technische universität
dortmund

Essays in Finance: Time-Frequency Decompositions and the Impact of Surprise

Dissertation zur Erlangung des akademischen Grades
Doctor rerum politicarum
der Fakultät für Wirtschaftswissenschaften
der Technischen Universität Dortmund

Timotheos Paraskevopoulos

February 5, 2019

Contents

List of Figures

List of Tables

Acknowledgment

I would like to express my gratitude to Professor Peter N. Posch for his continuous support and guidance in writing this thesis, and for his patience, motivation and the knowledge that has helped me tremendously. I have been fortunate to have a supervisor who gave me the freedom to explore on my own and at the same time the guidance to overcome many difficult situations as they presented.
I would also like to thank the whole research group (Johannes, Janis, Philipp, Vladislav, Gerrit, and Daniel) for their patience, criticism, and support during the entire time as a Ph.D. student, as well as our assistants.
Finally, I would like to thank my family who supported me in many aspects and smoothed the way I have taken. Most of all, I thank my wife Ana Lena for her unconditional support during all the ups and especially during all the downs of this thesis.
Thank you very much.

1 Introduction

The definition and detection of dependence structures is an essential building block in the model generating process. In this respect, revealing non-trivial patterns can be regarded as added value in the clarification of economic dynamical structures which also postulates a positive side-effect in forecasting. Even though the recent success of machine learning, signal processing, and natural language processing hold the key to empower the model generating process, the financial community is far from exploiting their full potential. In this thesis we seek to meet the demand for interdisciplinary techniques.

Predicting an unknown state is conducted based on the analysis of information that is believed to influence the future outcome. In finance, this process is often based on historical time-series data which also postulates the foundation of a model generating process. Its availability and costs determine the frequency at which economic and financial data are collected. The question whether the sampled frequency fully captures the movement of economic activity is of secondary importance. In case the sampled frequency is recognized as inadequate, the discrete time series is aggregated or partitioned on the time scale to obtain the ideal frequency. For example, summing daily logarithmic stock returns of five successive business days converts daily observations to weekly. However, this procedure shrinks the informational content of the time-series as the sequential structure is partially lost. This hampers, additionally to the adverse statistical properties and low signal-to-noise ratio of financial time series, the identification of a proper global model. It is therefore necessary to use a filter that creates linearly independent time-scale transformations without decreasing the informational value of the data.

Summarizing, this thesis focuses on the time-frequency decomposition of time series and temporal aggregation of information in unstructured data. The former is addressed in chapter 2 and 3, where we use wavelet analysis to investigate patterns of interdependence and create frequency-based features for forecasting. The latter is the topic of chapter 4,

as we use an interdisciplinary approach based on topic modeling and network theory to identify surprising events.

In the first part of the thesis, we investigate the effectiveness of an interdisciplinary methodology, known as wavelet analysis, to study temporal and frequency information simultaneously. Wavelets are small waves which have a finite length and are oscillatory as they grow at a point in time and decay at a subsequent point in time. Wavelets can be arbitrarily stretched and translated with a flexible resolution and are therefore localized in both the time and frequency domain which is one of the main advantages against base functions of the Fourier transformation. In comparison with econometric tools which operate solely in the time-domain, this approach can be used to reveal frequency specific features.

In the first analysis we explore the overlap between signal processing and machine learning within an application of time series forecasting and attempt to bridge their gaps. We focus on the decomposition of time-series into the time and frequency domain to alleviate the influence of noise in a forecasting model for stock market indices. The decomposition is conducted using wavelets as previous applications have shown a powerful feature extraction capability. One of the indispensable challenges in wavelet analysis is to choose the wavelet filter adequately. As there is no deterministic solution to this problem, it needs to be addressed by trial and error (Chernick 2001). Moreover, the redundancy of features created by the time-frequency decomposition hampers the model building process and requires an additional feature reduction step without deteriorating the signal-to-noise ratio of the data. We incorporate a recursive elimination algorithm to rank each feature in the context of the learning algorithm. Moreover, to avoid overfitting, we penalize model complexity by structural risk minimization as proposed by Vapnik, Golowich, and Smola (1997). The proposed model is evaluated on five years of daily out-of-sample data covering the worlds major stock indices. The outperformance of our algorithm against the SVR model with non-filtered forecasting features in the classification task supports the hypothesis that specific wavelet decompositions do reveal reoccurring patterns.

In the third chapter we make use of time-frequency decompositions to adopt a long- and short-term perspective on market linkages between the nominal effective European exchange rate and European stock markets. We estimate the spectral characteristics of the time series as a function of time and frequency using a continuous wavelet

transform. In a second step, we compute the continuous wavelet power spectrum and three cross-wavelet tools: the cross-wavelet power spectrum, the cross-wavelet coherency and the phase-difference of the complex wavelet coherency. This framework allows us straightforwardly to reveal how periodic components of the considered time series evolve over time. Moreover, we can include measures of co-movement between the foreign exchange market and the European stock market. Computing the phase-difference, we can identify lead-lag dynamics across frequencies for the observed co-movement patterns. The tools we apply reveal economic time-frequency relations that have cannot be captured by classical correlation or covariance-based approaches as we separately analyze each frequency component. We identify a significant bidirectional causal relationship between 2000 and 2003 as well as between 2007 and 2010 at 128-512 days time-scale. Additionally, we discover a high degree of interdependence during financial turmoil during 2002, 2007-2010 and 2012-2016. The observed phase pattern and thus the implicated non-stable lead-lag relationship favor solely for specific time intervals the Dornbusch model by Dornbusch and Fischer (1980) as well as the portfolio approach by Dominguez and Frankel (1993).

The above chapters look at the decompositions of economic time series to study temporal and frequency information simultaneously. The final chapter is devoted to the aggregation of information in unstructured data to explain the impact of surprising events on financial markets. To achieve this purpose, we model the informational content, produced by the S&P500 companies on Twitter, as a network. The extracted topic keywords stemming from the Twitter dataset are defined as nodes that form clusters which we define as events. The network projection allows us to observe the propagation of events and to quantify its' dynamics on the time axis. In this regard, we associate an unexpected event with a strongly expanding cluster in the network representation. Aggregating this process into a single indicator over the first and second moment allows quantifying the surprise level dynamics for each time interval. We discover that the proposed surprise indicator has an inverse effect on S&P500 returns and a positive impact on volatility. Additionally, for both the return and volatility models supporting a pre-report effect. Moreover, endogeneity cannot be completely ruled out in our study as we cannot answer the question of whether market volatility determines the level of Twitter activity or vice versa.

1.1 Publication details

Paper I (chapter 2):

A hybrid forecasting algorithm based on SVR and Wavelet Decomposition

Authors:

Timotheos Paraskevopoulos, Peter N. Posch

Abstract:

We present a forecasting algorithm based on support vector regression emphasizing the practical benefits of wavelets for financial time series forecasting. We utilize an effective de-noising algorithm based on wavelets feasible under the assumption that a systematic pattern plus random noise generate the data. The learning algorithm focuses solely on the decomposed time series components, leading to a more general approach. Our findings propose how machine learning can be used for data science applications in combination with signal processing methods. Applying the algorithm to real life financial data, we find wavelet decompositions to improve forecasting performance significantly.

Publication details:

Quantitative Finance and Economics, 2018, 2(3)

Paper II (chapter 3):

Time-frequency linkages and co-movements between the Euro and European stock market: A continuous wavelet analysis.

Authors:

Timotheos Paraskevopoulos, Peter N. Posch

Abstract:

We investigate the evolution of co-movement and lead-lag relationships between the nominal effective European exchange rate and the largest European stock markets in the time and frequency dimension. We decompose the financial return series into several time scales and apply the cross-wavelet coherence and phase difference. We observe patterns consistent with the notion of contagion, suggesting strong and sudden increases in the cross-market synchronization on particular frequency bands for the period between 2000 and 2016. Investigating the lead-lag relationships between both markets reveals that causality runs from one variable to the other and vice-versa depending on the frequency.

Publication details:

Working paper.

Paper III (chapter 4):

Event detection and the impact of surprise.

Authors:

Timotheos Paraskevopoulos, Peter N. Posch

Abstract:

This study integrates topic modeling into network theory to develop a metric of quantifying surprise. We propose to design an undirected network to model the flow of informational content published by the S&P 500 companies on Twitter. The network representation allows measuring the creation and propagation dynamics of real-world events from an agnostic point of view. To examine the overall effect of rare events on the S&P 500 we introduce a metric based on network indicators to quantify the daily level of surprise on the aggregated level. According to the results, we observe a positive impact of the aggregated surprise on the volatility and an inverse effect on the returns. Moreover, the results indicate for both the return and volatility models a pre-report effect. This study is the first attempt to quantify the level of surprise using topic modeling and network theory and to suggest how surprise can be quantified in user-generated content.

Publication details:

Working paper.

2 A hybrid forecasting algorithm based on SVR and Wavelet Decomposition

The following is based on Paraskevopoulos and Posch (2016).

2.1 Introduction

Forecasting the movements of stock market indices is challenging as the underlying process is characterized by significant fluctuations and exhibits nonlinear dynamics. There has been a tremendous amount of research on this topic over the past years, and a great deal of effort has been made to gain advantages in financial time series predictions. Atsalakis and Valavanis (2009), provide a cohesive presentation and classification of the research on stock market forecasting. The most widely used methods in forecasting such as autoregressive integrated moving average models, the exponential smoothing method, various models for seasonality, structural models, the Kalman filter and nonlinear models (including regime-switching models) could not reliably outperform the random walk (Atsalakis and Valavanis 2009). Machine learning provides a theory to construct algorithms that can learn, memorize and generalize from data. Generalization is essential to produce sufficiently accurate predictions from new and unseen patterns after having experienced learning on a related data set. However, before an algorithm can generalize it must be able to recognize the observed patterns which is achieved by the prior learning process. Therefore, the incremental informative content of the considered learning sample is crucial. Redundant, noisy or unreliable information impair the learning process and undermine the success of machine. Given the assumption that a systematic pattern plus random noise generate the considered data, a useful denoising algorithm provides a deeper understanding of the data generating process leading to a more precise forecast.

Wavelet transformation, as a signal processing technique, enables the decomposition

of a time series into the time and the frequency domain resulting to powerful feature extraction capabilities (Li et al. 2002; Kim and Han 2000; Dacorogna et al. 2001; Gonghui et al. 1999; Huang and Wu 2008; DeLurgio 1998; Seo et al. 2015). In this study, we introduce a rolling hybrid wavelet support vector regression with a precede recursive feature elimination (WL-RFE-SVR) algorithm, which is apart from traditional econometric and time series techniques. Our proposed algorithm consists of three stages which are conducted every walking forward step. In the first stage, we split the sample set into a training, validation and test set. Each sub-sample is subsequently decomposed into its low and high-frequency components using wavelet transform which enables us to extract additional information from the considered time series (Chang and Fan 2008; Li et al. 2002; Lu, Lee, and Chiu 2009; Ramsey 2002). In the second step, we identify the optimal feature set and support vector regression (SVR) hyperparameters using the recursive filter elimination (RFE) and hyperparameter optimization routine based on the training and validation set. In the third step, we apply the identified optimal feature set and SVR parameters on the test and obtain the out-of-sample (OOS) forecasts.
To best of our knowledge, the proposed WL-RFE-SVR methodology has not been presented in the literature; however, similar ideas have been discussed by Huang and Wu (2008). They apply a relevance vector machine on wavelet-based extracted features to forecast stock indices. In contrast to our work, the feature space dimension is not reduced by a feature selection algorithm. An algorithm based on wavelet decompositions and an MLP neural network with a financial application shows promising profitability when the data set is decomposed into low and high-frequency components (Zhang et al. 2001). Kao et al. (2013), provide a similar hybrid approach which, however, requires multiple optimization steps for feature selection and parametrization of the SVR. Also, their application is based on discrete wavelet transforms (DWT) which require the sample to be dyadic, i.e., observations of pairs each from disjunct sets. This restricts the application of the SVR as the size of the training set cannot be chosen arbitrarily.

