernels, Computational Management Science 14 (2) (2017) 179–196.

[29] M. Kumar, M. Thenmozhi, Forecasting stock index returns using arima-svm, arima-ann, and arima-random forest hybrid models, International Journal of Banking, Accounting and Finance 5 (3) (2014) 284–308.

[30] M. Göçken, M. Özçalıcı, A. Boru, A. T. Dosdoğru, Integrating metaheuristics and artificial neural networks for improved stock price prediction, Expert Systems with Applications 44 (2016) 320–331.

[31] R. S. Tsay, Analysis of financial time series, Vol. 543, John wiley & sons, 2005.

[32] B. G. Malkiel, E. F. Fama, Efficient capital markets: A review of theory and empirical work, The journal of Finance 25 (2) (1970) 383–417.

[33] Y. S. Abu-Mostafa, A. F. Atiya, Introduction to financial forecasting, Applied intelligence 6 (3) (1996) 205–213.

[34] T.-l. Chen, F.-y. Chen, An intelligent pattern recognition model for supporting investment decisions in stock market, Information Sciences 346 (2016) 261–274.

[35] R. J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, OTexts, 2018.

[36] R. Fildes, P. Goodwin, et al., Good and bad judgement in forecasting: Lessons from four companies, Foresight 8 (2007) 5–10.

[37] M. Hiransha, E. A. Gopalakrishnan, V. K. Menon, K. Soman, Nse stock market prediction using deep-learning models, Procedia computer science 132 (2018) 1351–1362.

[38] G. Box, G. Jenkins, Control, Halden-Day, San Francisco.

[39] J.-J. Wang, J.-Z. Wang, Z.-G. Zhang, S.-P. Guo, Stock index forecasting based on a hybrid model, Omega 40 (6) (2012) 758–766.

[40] Y.-S. Chen, C.-H. Cheng, W.-L. Tsai, Modeling fitting-function-based fuzzy time series patterns for evolving stock index forecasting, Applied intelligence 41 (2) (2014) 327–347.

[41] R. C. Cavalcante, R. C. Brasileiro, V. L. Souza, J. P. Nobrega, A. L. Oliveira, Computational intelligence and financial markets: A survey and future directions, Expert Systems with Applications 55 (2016) 194–211.

[42] K.-j. Kim, Financial time series forecasting using support vector machines, Neurocomputing 55 (1-2) (2003) 307–319.

[43] P.-C. Chang, C.-Y. Fan, A hybrid system integrating a wavelet and tsk fuzzy rules for stock price forecasting, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 38 (6) (2008) 802–815.

[44] V. Novák, M. Štěpnička, A. Dvořák, I. Perfilieva, V. Pavliska, L. Vavříčková, Analysis of seasonal time series using fuzzy approach, International Journal of General Systems 39 (3) (2010) 305–328.

[45] L. Nguyen, V. Novák, S. Mirshahi, Trend-cycle estimation using fuzzy transform and its application for identifying bull and bear phases in markets, Intelligent Systems in Accounting, Finance and Management.

[46] Y. Kara, M. A. Boyacioglu, Ö. K. Baykan, Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange, Expert systems with Applications 38 (5) (2011) 5311–5319.

[47] A. A. Adebiyi, C. Ayo, M. O. Adebiyi, S. Otokiti, Stock price prediction using neural network with hybridized market indicators, Journal of Emerging Trends in Computing and Information Sciences 3 (1) (2012) 1–9.

[48] J. Patel, S. Shah, P. Thakkar, K. Kotecha, Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques, Expert systems with applications 42 (1) (2015) 259–268.

[49] W. Huang, Y. Nakamori, S.-Y. Wang, Forecasting stock market movement direction with support vector machine, Computers & operations research 32 (10) (2005) 2513–2522.

[50] X.-Y. Qian, S. Gao, Financial series prediction: Comparison between precision of time series models and machine learning methods, arXiv preprint arXiv:1706.00948.

[51] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[52] S. Kim, S. Ku, W. Chang, J. W. Song, Predicting the direction of us stock prices using effective transfer entropy and machine learning techniques, IEEE Access 8 (2020) 111660–111682.

[53] L. Khaidem, S. Saha, S. R. Dey, Predicting the direction of stock market prices using random forest, arXiv preprint arXiv:1605.00003.

[54] D. Selvamuthu, V. Kumar, A. Mishra, Indian stock market prediction using artificial neural networks on tick data, Financial Innovation 5 (1) (2019) 16.

[55] S.-L. Ho, M. Xie, T. N. Goh, A comparative study of neural network and box-jenkins arima modeling in time series prediction, Computers & Industrial Engineering 42 (2-4) (2002) 371–375.

[56] J. L. Elman, Finding structure in time, Cognitive science 14 (2) (1990) 179–211.