We contribute to the literature by introducing a hybrid algorithm for processing data based on wavelet transform and SVR which integrates feature selection and parameter optimization in a single optimization step. Thereby we shed light on wavelet decompositions with the purpose to improve the forecast performance of an SVR by systematically comparing several wavelet filters and investigate the impact of short-term to long-term forecasting horizons on the performance of a trading application. The

remainder of the paper is as follows: We introduce and discuss the proposed method in the following section 2.2. In sections 2.2.1, 2.2.2 and 2.2.3 we give an introduction into wavelet transform, recursive feature selection and describe the concept of SVR. Section 2.3 focuses on the application of the method to real data, while the final section concludes.

2.2 Methodology

The proposed forecasting algorithm is a purely time-delayed model which captures the relationship between historical and subsequent index log returns. An effective adaptation to changing market conditions is achieved by optimization which follows the purpose of determining the most robust implementation of the forecasting algorithm within a specific time interval. A walk-forward routine is a well-suited systematic approach for this purpose as it allows to measure the robustness of the forecasting algorithm exclusively by out of sample forecasts. To effectively eliminate the likelihood that the out of sample results are just a matter of luck, the procedure is conducted over a time span of 4 years resulting in a series of 1200 successive out of sample tests. For each walking-forward step of 100 observations, we split the past 1100 observation into a training set (800 observations), a validation set (200 observations) and a test set (100 observations). Subsequently, we decompose each sub-sample separately into its low and high-frequency components. We use the training set to train our SVR with a specific feature set and hyperparameters while we asses its performance on the validation set. We pick the best performing SVR (based on the validation set), which is characterized by a specific feature set and its hyperparameters, and produce 100, depending on the forecast horizon, 1, 5, 10 and 20 days ahead, forecasts. In particular, the data preprocessing component follows the critical objective of denoising. We obtain the trend part by setting all high-frequency wavelet coefficients equal to zero and applying the inverse wavelet transform. To obtain the variation part, we set the low-frequency wavelet coefficients equal to zero and perform the same procedure. To avoid a looking forward bias, we decompose each sub-set using a rolling window approach with a window length of 200 observations and a step size of one observation. Subsequently, the algorithm decides based on an RFE algorithm the amount of trend and variation components that will be utilized to perform the forecasts. For each feature subset,

we perform a grid search optimization to identify the optimal SVR hyperparameters. The flowchart in Figure A.1 depicts the proposed methodology in detail. Instead of restricting ourselves to a single filter, we provide a comparative analysis of a wide range of wavelet transform filters which are adequate for financial time series analysis (Dacorogna et al. 2001). To ensure better comparability, we extended the study by a naive strategy. For this strategy, we omit the data preprocessing based on a wavelet transform. An important property of this algorithm is the scalability which allows additionally to point and classification tasks, the incorporation of trading strategies into the model.

2.2.1 Wavelet Transforms

Wavelet transforms are based on Fourier analysis, which represents any function as the sum of the sine and cosine functions. A Fourier transform is not applicable to non-stationary signals as it assumes the signal to be periodic. In contrast, a wavelet transform is a mathematical tool that can be applied to numerous applications, including non-stationary time series. The wavelet transform defines a finite domain which makes it well localized concerning both time and frequency. Our approach is based on the maximal overlap discrete wavelet transform (MODWT) as we can apply it to an arbitrarily large sample whereas a discrete wavelet transform restricts the sample size to be dyadic. Additionally, the trend and variation coefficients of a MODWT can be perfectly aligned with the original time series as they are associated with zero-phase filters (Chernick 2001). Depending on the normalization rules, there are two types of wavelets within a given function/family. The father wavelets which describe the low-frequency components of a signal and the mother wavelets which describe its' high-frequency components:

$$\Phi_{j,k} = 2^{-j/2}\Phi\left(\frac{t-2^j k}{2^j}\right) \tag{2.1}$$

$$\Psi_{j,k} = 2^{-j/2}\Psi\left(\frac{t-2^j k}{2^j}\right) \tag{2.2}$$

where $\Phi(t)$ represents the father wavelet, $\Psi(t)$ the mother wavelet and $j = 1, ..., J$ is the J-level wavelet decomposition (Ramsey and Lampart 1997). The two types of wavelets

stated above, satisfy

$$\int \Phi(t)dt = 1 \text{ and } \int \Psi(t)dt = 0. \tag{2.3}$$

The low-frequency parts can be understood as the smooth baseline trend from a signal whereas the high-frequency parts are any deviations from that. Given a signal $f(t)$, its wavelet transform is formed by (Chernick 2001):

$$f(t) \approx \sum_k s_{J,k}\Phi_{J,k}(t) + \sum_k d_{J,k}\Psi_{J,k}(t) + \sum_k d_{J-1}\Psi_{J-1,k}(t) \tag{2.4}$$

$$+ ... + \sum_k d_{1,k}\Psi_{1,k}(t) \tag{2.5}$$

The wavelet transform coefficients are given by $s_{J,k}$ and $d_{J,k}, ..., d_{1,k}$ where J denotes the maximum decomposition level sustainable by the number of data points. More precisely, 2^j is less than the total number of data points. The parameter k is defined as the dilation parameter ranging from one to the number of wavelet coefficients. The high and low frequency component coefficients $s_{j,k}$ and $d_{j,k}$ can be approximated by the following integrals:

$$s_{J,k} = \int f(t)\Phi_{J,k}(t)dt \tag{2.6}$$

$$d_{j,k} = \int f(t)\Psi_{j,k}(t)dt. \tag{2.7}$$

Given the coefficients, we obtain the low-frequency component of a signal by setting all high-frequency coefficients $d_{j,k}$ for $j = 1, ..., J$ to zero (Chang and Yadama 2010). Accordingly, we obtain the high-frequency components by setting the low-frequency coefficients $s_{J,k}$ to zero. The corresponding wavelet coefficients are approximated by the specific basis functions $\Phi_{J,k}(t)$ and $\Psi_{J,k}(t)$. Each basis function leads to a different transform as a given signal is weighted over several scales and center time differently. The issue of which basis function generates the best resolution for a specific forecasting application is considered as the most indispensable challenge in wavelet analysis and needs to be addressed by trial and error (Chernick 2001). On this account, we provide performance comparison for basis functions appropriate for time series analysis. In general, low order wavelets have good time resolution, but poor frequency resolution whereas high order wavelets has good frequency resolution but poor time resolution. As it is yet unknown how this, in particular, affects the performance of our

forecasting algorithm during different market regimes we analyze several filter lengths and basis functions: Daubechies (D) (4, 6, 8, 12), Least Asymmetric (LA) (8, 10, 12, 14), Best Localized (BL) (14, 18, 20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. Daubechies belong to the most common applied basis functions for time series denoising. CF filters allow specifying vanishing conditions on the associated scaling function leading to remarkably good phase properties (Daubechies 1992). LA filters have approximately linear phase while minimizing the maximum deviation of the phase function from the linear phase over the low and high frequencies (Oppenheim 1999). Based on the idea that deviations from the high-frequency phase are not as important as ones from low frequencies, Doroslovacki (1998), introduced the BL filter which penalizes deviations at low frequencies more heavily than those at high frequencies.

2.2.2 Feature Selection

To achieve the best possible performance with a particular learning algorithm, it is necessary for the learning algorithm to focus on the most relevant information in a potentially overwhelming quantity of input variables. Proper feature selection algorithms provide a decisive countermeasure to reduce the dimensionality of the input variables. Feature selection approaches are divided into two categories: filter and wrapper methods (John, Kohavi, and Pfleger 1994). Filter based methods use a mathematical evaluation function to justify a plausible relationship between the input and output variables. The interested reader is referred to Saeys, Inza, and Larrañaga (2007), for a survey regarding filter methods. In contrast to filter methods, wrapper methods are brute-force search algorithms which maximize model's performance by adding or removing features while evaluating each feature subset in the context of the learning algorithm. The subset selection takes place based on the learning algorithm used to train the model itself and allows to consider how the algorithm and the training set interact. For feature selection, we choose an RFE algorithm which repeatedly constructs our proposed model and ranks features sets according to the achieved root mean square error (RMSE) for the point forecast task and according to accuracy for the classification task. However, an adverse effect on the generalization capability of our forecasting algorithm is the potential issue of overfitting. This occurs when the learning algorithm focuses on nuances of the

training data which are not found in future samples (i.e., overfitting to predictors and samples). For example, suppose a substantial number of uninformative features were collected, and one such feature set randomly correlates with the outcome. The RFE algorithm would give a good rank to this feature set, and the prediction error (on the same data set) would be lowered. It would take a different test/validation sample set to find out that this predictor was uninformative. This is referred to as selection bias by Ambroise and McLachlan (2002). To avoid the potential issue of overfitting and selection bias, we resampled each training set using a 10-fold cross-validation (Svetnik et al. 2004; Ambroise and McLachlan 2002).

2.2.3 Support Vector Regression

The supervised learning technique SVR is an application form of Support Vector Machines which deals with estimating real-valued functions by allowing the output to take any real number (Vapnik, Golowich, and Smola 1997). The key advantage of the SVR is that it allows for noise in the data by addressing the loss of accuracy from large residuals caused by outliers using structural risk minimization.

Structural Risk Minimization

Given a training set of labeled data drawn from an unknown distribution $P(x, y)$

$$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), ..., (x_l, y_l)\}, x \in \mathbb{R}^n \tag{2.8}$$

where the input variable $x_i \in \mathbb{R}^l$ is a n-dimensional vector and the output variable $y_i \in \mathbb{R}$ is a continuous value. The problem of learning is to determine a function $f(x, \alpha)$ which approximates the relationship between x_i and y_i with a parameter vector α as accurate as possible. The best approximation of y_i minimizes the following expected risk $R(\alpha)$:

$$R(\alpha) = \int L(y, f(x, \alpha)) dP(x, y), \tag{2.9}$$

where $L(y, f(x, \alpha))$ is a loss function describing the difference between the true y_i and its approximation:

$$L(y, f(x, \alpha)) = (y - f(x, \alpha))^2 \tag{2.10}$$

consequently Equation (2.9) can be written as

$$R(\alpha) = \int (y - f(x, \alpha))^2 dP(x, y) \tag{2.11}$$

Since x_i and y_i follow an unknown distribution, Equation (2.9) can solely be minimized by an inductive step. Hence, we replace the expected risk $R(\alpha)$ by the empirical risk

$$R_{emp}(\alpha) = \frac{1}{N} \sum_{i=1}^{l} (y - f(x, \alpha))^2 \tag{2.12}$$

However, minimizing the empirical risk does not necessarily imply a small expected risk in small samples which hampers generalizing from unseen data. A solution to this issue is provided by structural risk minimization principle which simultaneously minimizes the empirical risk and a confidence interval Ω (Vapnik 1999). Given the VC-dimension, defined as the parameter h, the bound on the expected return can be written as:

$$R(\alpha) \leq R_{emp}(\alpha) + \Omega \left(\frac{l}{h} \right), \tag{2.13}$$

whereby the VC-dimension can be interpreted as the model capacity or complexity of the learning machine. For a given amount l of data and predetermined VC-dimension h^*, the generalization ability of the machine depends on the confidence interval $\Omega(l/h^*)$. In case that we construct the learning machine too complex, we may be able to minimize the empirical risk $R_{emp}(\alpha)$ down to zero. However, the confidence interval Ω would be in this case large implying a large expected risk. This case is called overfitting and can be avoided by choosing the VC-dimension that minimizes the expected risk while controlling both terms in Equation (2.13). In support vector methods, the tradeoff between the quality of the approximation and the complexity of the approximating function is determined by the parameter C. This parameter can be interpreted as the degree of penalized loss when a training error occurs and is described in more detail in the following section.

Linear Support Vector Regression

To approximate the output vector y, the SVR model performs a linear regression with an error tolerance ϵ. The regression function has the following form:

$$\hat{y}_i = f(x_i) = w_0 + w \cdot x, \tag{2.14}$$

where w_0 and w^T are the parameter vectors of the function. The tolerated losses within extent of the ϵ-tube as well the penalized losses L_ϵ are defined by the ϵ-insensitive loss function,

$$L_\epsilon(y, \hat{y}) = \begin{cases} 0 & \text{if} \quad |y - \hat{y}| \leq \epsilon \\ |y - \hat{y}| - \epsilon & \text{if} \quad |y - \hat{y}| > \epsilon \end{cases} \tag{2.15}$$

The linear regression $f(x_i)$ is estimated by simultaneously minimizing $||w||^2$ and the sum of the linear ϵ-insensitive loss function that is given by Equation (2.15). Hence the objective function can be written in the following, not differentiable, convex and unconstrained form:

$$L_\epsilon = C \sum_{i=1}^{N} L_\epsilon(y_i, \hat{y}_i) + \frac{1}{2}||w||^2 \tag{2.16}$$

where $C = 1/\lambda$ is a regulation constant which controls the trade-off between the complexity and the approximation accuracy of the regression model. By introducing two slack variables ξ_i^+ and ξ_i^-, $i = 1, 2, ..., n$, which represent the upper and lower training errors subject to the ϵ-tube we are able to formulate the problem as a constrained optimization problem:

$$y_i \leq f(x_i) + \epsilon + \xi_i^+ \tag{2.17}$$

$$y_i \geq f(x_i) - \epsilon - \xi_i^-. \tag{2.18}$$

Consequently, Equation (2.16) can be transformed into the following constrained form:

$$\min_{w, w_0, \xi_i^+, \xi_i^-} \quad R_{reg}(f) = \frac{1}{2}||w||^2 + C \sum_{i=1}^{l} (\xi_i^+ + \xi_i^-) \tag{2.19}$$

with subject to:

$$y_i - (w \cdot x_i) - w_0 \leq \epsilon + \xi_i^+ \tag{2.20}$$

$$-y_i + (w \cdot x_i) + w_0 \leq \epsilon + \xi_i^- \tag{2.21}$$

$$\xi_i^+, \xi_i^- \geq 0, \quad for \quad i = 1, ..., l \tag{2.22}$$

Minimizing the first term is equivalent to minimizing the confidence interval Ω since we minimize the complexity, whereas minimizing the second term is equivalent to minimizing the empirical risk. The constrained optimization problem can be solved by applying the Lagrangian theory and the Karush Kuhn-Tucker condition as has been done for SVM classifiers (Chang and Fan 2008; Schölkopf and Smola 2002). Introducing a dual set of Langrange multipliers $\underline{a}_i$ and $\bar{\alpha}_i$ enables applying the standard quadratic programming algorithm and thus to solve the optimization problem in the dual form:

$$\min_{\underline{\alpha}_i, \bar{\alpha}_i} \quad \frac{1}{2} \sum_{i,l=1}^{l} (\underline{\alpha}_i - \bar{\alpha}_i)(\underline{\alpha}_i - \bar{\alpha}_i)\langle x_i \cdot x_j \rangle \tag{2.23}$$

$$+ \epsilon \sum_{i=1}^{l} (\underline{\alpha}_i + \bar{\alpha}_i) - \sum_{i=1}^{l} y_i (\underline{\alpha}_i - \bar{\alpha}_i) \tag{2.24}$$

with subject to:

$$\sum_{i=1}^{l} (\underline{\alpha}_i - \bar{\alpha}_i) = 0 \tag{2.25}$$

$$0 \leq \underline{\alpha}_i \leq C, \quad i = 1, 2, ..., l \tag{2.26}$$

$$0 \leq \bar{\alpha}_i \leq C, \quad i = 1, 2, ..., l \tag{2.27}$$

Once the model is trained, by calculating the Lagrangian multipliers $(\underline{\alpha}_i, \bar{\alpha}_i)$, we are able to make predictions using

$$\hat{y}(x) = \hat{w_0} + \sum_{j,i=1}^{l} (-\underline{\alpha}_i + \bar{\alpha}_i)\langle x_i \cdot x_j \rangle \tag{2.28}$$

where $\langle x_i \cdot x_j \rangle$ is the inner product of x_i and x_j. The support vectors are given by x_i for which the corresponding coefficients $(-\underline{\alpha}_i + \bar{\alpha}_i)$ are nonzero. These are the data points

for which the errors lie outside the ϵ-insensitive tube and thus, support the definition of the approximate function.

Nonlinear Support Vector Regression

The practical use of a linear approximation is very limited when the data is nonlinear. However, the input data x_i can be mapped in a higher dimensional feature space by a nonlinear function $\Phi(x)$ in which it exhibits linearity. Consequently, the higher dimensional feature space can be lineary approximated. The decision function is now given by

$$\hat{y}_i = f(x_i) = w_0 + \langle w \cdot \Phi(x) \rangle, \tag{2.29}$$

The optimization problem as stated in Equation (2.19) can be expressed as:

$$\min_{w,w_0,\xi_i^+,\xi_i^-} \quad R_{reg}(f) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{l}(\xi_i^+ + \xi_i^-) \tag{2.30}$$

with subject to:

$$y_i - (w \cdot \Phi(x_i)) - w_0 \leq \epsilon + \xi_i^+ \tag{2.31}$$

$$-y_i + (w \cdot \Phi(x_i)) + w_0 \leq \epsilon + \xi_i^- \tag{2.32}$$

$$\xi_i^+, \xi_i^- \geq 0, \quad for \quad i = 1, ..., l \tag{2.33}$$

The dual form of the optimization problem for the nonlinear SVR is given by:

$$\min_{\underline{\alpha}_i,\bar{\alpha}_i} \quad \frac{1}{2}\sum_{i,l=1}^{l}(\underline{a}_i - \bar{\alpha}_i)(\underline{\alpha}_i - \bar{\alpha}_i)\langle\Phi(x_i) \cdot \Phi(x_j)\rangle \tag{2.34}$$

$$+ \epsilon\sum_{i=1}^{l}(\underline{\alpha}_i + \bar{\alpha}_i) - \sum_{i=1}^{l} y_i(\underline{\alpha}_i - \bar{\alpha}_i) \tag{2.35}$$

with subject to:

$$\sum_{i=1}^{l}(\underline{\alpha}_i - \bar{\alpha}_i) = 0 \tag{2.36}$$

$$0 \leq \underline{\alpha}_i \leq C, \quad i = 1, 2, ..., l \tag{2.37}$$

$$0 \leq \bar{\alpha}_i \leq C, \quad i = 1, 2, ..., l \tag{2.38}$$

In the final form of the prediction function the scalar products $\langle x_i \cdot x_j \rangle$ are replaced by $\langle \Phi(x_i) \cdot \Phi(x_j) \rangle$ while the scalar product is computed by a kernel function $K(x_i, x_j)$ which enables to avoid a mapping $\Phi(x)$. The key advantage here is that one does not necessarily need to specify the $\Phi(x)$ function in order to yield the inner products $\langle \Phi(x_i) \cdot \Phi(x_j) \rangle$. Once the Lagrangian multipliers are calculated, the kernelized SVR regression function can be written as

$$\hat{y}(x) = \hat{w_0} + \sum_{j,i=1}^{l} (-\underline{\alpha}_i + \bar{\alpha}_i) K(x_i, x_j) \tag{2.39}$$

Practically, the choice of the kernel function is restricted to those which satisfy Mercer's theorem (Vapnik 1999; Vapnik, Golowich, and Smola 1997). A popular kernel function which we also apply in our study is the Radial Basis Function (RBF):

$$K(x_i, x_j) = exp(-\gamma ||x_i - x_j||^2). \tag{2.40}$$

The choice of the RBF kernel relies on the results of a trial-error-process in which we identified the RBF kernel as the most suitable kernel among the linear, polynomial and sigmoid kernel. This finding is confirmed by further applications on financial time series in the literature (Ince and Trafalis 2008; Lu, Lee, and Chiu 2009).

Parametrization of the nonlinear SVR

The generalization performance of the SVR machine strongly depends on the parameter C, the insensitive tube ϵ and the kernel parameter γ (Cherkassky and Ma 2004). Additionally, the parameters are mutually dependent which hampers the optimization process. The greater the parameter C, the greater the penalty for errors which forces the machine to become more complex to achieve a low empirical risk. A smaller C

leads to an increased error tolerance which in return leads to a poor approximation. However, such behavior may be preferred in case of noisy data. The ϵ-insensitive tube affects the learning machine complexity as it determines the number of support vectors. A greater ϵ leads to a greater ϵ-insensitive tube. As a result, a larger amount of data points which are within the ϵ-insensitive are ignored by the learning machine resulting in flattening the regression function. We identify adequate hyperparameters for our SVR based on a grid search within a cross-validation routine which we conduct at each walking-forward step on historical data. By doing this, we use estimated parameters from the literature to restrict the scope of our grid search (Cherkassky and Ma 2004; Mattera and Haykin 1999). As the scope of this study does not rely on the optimal parametrization of an SVR, the grid search results will not be further discussed.

2.3 Empirical Results

Our data set consists of daily bid and ask closing prices from May, 18th 2007 to Dec, 30th 2015 from nine major stock indices: US (S&P 500), United Kingdom (FTSE100), Japan (Nikkei225), Germany (DAX30), France (CAC40), Switzerland(SMI20), Australia (ASX200) and Hong Kong (HS33). The selection was made with the premise of providing comprehensive analysis on the most traded indices based on 2200 observations per each index. From the daily closing price P_t the calculate the log returns for a period of l days at time τ as $r_\tau = ln\left(P_\tau / P_{\tau-l}\right)$

We develop the WL-RFE-SVR algorithm based on the median price while conducting the trading application based on the bid and ask price to account for transaction costs. As we aim at predicting the logarithmic return 1, 5, 10, 15 and 20 trading days for each index, the parameter l varies respectively from 1 to 20. The first 1000 observations are used to initialize the model. We re-train the model every 100 observations thereby producing 1200 out-of-sample forecast for every index. The descriptive statistics for each index and forecasting horizon are given in Table 2.1. We observe, except for DAX, a negative skewness for every time series. Additionally, every time series prevails a kurtosis greater than 2.4 and is non-normal as confirmed by the Jarque-Bera statistics at the 99 percent confidence interval.