[57] D. G. Torres, H. Qiu, Applying recurrent neural networks for multivariate time series forecasting of volatile financial data (2018).

[58] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.

[59] D. M. Nelson, A. C. Pereira, R. A. de Oliveira, Stock market's price movement prediction with lstm neural networks, in: 2017 International joint conference on neural networks (IJCNN), IEEE, 2017, pp. 1419–1426.

[60] K. Chen, Y. Zhou, F. Dai, A lstm-based method for stock returns prediction: A case study of china stock market, in: 2015 IEEE international conference on big data (big data), IEEE, 2015, pp. 2823–2824.

[61] S. McNally, Predicting the price of bitcoin using machine learning, Ph.D. thesis, Dublin, National College of Ireland (2016).

[62] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, European Journal of Operational Research 270 (2) (2018) 654–669.

[63] W. Bao, J. Yue, Y. Rao, A deep learning framework for financial time series using stacked autoencoders and long-short term memory, PloS one 12 (7) (2017) e0180944.

[64] T. C. Mills, R. N. Markellos, The econometric modelling of financial time series, Cambridge University Press, 2008.

[65] R. S. Tsay, Multivariate time series analysis: with R and financial applications, John Wiley & Sons, 2013.

[66] C. A. Sims, Macroeconomics and reality, Econometrica: journal of the Econometric Society (1980) 1–48.

[67] C. Hanck, M. Arnold, A. Gerber, M. Schmelzer, Introduction to econometrics with r, Obtenido de https://www. econometrics-with-r. org/ITER. pdf.

[68] R. F. Engle, C. W. Granger, Co-integration and error correction: representation, estimation, and testing, Econometrica: journal of the Econometric Society (1987) 251–276.

[69] R. S. Tsay, Financial time series, Wiley StatsRef: Statistics Reference Online (2014) 1–23.

[70] T.-C. Fu, A review on time series data mining, Engineering Applications of Artificial Intelligence 24 (2011) 164–181.

[71] D. Q. Goldin, P. C. Kanellakis, On similarity queries for time-series data: constraint specification and implementation, in: International Conference on Principles and Practice of Constraint Programming, Springer, 1995, pp. 137–153.

[72] R. Agrawal, C. Faloutsos, A. Swami, Efficient similarity search in sequence databases, in: International conference on foundations of data organization and algorithms, Springer, 1993, pp. 69–84.

[73] D. J. Berndt, J. Clifford, Using dynamic time warping to find patterns in time series., in: KDD workshop, Vol. 10, Seattle, WA, 1994, pp. 359–370.

[74] T. B. N. Y. M. Ozsoyoglu, Matching and indexing sequences of different lengths, in: Proceedings of the Sixth International Conference on Information and Knowledge Management: CIKM'97, November 10-14, 1997, Las Vegas, Nevada, Vol. 6, Citeseer, 1997, p. 128.

[75] R. J. Alcock, Y. Manolopoulos, et al., Time-series similarity queries employing a feature-based approach, in: 7th Hellenic conference on informatics, 1999, pp. 27–29.

[76] E. H. Ruspini, I. S. Zwir, Automated qualitative description of measurements, in: IMTC/99. Proceedings of the 16th IEEE Instrumentation and Measurement Technology Conference (Cat. No. 99CH36309), Vol. 2, IEEE, 1999, pp. 1086–1091.

[77] J. P. Morrill, Distributed recognition of patterns in time series data: from sensors to informed decisions, Communications of the ACM 41 (5) (1998) 45–52.

[78] R. N. Mantegna, H. E. Stanley, Introduction to econophysics: correlations and complexity in finance, Cambridge university press, 1999.

[79] G. Marti, F. Nielsen, M. Bińkowski, P. Donnat, A review of two decades of correlations, hierarchies, networks and clustering in financial markets, arXiv preprint arXiv:1703.00485.

[80] M. Tumminello, C. Coronnello, F. Lillo, S. Micciche, R. N. Mantegna, Spanning trees and bootstrap reliability estimation in correlation-based networks, International Journal of Bifurcation and Chaos 17 (07) (2007) 2319–2329.

[81] R. N. Mantegna, Hierarchical structure in financial markets, The European Physical Journal B-Condensed Matter and Complex Systems 11 (1) (1999) 193–197.

[82] M. Billio, M. Getmansky, A. W. Lo, L. Pelizzon, Econometric measures of connectedness and systemic risk in the finance and insurance sectors, Journal of financial economics 104 (3) (2012) 535–559.

[83] D. Y. Kenett, M. Tumminello, A. Madi, G. Gur-Gershgoren, R. N. Mantegna, E. Ben-Jacob, Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market, PloS one 5 (12) (2010) e15032.