The results are solely based on out-of-sample (OOS) forecasts to avoid any forward-looking bias.

Symbol	DAX	FTSE100	SP500	ASX200	SMI20	CAC40	HangSeng33	Nikkei225
1 Trading Day								
Mean	0.00033	0.00012	0.00017	0.00012	0.00016	0.00011	0.00021	0.00016
Median	0.00096	0.00052	0.00037	0.00057	0.00065	0.00057	0.0005	0.00045
Std Dev	0.01452	0.01294	0.01159	0.01102	0.01146	0.0148	0.01591	0.01511
Skewness	0.02288	-0.14179	-0.17243	-0.4901	-0.52552	-0.22491	-0.0212	-0.16843
Kurtosis	11.11564	12.82436	13.74693	7.29642	11.57568	7.50848	6.22434	8.7834
Max	0.15788	0.12501	0.11211	0.07747	0.10545	0.11652	0.11143	0.1309
Min	-0.1041	-0.11063	-0.10506	-0.09212	-0.12067	-0.1065	-0.11613	-0.11055
5 Trading Days								
Mean	0.00156	0.00059	0.00082	0.0006	0.00076	0.00048	0.00098	0.00077
Median	0.00403	0.00219	0.00282	0.00258	0.00289	0.0029	0.00311	0.00296
Std Dev	0.0303	0.02501	0.0233	0.0226	0.02482	0.02956	0.03382	0.03138
Skewness	-0.81624	-0.75415	-0.94196	-0.76561	-0.99497	-0.69869	-0.50057	-1.00642
Kurtosis	4.90999	6.7913	8.65197	4.45807	8.16449	4.08987	6.1168	10.46062
Max	0.1723	0.1726	0.15449	0.1096	0.15036	0.15219	0.21578	0.24451
Min	-0.20055	-0.17495	-0.2041	-0.16135	-0.23922	-0.1904	-0.31876	-0.31106
10 Trading Days								
Mean	0.00307	0.00118	0.0016	0.00119	0.0015	0.00093	0.00188	0.00158
Median	0.00851	0.0044	0.00526	0.00479	0.0049	0.00668	0.00454	0.00446
Std Dev	0.04113	0.03353	0.03134	0.03094	0.03291	0.03976	0.04703	0.04253
Skewness	-1.24353	-1.13694	-1.52812	-1.05393	-1.08994	-1.0033	-0.63602	-1.12514
Kurtosis	4.42924	5.33443	9.42939	3.64628	2.89532	2.74262	4.47447	6.45829
Max	0.15569	0.15157	0.13999	0.11462	0.13956	0.16402	0.26098	0.23354
Min	-0.27998	-0.25969	-0.28763	-0.23946	-0.22718	-0.24877	-0.38096	-0.41482
15 Trading Days								
Mean	0.00453	0.00177	0.00237	0.0018	0.00223	0.00136	0.00272	0.00241
Median	0.01208	0.0066	0.00723	0.00716	0.00711	0.00938	0.00556	0.008
Std Dev	0.05	0.04005	0.03779	0.03721	0.03979	0.04739	0.0567	0.05243
Skewness	-1.27906	-1.27929	-1.6261	-0.98944	-0.97344	-1.05765	-0.67646	-1.17908
Kurtosis	4.42924	5.33443	9.42939	3.64628	2.89532	2.74262	4.47447	6.45829
Max	0.15283	0.15387	0.20458	0.15854	0.14684	0.14116	0.26013	0.2174
Min	-0.29832	-0.28481	-0.30758	-0.24146	-0.25497	-0.28256	-0.47736	-0.44401
20 Trading Days								
Mean	0.00592	0.00229	0.00312	0.00241	0.00288	0.00168	0.00351	0.00326
Median	0.01329	0.00769	0.01028	0.00933	0.00861	0.01041	0.00772	0.00815
Std Dev	0.05691	0.045	0.04332	0.04243	0.04515	0.05387	0.06571	0.06048
Skewness	-1.21703	-1.24122	-1.6735	-0.88753	-1.09631	-1.02127	-0.70065	-1.12914
Kurtosis	4.07509	4.77541	8.23162	2.5869	3.4117	2.42166	4.52242	5.9459
Max	0.18674	0.16214	0.18365	0.15382	0.17002	0.1682	0.26912	0.26389
Min	-0.35076	-0.29631	-0.30981	-0.24845	-0.33676	-0.28736	-0.5371	-0.46988

Table 2.1: Descriptive statistics for each symbol and granularity. The granularity varies between +1, +5, +10, +15, +20 days.

2.3.1 Point Forecast Results

We measure the point forecast performance based on the mean absolute percentage error (MAPE) and the RMSE which are given by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=0}^{N}(y_t - \hat{y}_t)^2}, \tag{2.41}$$

$$MAPE = \frac{1}{N}\sum_{t=0}^{N}|y_t - \hat{y}_t| \tag{2.42}$$

where y_t is the observed value at time t, $\hat{y}_t$ the modelled value at time t and N the sample size. The advantage of MAPE against RMSE is that it allows comparing the point forecast accuracy across different time series as it is scale-independent. Nevertheless, reporting RMSE appears necessary to enhance the comparability with

other contributions on the topic. The point forecast results are illustrated in Figure A.2 to A.5. First, it becomes evident that the naive strategy is equal or better than any wavelet filter regarding MAPE and RMSE. This extends across every forecasting horizon and index. Notwithstanding the naive strategy, the wavelet filters $CF6$, $D4$ and $D6$ outperform every other filter, again, regarding MAPE and RMSE. We confirm this observation by every forecasting horizon and index. It also becomes evident that the RFE algorithm has a positive effect on the MAPE and the RMSE. This holds for every forecasting horizon and index, which strongly emphasizes the need for a feature selection algorithm. As to the optimal forecasting horizon, the Figures A.4 and A.5 show that a forecasting horizon of five days delivers the best point forecasting results regarding MAPE and RMSE. Nevertheless, wavelet transforms appear to be inappropriate for point forecasting as the naive approach outperforms each wavelet filter regarding MAPE and RMSE.

2.3.2 Clasification Results

The forecasted sign of an asset return plays for several financial applications a more important role than its point forecast. For this reason, we additionally asses the performance of our model for a binary classification task based on the confusion matrix. Consequently, returns greater than zero belong to class A and returns smaller than zero to class B. Subsequently, we evaluate the classification performance based on the accuracy and the Cohen's Kappa which are given by:

$$Accuracy = \frac{\sum \text{True Positive} + \sum \text{True Negative}}{\text{Total Population}} \tag{2.43}$$

$$Kappa = \frac{\text{Accuracy} - \text{Random Accuracy}}{1 - \text{Random Accuracy}}, \tag{2.44}$$

where $\text{RandomAccuracy} = N^{-1}(\text{Act.False} \times \text{Pred.False} + \text{Act.True} \times \text{Pred.True})$. In contrast to Accuracy, Cohen's Kappa provides a more reliable measure of agreement as it takes the amount of agreement that could be expected due to chance alone into account. The higher the Cohen's Kappa is, the better is our classifier compared to a random chance classifier. As we can see in Figure A.6 and A.7, the wavelet filters $CF6$, $D4$ and $D6$ slightly outperform the naive strategy regarding Accuracy for most of the indices and forecasting horizons. Additionally, we observe that the RFE algorithm

has a positive impact on the Accuracy which is in accordance with the point forecast results. The Cohen's Kappa results (Figure A.8 and A.9) align with the Accuracy results. There is no clear-cut differentiation between the dominating wavelet filters and the naive strategy.

2.3.3 Trading Results

In practice, an ultimate goal of a forecasting algorithm in finance is to confirm its OOS profitability as statistical accuracy is not always synonymous with the financial profitability of the deriving forecasts. Therefore, based on the above-defined classification rule, we implement a most basic trading simulation to extend the results by a financial application and improve the comparability of the applied wavelet filters. We measure the profitability of our trading algorithm based on the following metrics:

Total return

$$tr = \sum_{t=0}^{T} r_t. \tag{2.45}$$

Annualized return

$$ar = 252 \frac{1}{T} \sum_{t=1}^{T} r_t. \tag{2.46}$$

Sharpe Ratio

$$sr = \frac{\frac{\sum_{t=1}^{T} r_t}{T} - \frac{r_{fix}}{252}}{\sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (r_t - \bar{r})^2}} \tag{2.47}$$

Standard deviation

$$sd = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (r_t - \bar{r})^2} \tag{2.48}$$

Maximum Drawdown

$$mdd = max_{\tau \in (0,T)} \left[max_{\tau \in (0,T)} X(t) - X(\tau) \right] \tag{2.49}$$

where r_t denotes the realized return and $X(t)$ the realized equity at time t. The trading strategy is identical for all models and is defined as follows: enter a long position when the return forecast is greater zero and enter a short position when the return forecast

is below zero. Trading costs have been considered to the extent that long positions are executed based on the ask quote and short positions based on the bid quote of the respective instrument. As shown in Figure A.10 and A.11, the wavelet filters $D4$, $D6$ and $CF6$ outperform every other considered wavelet filter including the naive strategy. In contrast to the point forecast and classification results we are able to make a clear distinction in the performance between the $D4$, $D6$ and $CF6$ filters and the remainder. The highest total return and annualized return is achieved for the forecasting horizon of 20 days. On the other hand, the increasing drawdown in forecasting horizon illustrates the increasing downside risk which comes along. This trade-off between return on investment and downside risk is present in all forecasting horizons. Again, the RFE algorithm has a positive effect on the overall return and maximum drawdown for the dominant wavelet filters.

2.4 Conclusion

In this study, we introduce a WL-RFE-SVR algorithm with an optimal parameter and feature selection which is applied to the task of forecasting and trading multiple steps ahead of the worlds major stock market indices. The proposed model is trained during the period 16/01/2007-22/11/2010 while its out-of-sample performance is validated over the period from 23/11/2010 through 31/12/2015 based on a walking forward analysis. The study shows that despite the adverse statistical properties of daily data, proper transformation and pre-processing can reveal reoccurring patterns that can be a source for a trading algorithm. Even though we are not able to obtain an advantage over the naive strategy in the task of point forecasting, the classification and especially the trading results reveal a different picture. The positive results for three wavelet filter out of fifteen, namely $D4$, $D6$ or a $CF6$, show that noise does not completely govern the process of price formation in major stock indices. Furthermore, results provide evidence to support the hypothesis of a serial dependence at least when the data is preprocessed by a $D4$, $D6$ or a $CF6$ wavelet filter. In the light of limited computational power, we are not able to exploit the full potential of the proposed hybrid method. Hence, the results can be considered as a lower bound on profits that can be achieved with a WL-RFE-SVR trading algorithm. A more sophisticated implementation of entry and exit rules and the testing of a wider/smaller range of inputs may increase forecasting

and trading performance. The same applies to the utilized money management. An extension would include an adjustment of the size of the position depending on the strength of the trading signal or volatility. Concerning some future directions of research, the proposed hybrid method could be applied to other forecasting problems in the domain of finance to prove its effectiveness. Additional research may be conducted towards the implementation of more refined feature selection algorithms and a deeper investigation of how wavelet filter influence the performance of the algorithm.

3 Time-frequency linkages and co-movements between the Euro and European stock market: A continuous wavelet analysis.

The following is based on Paraskevopoulos and Posch (2017).

3.1 Introduction

In times of complex interconnectedness between financial institutions, we observe the phenomenon of propagation of financial market fluctuations from one market to another due to financial market linkages. As a consequence, a shock in one financial market can spread to multiple markets, leading to higher economic damage than originally expected. As each market participant acts based on an individual investment horizon, its impact on the market should be visible at specific frequencies. This favors the application of a time-frequency decomposition with the aim to identify frequency specific co-movement patterns. In this paper we aim at adopting a long-and short-term perspective on the impact of crises on market linkages between the nominal effective European exchange rate and European stock markets[1]. Our data set includes the dot-com bubble in September 2001, the financial crisis in 2007 and the European sovereign debt crisis since 2009.

An investigation into transmission processes between international financial markets requires a short introduction into the terminology as the literature specifically relevant in our study is not used coherently. The term financial market "linkages", "relations" and "interdependence" are subdivided by a recent branch of literature into "interdependence"

[1]The nominal effective exchange rate of the Euro is a summary measure of the Euro currency value vis-a-vis the currencies of the 19 most trading European partners (Austria, Canada, Denmark, Hong-Kong, Japan, Norway, Singapore, South Korea, Sweden, Hungary, Great Britain, USA, Bulgaria, Czech Republic, Poland, Romania, Croatia and China).

and "contagion" (Forbes and Rigobon 2002; Corsetti, Pericoli, and Sbracia 2005; Gebka and Karoglou 2013). Proceeding in this way, Forbes and Rigobon (2002) define contagion as a sudden increase in measured cross-market linkages among assets or asset classes due to a shock which disappears within a short period. Further studies such as Caporale, Cipollini, and Spagnolo (2005), Mcclelland and Kokoski (1994), and Billio and Caporin (2006) rely on this definition to describe contagion in the context of cross-market co-movement in returns. In contrast, interdependence is characterized by a fundamental-driven state of stable market linkages which are subdivided into a continuous, normal and tranquil-period relations between the markets (Baele and Inghelbrecht 2010; Kallberg and Pasquariello 2008). In this state, measured stock market linkages can then be entirely explained by common observed factors due to real or financial linkages. Phenomena such as sudden expectation shifts or herding are excluded. However, high levels of co-movement and some limited time-variation in measured linkages are well in line with the notion of interdependence. This is acknowledged, for example, by Forbes and Rigobon (2002), Billio and Caporin (2006), and Baele and Inghelbrecht (2010), who state that fundamentals vary over time too.
Our study aims at answering three questions. Firstly, are the dynamics of co-movement between the nominal effective European exchange rate and the European stock markets stable? Secondly, can we find evidence for contagion during our sample period and in particular during the financial crisis of 2007 for specific frequencies? Thirdly, does measured co-movement reveals any lead-lag dynamics at specific frequencies, such as those in business cycles?
To address these questions empirically, we estimate the spectral characteristics of the time series as a function of time and frequency using a continuous wavelet transform. In a second step, we compute the continuous wavelet power spectrum (WPS) and three cross-wavelet tools: the cross-wavelet power spectrum (CWP), the cross-wavelet coherence (CWC) and the phase difference (PD) of the complex wavelet coherence. This framework allows us straightforwardly to reveal linkages and co-movement patterns between periodic components. Moreover, we can include measures which quantify the level of co-movement and direction of lead-lag relationships between the foreign exchange market and the European stock market.
Our work is mainly encouraged by Aguiar-Conraria and Soares (2014) as they utilized the phase-difference to identify lead-lag dynamics across frequencies. These tools help

to unearth economic time-frequency relations that have not been captured thus far by classical correlation or covariance-based approaches. The dynamic interdependence patterns over different time-scales and frequencies of stock markets cannot be identified if the information is averaged among frequencies. We find causal and reverse causal relation between share prices and the nominal effective European exchange rate which varies across scale and time in Europe. The evidence supports both cyclical and anti-cyclical relationships between the series. Share prices lag and receive cyclical effects from exchange rate shocks at higher time scales over the study period. The findings offer fresh insights into the causal link between the Euro and stock returns in Europe. The methodological approach we apply succeeds in capturing the simultaneous influence of the two series and thus, helped to detect bidirectional causality at different time scales. This paper contributes to the literature as the first study applying the methodology of cross-wavelet coherence and phase-difference to the given economic time series.
The rest of the paper is organized as follows. Section 3.2 reviews the relevant literature and defines important terminology. Section 3.3 describes the applied methodology. In section 3.4, the empirical results are reported. Section 3.5 draws conclusions based on the findings.

3.2 Related Literature

Macroeconomic theory offers two perspectives on causal relationships from exchange rates to stock markets and vice versa, which extend to the phenomenon of contagion and interdependence. According to Dornbusch and Fischer (1980), the exchange rate movements affect a firm's international competitiveness and real income which in turn affects firm's future cash flow. Since stock prices are interpreted as the present value of the firm's future cash flows, it is safe to assume the stock prices are affected by exchange rates. This implies a negative correlation between both instruments in which stock prices act as a leading driver within a lead-lag relationship. On the other hand, the portfolio approach by Dominguez and Frankel (1993) assumes that stock markets affect the behavior of exchange rates through the demand for money. According to the portfolio-balance model, a stock price depreciation leads to a reduction in wealth of domestic investors, which leads to lower demand for money, and thus lower interest rates. C. p., this must be followed by capital outflows leading to currency depreciation.

Thus, the correlation between exchange rates and stock markets will be positive. This is line with Karoui (2006), as stocks become due to the exchange rate depreciation more attractive to foreign investors which again leads to a positive correlation between exchange rates and stock markets in which exchange rates act as a driver of a lead-lag relationship. Despite the sign of the effect, both models provide an economic foundation in which investor's behavior drives the correlation between exchanges rates and stock markets. However, it is unclear which of both effects dominates in a lead-lag relationship. The first branch of literature, consisting of Franck and Young (1972), Solnik (1987), Smith (1992), and Phylaktis and Ravazzolo (2005), provides evidence for significant and positive causal relationship between exchange rates and stock indices. The second branch of literature provides evidence for a contradicting, although significant, negative causal relationship (Soenen and Hennigar 1988; Ajayi and Mougouė 1996; Ma and Kao 1990). Concerning lead-lag relationships between both time-series, Granger, Huangb, and Yang (2000) provide empirical evidence supporting the hypothesis that stock prices lead exchange rates in some countries while the opposite is true for South Korea. Abdalla and Murinde (1997), produced similar results showing unidirectional causality from exchange rates to stock prices in developing countries, while Ajayi, Friedman, and Mehdian (1998), extend the empirical evidence to advanced countries. The studies on Europe produced mixed results, most supporting a unidirectionally Granger-cause between stock prices and exchange rates for the post-Euro era while the opposite is true for the pre-Euro era (Murinde and Poshakwale 2004; Stavarek 2005). At least one reason for that is the observed heterogeneous integrity of each market with the EU which is in line with the dynamic nature of the transition process. According to Stavarek (2005), exchange rates and stock market indices stemming from more developed economies (e.g., UK, Germany, and France) are more likely to result in a unidirectional Granger-causal relationship than from post-communistic EU-member countries (e.g., Hungary, Poland, Slovakia, and the Czech Republic). In summary, the causal relationship between exchanges rate and equities in Europe is inconclusive. Therefore, especially in the light of forecasting and policy implications shedding additional insights from a time-domain framework is more than necessary. Based on the idea that each market component has its own reaction time to the information, causal and lead-lag relations might exist at different frequencies (Dacorogna et al. 2001). Hence, the linkage bonding stock markets and exchange rates can vary across frequencies and time. Decomposing a time

series into several time scales unearths some economic time-frequency relations and the transient effect that traditional methods do not deliver.

3.3 Methodology

We use the continuous wavelet transform to detect bidirectional causalities in the time-frequency domain. Since the use of wavelets is a well-established technique, in this section we only present the essential methods applied here.
We make use of the WPS, CWP, CWC, and PD, as proposed by Hudgins, Friehe, and Mayer (1993) and Torrence and Compo (1998) to analyze possible transient effects between the nominal effective European exchange rate and the European stock market. The tools we use to analyze and quantify time-frequency dependencies between two time-series are functions of time and scale/frequency. The WPS describes the evolution of the variance of a time series at different frequencies, which is to identify volatility hotspots in the time-frequency domain. The CWP is a multivariate method, which depicts the local covariance at each time and frequency, analogous to the traditional covariance. The CWC is the localized (both in time and frequency) correlation between two time-series and is normalized by its power spectrum. The key advantage of CWC over CWP lies in its ability to identify strong peaks even for the realization of independent processes suggesting the possibility of spurious significance tests (Maraun and Kurths 2004). Also, we make use of the PD to reveal lead-lag relationships between the two time series. As the phase of a time series describes its position in the pseudo-cycle in the time-frequency domain, the PD describes the relative position of the two times series.
In the following, we solely focus on the key elements of the wavelet transform. However, the interested reader is referred to Dacorogna et al. (2001) and Chernick (2001) for a more complete and insightful development of the theory.
For a given time-series $x_t \in L^2(\mathbb{R})$, its continuous wavelet transform is a function of time t and a scaling factor s which controls the width of the wavelet:

$$W_x(\tau, s) = \int x_t \left[\frac{1}{\sqrt{|s|}} \overline{\psi} \left(\frac{t-\tau}{s} \right) \right] dt, \quad s, \tau \in \mathbb{R}, \tag{3.1}$$

where $\overline{\psi}$ denotes the complex conjugation, τ a translation parameter controlling the location in time and ψ is defined as the mother wavelet satisfying $\int_{-\infty}^{\infty} \psi(t)dt = 0$. The normalization factor $\frac{1}{\sqrt{|s|}}$ ensures comparability across frequency bands and time. Scaling a wavelet means stretching it ($|s| > 1$) or compressing it ($|s| < 1$) while translating it means shifting its position in time.
The choice of the mother wavelet depends on the particular application. As we intend to analyze bidirectional causalities between time series, we need a complex-valued wavelet to preserve information on amplitude and phase during the wavelet transform. The choice of the mother wavelet is limited to analytical wavelets as they provide an estimate of the instantaneous amplitude and instantaneous phase. The wavelet is further characterized by how it is localized in time (δt) and frequency ($\delta\omega$). According to the Heisenberg uncertainty principle, the real trade-off between localization in time and frequency limits how small the uncertainty product $(\delta t) \cdot (\delta\omega)$ can be (Heisenberg 1927). In case of the Morlet wavelet, which is an analytical wavelet as long as it is conveniently parametrized, the uncertainty product attains the lowest possible theoretical value $\sigma_{t,\psi_{\omega_0}}\sigma_{\omega;\psi_{\omega_0}} = \frac{1}{2}$. Besides, the time radius and frequency radius are equally providing a good balance between time and frequency concentration (Aguiar-Conraria and Soares 2011; Grinsted, Moore, and Jevrejeva 2004). The Morlet mother wavelet is given by:

$$\psi_{\omega_0}(t) = Ke^{i\omega_0 t}e^{-\frac{t^2}{2}}, \tag{3.2}$$

with $i := \sqrt{-1}$ as an imaginary unit, t as a dimensionless time parameter and ω_0 as a dimensionless frequency parameter controlling the number of oscillations within the Gaussian envelope. The normalization parameter K is restricted to $\pi^{-\frac{1}{4}}$ as it ensures $\psi_{\omega_0}(t)$ to have unit energy. We set $\omega_0 = 6$ to ensure an inverse relation between wavelet scales and frequencies, $f = \frac{6}{2\pi} \approx \frac{1}{s}$ which simplifies the interpretation of the results. A robustness check has been performed for a range of parameters values which imply that the Heisenberg box area is close to its lower bound as proposed by Foufoula-Georgiou and Kumar (1994). In summary, for any reasonable parameter for ω_0, the results were similar or almost indistinguishable from the results presented in section 3.4.