[84] J. Rocchi, E. Y. L. Tsui, D. Saad, Emerging interdependence between stock values during financial crashes, Plos one 12 (5) (2017) e0176764.

[85] P. Fiedor, Information-theoretic approach to lead-lag effect on financial markets, The European Physical Journal B 87 (8) (2014) 168.

[86] X. Zhang, B. Podobnik, D. Y. Kenett, H. E. Stanley, Systemic risk and causality dynamics of the world international shipping market, Physica A: Statistical Mechanics and its Applications 415 (2014) 43–53.

[87] H. Wang, R. Pappadà, F. Durante, E. Foscolo, A portfolio diversification strategy via tail dependence clustering, in: International Conference on Soft Methods in Probability and Statistics, Springer, 2016, pp. 511–518.

[88] G. Marti, F. Nielsen, P. Donnat, Optimal copula transport for clustering multivariate time series, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2016, pp. 2379–2383.

[89] H. M. Mok, Causality of interest rate, exchange rate and stock prices at stock market open and close in hong kong, Asia Pacific Journal of Management 10 (2) (1993) 123–143.

[90] R. Marschinski, H. Kantz, Analysing the information flow between financial time series, The European Physical Journal B-Condensed Matter and Complex Systems 30 (2) (2002) 275–281.

[91] O. Kwon, J.-S. Yang, Information flow between stock indices, EPL (Europhysics Letters) 82 (6) (2008) 68003.

[92] J. G. Brida, W. A. Risso, Multidimensional minimal spanning tree: The dow jones case, Physica A: Statistical Mechanics and its Applications 387 (21) (2008) 5205–5210.

[93] G. S. Lee, M. A. Djauhari, Multidimensional stock network analysis: An escoufier's rv coefficient approach, in: AIP Conference Proceedings, Vol. 1557, American Institute of Physics, 2013, pp. 550–555.

[94] K. Nam, N. Seong, Financial news-based stock movement prediction using causality analysis of influence in the korean stock market, Decision Support Systems 117 (2019) 100–112.

[95] L. Sidi, Improving s&p stock prediction with time series stock similarity, arXiv preprint arXiv:2002.05784.

[96] A. Yaglom, An Introduction to The Theory of Stationary Random Functions. Revised English ed. Translated and edited by Richard A. Silverman., Englewood Cliffs, NJ: Prentice-Hall, Inc. XIII, 1962.

[97] W. W. Wei, Time series analysis, in: The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2, 2006.

[98] V. Novák, I. Perfilieva, A. Dvorak, Insight into fuzzy modeling, John Wiley & Sons, 2016.

[99] S. Gottwald, Fuzzy Sets and Fuzzy Logic. The Foundations of Application – from a Mathematical Point of View, Vieweg, Braunschweig/Wiesbaden and Teknea, Toulouse, 1993.

[100] V. Novák, I. Perfilieva, J. Močkoř, Mathematical Principles of Fuzzy Logic, Kluwer, Boston, 1999.

[101] I. Perfilieva, M. Daňková, B. Bede, Towards a higher degree f-transform, Fuzzy Sets and Systems 180 (1) (2011) 3–19.

[102] I. Perfilieva, Fuzzy transforms: theory and applications, Fuzzy Sets and Systems 157 (2006) 993–1023.

[103] V. Novák, I. Perfilieva, A. Dvořák, Insight into Fuzzy Modeling, Wiley & Sons, Hoboken, New Jersey, 2016.

[104] I. Perfilieva, M. Daňková, B. Bede, Towards a higher degree F-transform, Fuzzy Sets and Systems 180 (2011) 3–19.

[105] V. Novák, A comprehensive theory of trichotomous evaluative linguistic expressions, Fuzzy Sets and Systems 159 (22) (2008) 2939–2969.

[106] L. A. Zadeh, A fuzzy-set-theoretic interpretation of linguistic hedges.

[107] L. G. Hedges, A study in meaning criteria and the logic of fuzzy concepts, Chicago Linguistic Society Papers 8 (1972) 183–228.

[108] R. Bělohlávek, V. Novák, Learning rule base of the linguistic expert systems, Soft Computing 7 (2002) 79–88.

[109] M. Burda, lfl: Linguistic fuzzy logic (r package on cran), 2015, R package version 1.

[110] V. Novák, S. Mirshahi, V. Pavliska, Lfl forecaster: Analysis, forecasting and mining information from time series, in: 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2019, pp. 1–6.

[111] L. Nguyen, V. Novák, Forecasting seasonal time series based on fuzzy techniques, Fuzzy Sets and Systems.

[112] L. Nguyen, V. Novák, Filtering out high frequencies in time series using F-transform with respect to raised cosine generalized uniform fuzzy partition, in: Proc. Int. Conference FUZZ-IEEE 2015, IEEE Computer Society, CPS, Istanbul, 2015.