In the wavelet domain, we can measure the relative contribution at each frequency

and at each scale to the time series variance by the WPS:

$$(WPS)_x(\tau, s) = |W_x(\tau, s)|^2 \tag{3.3}$$

In order to identify dominant frequencies in term of variation we average equation 3.3 over time. In a similar way, we are able to obtain the local covariance between a time series $x(t)$ and $y(t)$ at each time and frequency by the CWP, defined as:

$$W_{xy}(\tau, s) = W_x(\tau, s)\overline{W}_y(\tau, s) \tag{3.4}$$

where W_x and W_y are the wavelet transforms of the respective time series $x(t)$ and $y(t)$. While the CWP measure reveals areas with high common power, we understand the CWC, as defined by Torrence and Compo (1998), as a localized correlation coefficient in time frequency space. More specifically, it measures the co-movement of two time series over time and across frequencies. The key advantage of the cross wavelet measure is that it is normalized by the power spectrum of the two time series. A coherence factor close to one implies a strong co-movement between the time series while the opposite is true for a coherence factor close to zero. For the given time series $x(t)$ and $y(t)$ the CWC is defined as:

$$R_{xy}(\tau, s) = \frac{|S(W_{xy}(\tau, s)|}{\sqrt{S(|W_{xx}(\tau, s)|)S(|W_{yy}(\tau, s)|}} \tag{3.5}$$

where S denotes a smoothing operator in both time and frequency. Following Torrence and Compo (1998), a suitable smoothing operator with a similar footprint as the the Morlet wavelet is given by:

$$S(W) = S_{scale}(S_{time}(W_n(s))) \tag{3.6}$$

where S_{scale} denotes smoothing along the wavelet scale axis and S_{time} smoothing in

time with:

$$S_{scale}(W)|n = (W_n(s) * c_2\Pi(0.6s))\Big|_n \tag{3.7}$$

$$S_{time}(W)|s = \left(W_n(s) * c_1^{-\frac{t^2}{2s^2}}\right)\Big|_s \tag{3.8}$$

where c_1 and c_2 are denoted as normalization constants and Π as the rectangle function. Both convolution are done discretely which enables to determine the normalizations coefficients numerically. We asses the statistical inference of the wavelet power and coherence based on Monte Carlo simulations.
The defintion of CWC is essential to our application. Despite the fact that it indicates comovement, it enables to distinguish between cross-market interdependence and contagion depending on the frequency. Significant CWC at the lower frequencies indicates interdependence whereas significant CWC at higher frequencies indicate cross-market contagion (Bodart and Candelon 2009; Saiti, Bacha, and Masih 2016).

Based on wavelet transform in equation 3.1 we obtain the phase angle of the time series $x(t)$ by:

$$\rho_x(\tau, s) = tan^{-1}\left[\frac{\mathfrak{I}\{W_x(\tau, s)\}}{\mathfrak{R}\{W_x(\tau, s)\}}\right], \tag{3.9}$$

where $\mathfrak{R}(W_x)$ and $\mathfrak{I}(W_x)$ are the real and imaginary parts of the wavelet tranform W_x, respectively. The phase angle reveals the oscillation position of a time series $x(t)$ as a function of time and scale/frequency. The difference between the phases angles of two time-series reveals possible delays of their oscillations. In order to evaluate the relationship between two time series we compute the difference from the phase angle of the cross-wavelet transform given by equation 3.4:

$$\rho_{xy}(\tau, s) = \rho_x(\tau, s) - \rho_y(\tau, s) = tan^{-1}\left[\frac{\mathfrak{I}\{W_{xy}(\tau, s)\}}{\mathfrak{R}\{W_{xy}(\tau, s)\}}\right], \tag{3.10}$$

with $\rho_{xy} \in [-\pi, \pi]$. We adhere to previous studies in relation to the interpretation of the phase difference ρ_{xy} with respect to the sign of correlation between $x(t)$ and $y(t)$:

- If $\rho_{xy} \in (-\frac{\pi}{2}, \frac{\pi}{2})$ *then* in-phase (positive correlation).
- If $\rho_{xy} \in (\frac{\pi}{2}, \pi) \cup (-\pi, -\frac{\pi}{2})$ *then* out-of-phase (negative correlation).

- If $\rho_{xy} = \pi$ or $\rho_{xy} = -\pi$ *then* anti-phase (no correlation).

The phase difference contains information regarding lead-lag relationships. However, the interpretation of the phase difference in relation to lead-lag relationships has been subjected to multiple interpretations in the literature. As a result, studies analysing lead-lag relationships based on the PD have arrived at incorrect conclusions due to the misinterpretation of the PD. It is unclear why different interpretations of the PD exist in the literature. In fact, they have been presented without any explanation. Funashima (2017) provides an overview of the three strands of interpretation used in the literature and identifies a plausible interpretation of the PD. His work addresses the gaps in the existing literature, premised on an analysis based on simulated as well as on empirical data. In accordance with the empirical results in Funashima (2017), in terms of lead-lag relationships, we interpret ρ_{xy} as following:

- If $\rho_{xy} \in (0, \frac{\pi}{2}) \cup \rho_{xy} \in (-\pi, -\frac{\pi}{2})$ *then* $x(t)$ leads $y(t)$.
- If $\rho_{xy} \in (-\frac{\pi}{2}, 0) \cup \rho \in (\frac{\pi}{2}, \pi)$ *then* $y(t)$ leads $x(t)$.

3.4 Empirical Results

In this section we present the empirical results of the wavelet decomposition for the nominal effective European exchange rate and the European stock market represented by the eleven most traded stock indices. Each time-series represents thus daily logarithmic returns. Table 3.1 provides the descriptive statistics. The dataset used consists of 4000 observations covering the period 05.01.2000 - 18.05.2017.

3.4.1 Co-Movement and Lead-Lag Relationships

Despite the location of potential co-movement in the time-frequency domain, the phase-difference provides detailed information on lead-lag relationships. This measure gives us information on the delay or synchronization between the two time-series as it implies the position of a time-series in the pseudo-cycle as a function of frequency. Since the theoretical distribution for the wavelet coherence is unknown, we asses the statistical significance based on a Monte Carlo simulation with 10,000 trials, following the approach of Torrence and Compo (1998). We present the results of the cross-wavelet analysis between the Euro and the Euro-12 stock market in figures 3.3-3.6. The horizontal

	Mean	Std	Min	Max	Skewness	Kurtosis
Euro	-0.00002	0.00372	-0.02772	0.02780	0.02518	3.77828
Belgium	-0.00001	0.01274	-0.08319	0.09334	-0.01536	5.65786
France	0.00002	0.01482	-0.09472	0.10595	-0.04260	4.58415
Germany	0.00015	0.01528	-0.08875	0.10797	-0.05521	4.10007
Greece	-0.00036	0.01932	-0.17713	0.13431	-0.30092	6.02502
Ireland	0.00005	0.01476	-0.14955	0.10823	-0.62558	8.71025
Italy	-0.00018	0.01552	-0.13331	0.10877	-0.21653	4.85545
Luxembourg	0.00010	0.01307	-0.11550	0.09104	-0.46431	7.62172
Netherlands	-0.00005	0.01458	-0.09590	0.10028	-0.10984	5.99588
Portugal	-0.00020	0.01216	-0.10379	0.10196	-0.24411	5.70944
Spain	-0.00004	0.01508	-0.13185	0.13484	-0.10271	5.63121
United Kingdom	0.00004	0.01212	-0.09266	0.09384	-0.15437	5.80512

Table 3.1: Descriptive Statistics

axis of these figures denotes the time component while the vertical axis denotes the frequency component in business days beginning from scale 1 (one business day) up to 1024 (four years). Regions of high coherence (dark grey) indicate a strong local correlation while the opposite holds for regions of low coherence (light grey). Time and frequency regions with a significant coherence are framed in a white contour line.
We can observe that co-movement patterns do not only vary among time but also among frequencies. Between 2000 and 2003 we have high coherence regions at 128-512 day frequency band. For the Greek, Portuguese and Irish stock market we observe significant high coherence regions for a greater time span than for the remaining nine stock markets. Between 2007 and 2010, we observe high coherence regions at similar frequency bands as well. The significant coherence patterns beginning in 2007 are more persistent in time for the Spanish, Portuguese and Greek time series than for the remaining dataset. In figure 3.1 we plot the average coherence per frequency over the whole period. This perspective allows us to localize significant correlation patterns in the frequency domain. We observe for every examined pair a significant co-movement between 64 business days (3 months) and 256 business days (one year) as well for the frequencies below 32 days. Interpreting the results in figure 3.1 leads to the conclusion that significant coherence patterns for frequencies lower than 128 days appear to be more severe but not very persistent in time. As illustrated in figure 3.1, the average coherence across frequencies follows an almost uniform pattern for every

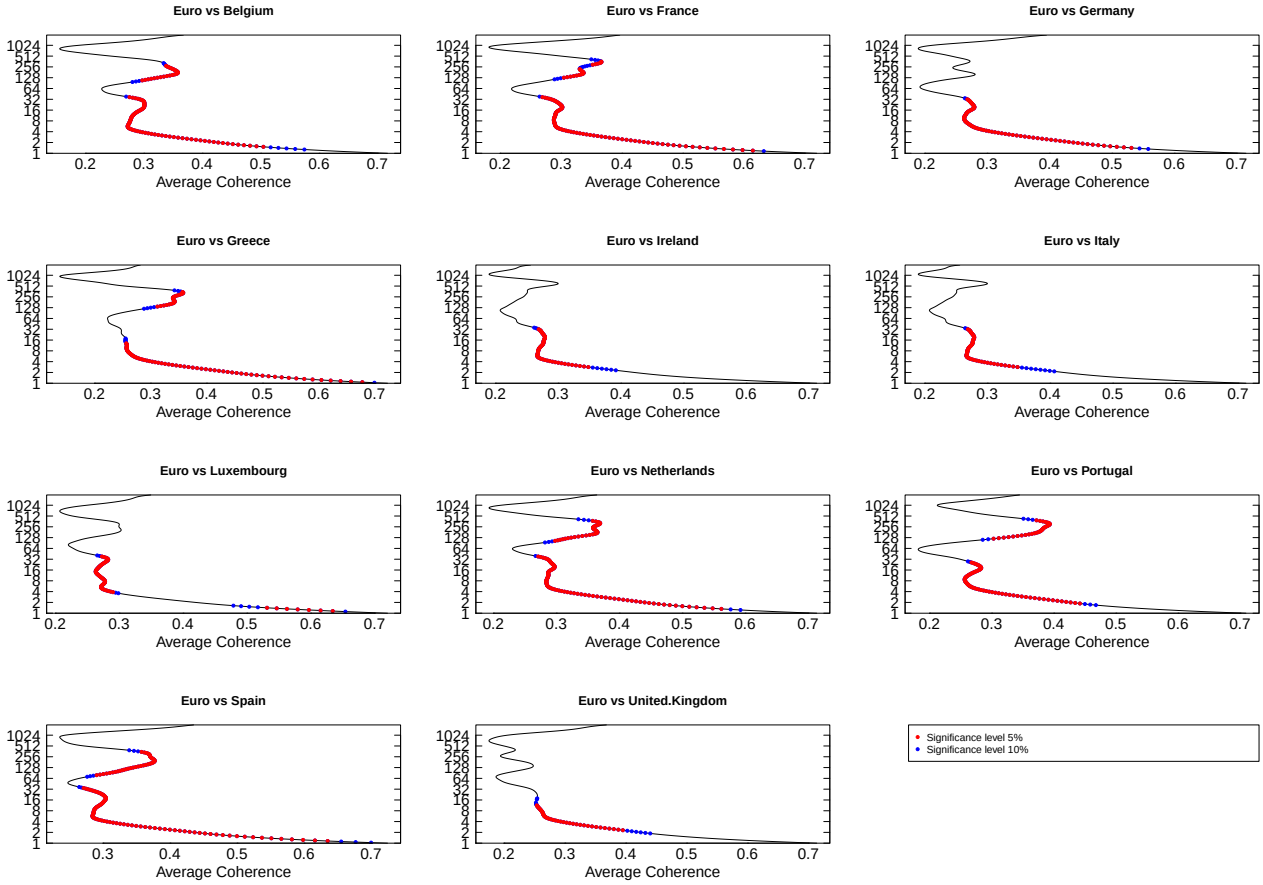


Figure 3.1: Average wavelet coherence per period. The red dots denote a significance level of at least 5% while the blue dots denote a significance level of 10%.

European stock market. The highest significant coherence is observable at the highest frequencies. Between the frequency band 1 and 4, the coherency falls from 0.7-0.6 to 0.3 for every considered stock market. For Belgium, France Greece, Netherlands, Portugal and Spain we observe an increase of the average coherence for the frequency band 64-256 days. Additionally, since the crisis in 2007 we observe for Greece, Portugal and Ireland very slow decaying significant coherence patterns on frequencies lower than six months. This observation favors the existence of time- and frequency-varying cross-market interdependence implying far more stable cross-market linkages than for the remaining dataset.

The findings from the coherence analysis reveal a highly dynamic and significant co-movement pattern for the frequencies bands between 128 and 256 days as well as between 8 and 32 days. Based on this observation we shift our focus during the phase-difference analysis solely between these two frequency bands. The PD reveals an unstable phase relation over the considered time interval, as it constantly varies between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ while we rarely observe anti-phase relations. Also, it is evident that the fluctuation is more severe in the higher frequencies. We observe for each time-series pair a sharp decline in the PD around between 2002 and 2003 towards $\frac{-\pi}{2}$ at the frequency band 128-256 days. This observation indicates a cause-effect relationship in which the Euro leads the respective stock market. The opposite pattern occurs between 2007 and 2008 for each time-series couple at the same frequency band. These phenomena are not observable in the higher frequency band 8-31 days as the PD is constantly fluctuating between $\frac{\pi}{2}$ and $\frac{-\pi}{2}$. Comparing the PD of both frequency bands, we find contradicting results concerning lead-lag relationships. For the lower frequency band (128-256 days) we observe with some exceptions during 2007, a cause-effect relation between the Euro and the European stock markets in which the stock market leads the Euro up to the time of the European debt crisis in 2012. For this particular frequency-band (128-256 days) within the time-interval 2007-2012, the observed lead-lag relationship between both time series is in line with the portfolio approach by Dominguez and Frankel (1993). From that moment we observe for each time-series a PD between 0 and $\frac{-\pi}{2}$ indicating that the Euro leads the stock market, which is in line with Dornbusch and Fischer (1980). In summary, the existing lead-lag relationships as identified by the PD are not persistent in time and do not indicate a dominating directional influence.

3.4.2 Link to the time domain

Wavelet analysis provides with the cross wavelet coherence a tool to measure co-movement in the time-frequency domain. Existing co-movement patterns can be decomposed into several frequencies, which can economically be interpreted as the investment horizon of market participants. As this procedure enables to capture time-frequency regions in which two time-series co-move, we can interpret it as the local correlation while neglecting the complex component of the wavelet. In a second step, we link the observed coherence pattern to econometric tools which operate solely in the time-domain. For this purpose, it is necessary to reduce the coherence to the time domain as well, which is done by averaging it for each point in time over all frequencies. The non-significant coherence values (sign. level $< 95\%$) have been neglected during the averaging. Since the coherence coefficients are in absolute value, we adjust them by the phase difference of the examined time series pair to distinguish between positive and negative correlations. The DCC-GARCH, as well as the unconditional correlation, have been considered as representatives of standard econometric methods. Table 3.2 compares the mean values of correlation between the Euro and the major European stock markets using the whole sample.

	Unconditional	DCC	WTC
Belgium	0.000	-0.004	-0.059
France	-0.015	-0.020	-0.081
Germany	-0.014	-0.013	-0.069
Greece	0.062	0.022	-0.003
Ireland	0.040	-0.060	-0.124
Italy	0.040	-0.060	-0.124
Luxembourg	-0.023	-0.034	-0.075
Netherlands	-0.047	-0.048	-0.130
Portugal	0.031	0.013	0.063
Spain	0.037	0.020	0.015
United.Kingdom	-0.008	0.015	-0.020

Table 3.2: Comparison of the unconditional correlation, DCC-GARCH and the wavelet coherence. The standard deviation is provided in the parentheses.

We observe that the correlation estimates strongly differ from one another. A part of the deviation can be explained by the standard deviation of arithmetic mean of

correlations from the DCC-GARCH as well as from the coherence. The difference is most striking in figure 3.2 in which we plot the unconditional correlation, the DCC-GARCH correlation and the time domain based coherence. We observe that the average wavelet coherence in the time domain has a stronger tendency to switch the direction and to take boundary values. This can be explained by the fact that non significant coherence pattens have been removed a priori.

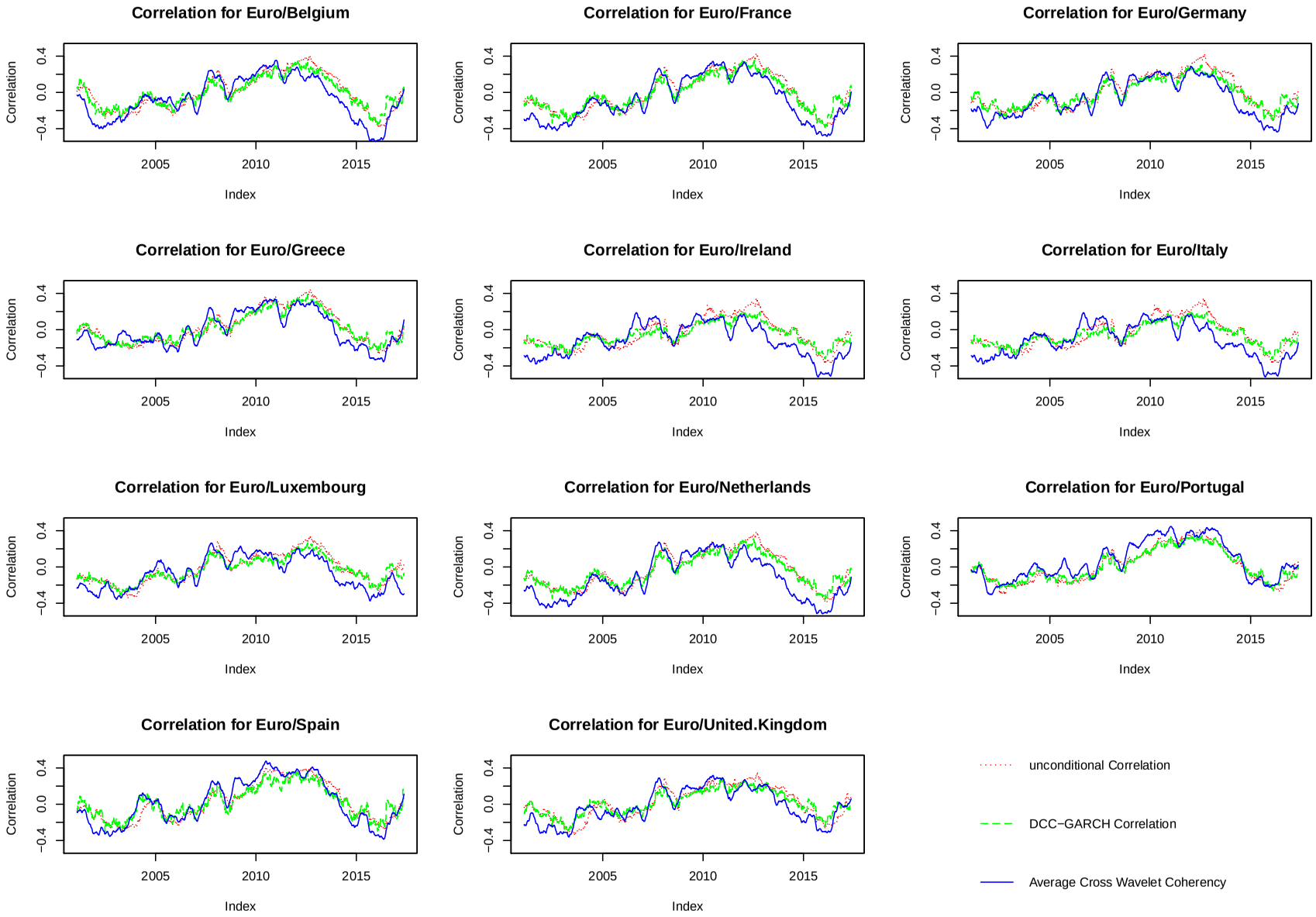


Figure 3.2: Comparison of the time-varying correlations from the unconditional correlation (dotted line), the DCC-GARCH (dashed line) and the wavelet coherence (solid line) between the nominal effective Euro exchange rate and European stock indices pairs. The results are smoothed using a 100-day moving average.