[113] J. Anděl, Statistical Analysis of Time Series, SNTL, Praha, 1976 (in Czech).

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

[115] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, E. Keogh, Experimental comparison of representation methods and distance measures for time series data, Data Mining and Knowledge Discovery 26 (2) (2013) 275–309.

[116] M. D. Morse, J. M. Patel, An efficient and accurate method for evaluating time series similarity, in: Proceedings of the 2007 ACM SIGMOD international conference on Management of data, ACM, 2007, pp. 569–580.

[117] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, E. Keogh, Indexing multidimensional time-series, The VLDB Journal 15 (1) (2006) 1–20.

[118] P. E. Wang (Ed.), Computing with Words, J. Wiley, New York, 2001.

[119] L. Kaufman, P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis, Vol. 344, John Wiley & Sons, 2009.

[120] D. L. Davies, D. W. Bouldin, A cluster separation measure, IEEE transactions on pattern analysis and machine intelligence (2) (1979) 224–227.

[121] W. S. Chan, Stock price reaction to news and no-news: drift and reversal after headlines, Journal of Financial Economics 70 (2) (2003) 223–260.

[122] G. Zervas, S. M. Ruger, The curse of dimensionality and document clustering.

[123] R. Roll, Industrial structure and the comparative behavior of international stock market indices, The Journal of Finance 47 (1) (1992) 3–41.

[124] A. F. Herbst, J. P. McCormack, E. N. West, Investigation of a lead-lag relationship between spot stock indices and their futures contracts, Journal of Futures Markets 7 (4) (1987) 373–381.

[125] T. S. Mech, Portfolio return autocorrelation, Journal of Financial Economics 34 (3) (1993) 307–344.

[126] G. McQueen, M. Pinegar, S. Thorley, Delayed reaction to good news and the cross-autocorrelation of portfolio returns, The Journal of Finance 51 (3) (1996) 889–919.

[127] A. W. Lo, A. C. MacKinlay, When are contrarian profits due to stock market overreaction?, The review of financial studies 3 (2) (1990) 175–205.

[128] K. Hou, Industry information diffusion and the lead-lag effect in stock returns, The Review of Financial Studies 20 (4) (2007) 1113–1138.

[129] C. W. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica: journal of the Econometric Society (1969) 424–438.

[130] L. Mandel, E. Wolf, Spectral coherence and the concept of cross-spectral purity, JOSA 66 (6) (1976) 529–535.

[131] C. W. Granger, Testing for causality: a personal viewpoint, Journal of Economic Dynamics and control 2 (1980) 329–352.

[132] 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.

[133] 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.

[134] M. Holčapek, L. Nguyen, Suppression of high frequencies in time series using fuzzy transform of higher degree, in: 16th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2016, Eindhoven, The Netherlands, 2016, pp. 705–716.

[135] M. Holčapek, L. Nguyen, Trend-cycle estimation using fuzzy transform of higher degree, Iranian Journal of fuzzy systems 15 (2018) 23–54.

[136] L. Nguyen, M. Holčapek, Higher degree fuzzy transform: Application to stationary processes and noise reduction, in: In: Advances in Fuzzy Logic and Technology 2017. EUSFLAT 2017, IWIFSGN 2017, Springer, 2018, pp. 1–12. URL `https://doi.org/10.1007/978-3-319-66827-7_1`

[137] L. Nguyen, V. Novák, S. Mirshahi, Trend-cycle estimation using fuzzy transform and its application for identifying bull and bear phases in markets, Intelligent Systems in Accounting, Finance and Management 27 (3) (2020) 111–124.

[138] M. Štěpnička, A. Dvořák, V. Pavliska, L. Vavříčková, A linguistic approach to time series modeling with the help of f-transform, Fuzzy sets and systems 180 (1) (2011) 164–184.

[139] 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.

[140] 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.

Table 9: **List of stock indices**

Number	**Ticker**	**Country**
1	GDAXI	Germany
2	STOXX50E	Euro Zone 50
3	FCHI	France
4	N100	Euro Zone 100
5	AEX	Netherlands
6	BFX	Belgium
7	IBEX	Spain
8	OMX	Sweden
9	OMXH25	Finland
10	GSPC	USA ($S\&P500$)
11	DJI	USA (Dow Jones)
12	IXIC	USA (NASDAQ)
13	RUT	USA(Russell 2000)
14	GSPTSE	Canada
15	BVSP	Brazil
16	MXX	Mexico
17	N225	Japan
18	HSI	Hong Kong
19	000001.SS	China (Shanghai)
20	399001.SZ	China(Shenzhen)
21	TWII	Taiwan
22	KS11	South Korea
23	AXJO	Australian
24	NZ50	New Zealand
25	BSESN	India
26	TA125.TA	Israel