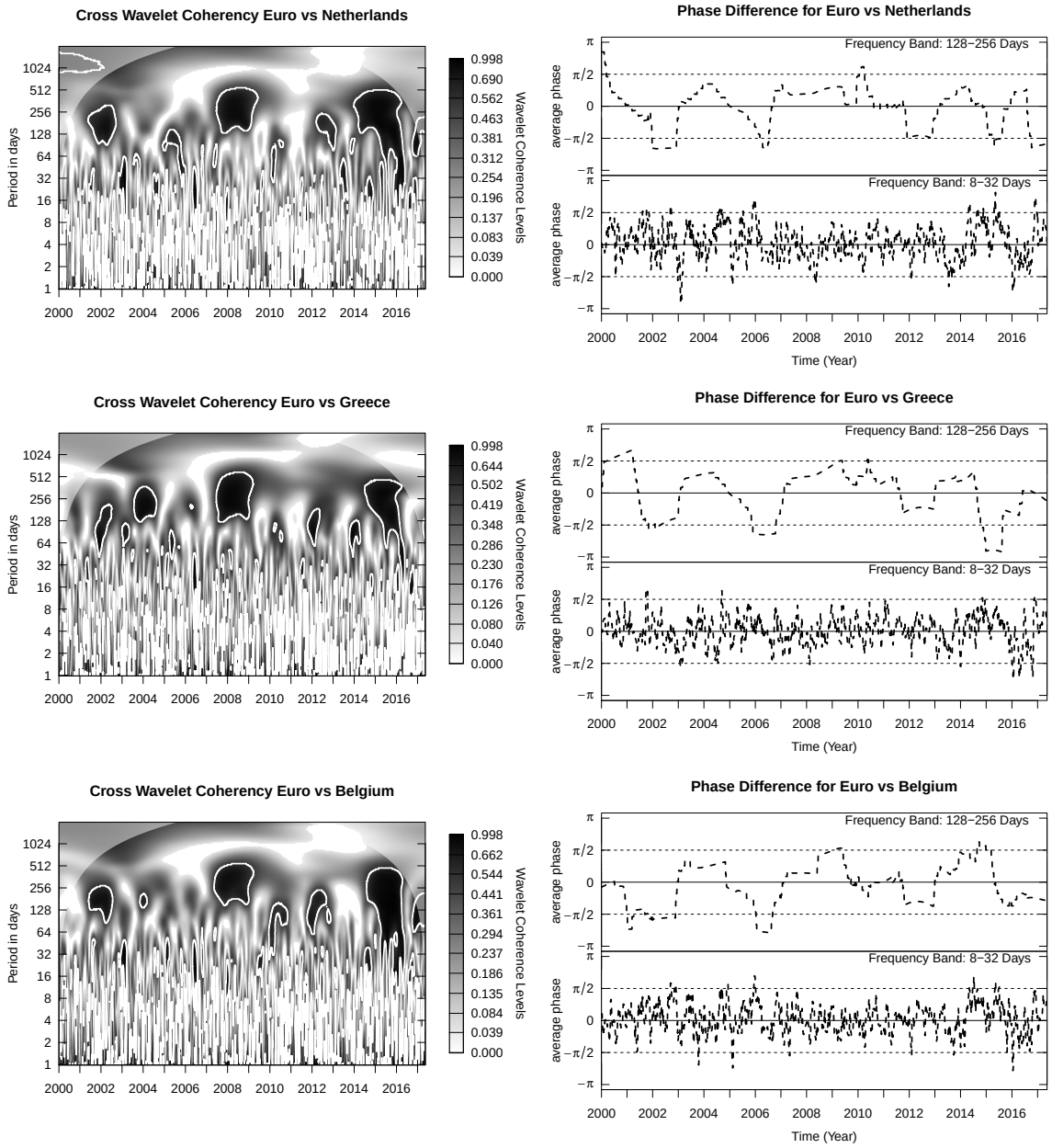


Figure 3.3: On the left we plot the wavelet coherence where the white contour indicates the 5% significance level estimated by Monte Carlo simulations (10,000 trials). The code for coherence ranges from dark grey (high power) to light gray (low power). On the right we plot the phase difference for the frequency bands 256-512 days and 128-256 days..

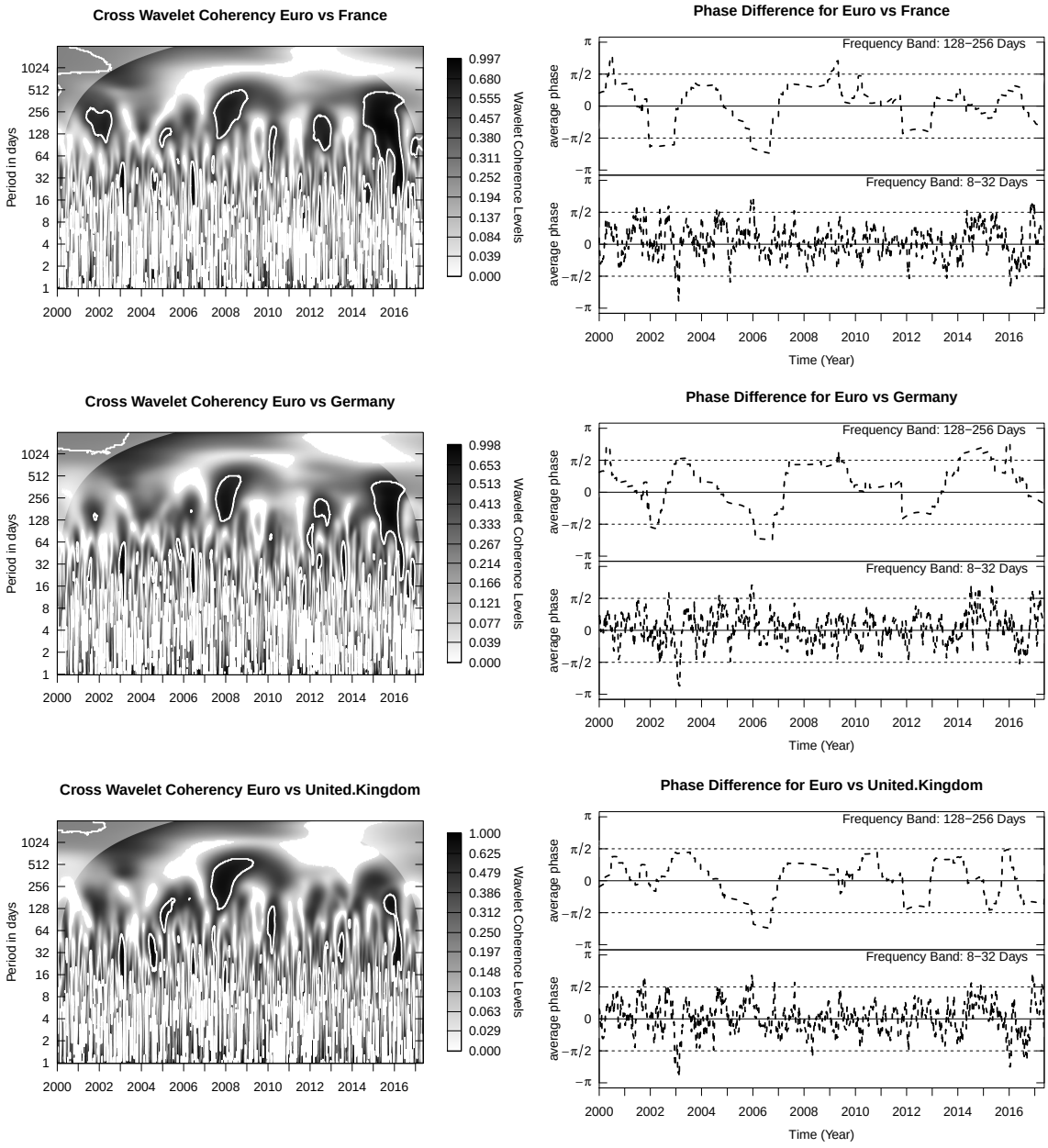


Figure 3.4: On the left we plot the wavelet coherence where the white contour indicates the 5% significance level estimated by Monte Carlo simulations (10,000 trials). The code for coherence ranges from dark grey (high power) to light gray (low power). On the right we plot the phase difference for the frequency bands 256-512 days and 128-256 days.

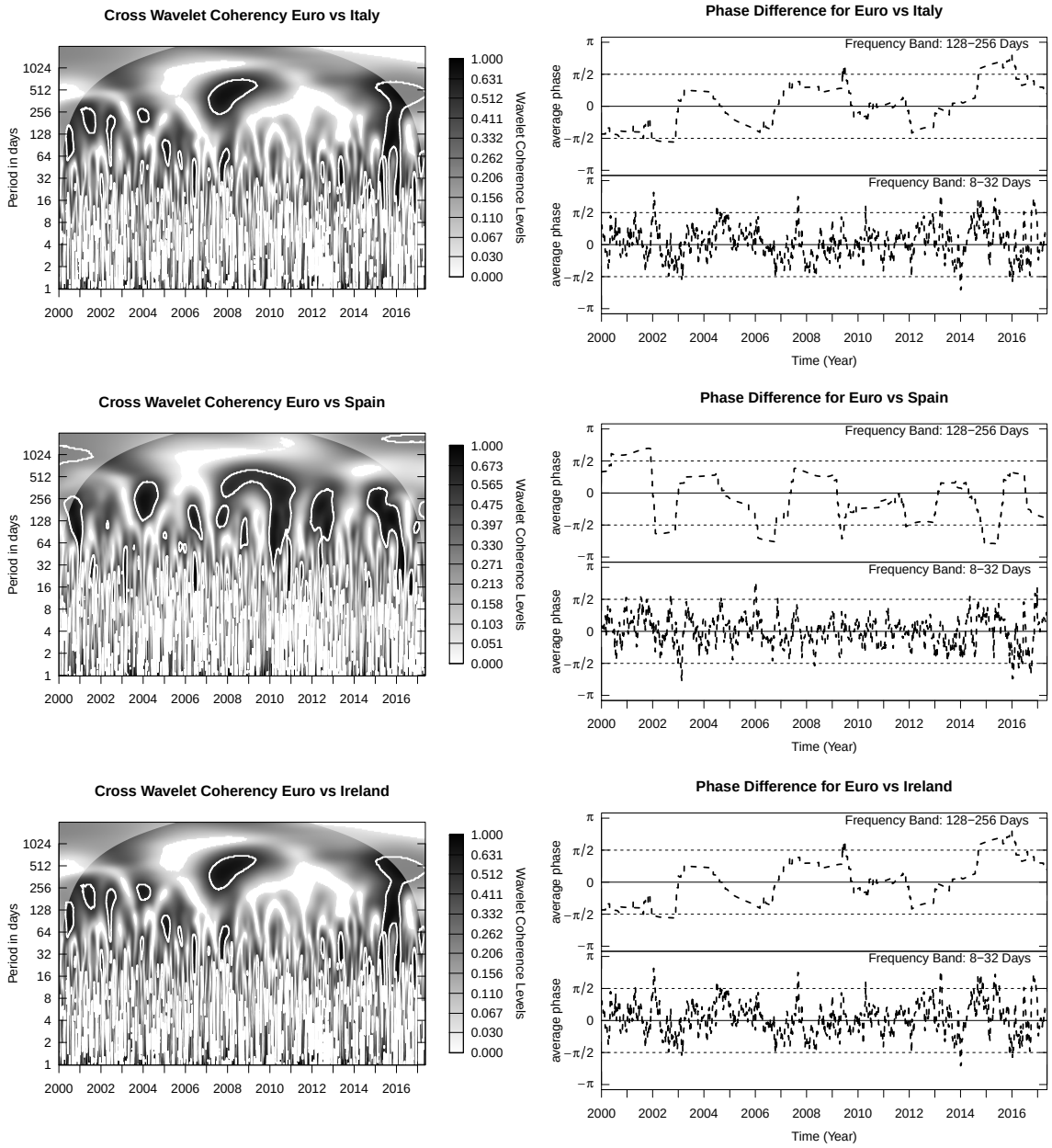


Figure 3.5: On the left we plot the wavelet coherence where the white contour indicates the 5% significance level estimated by Monte Carlo simulations (10,000 trials). The code for coherence ranges from dark grey (high power) to light gray (low power). On the right we plot the phase difference for the frequency bands 256-512 days and 128-256 days..

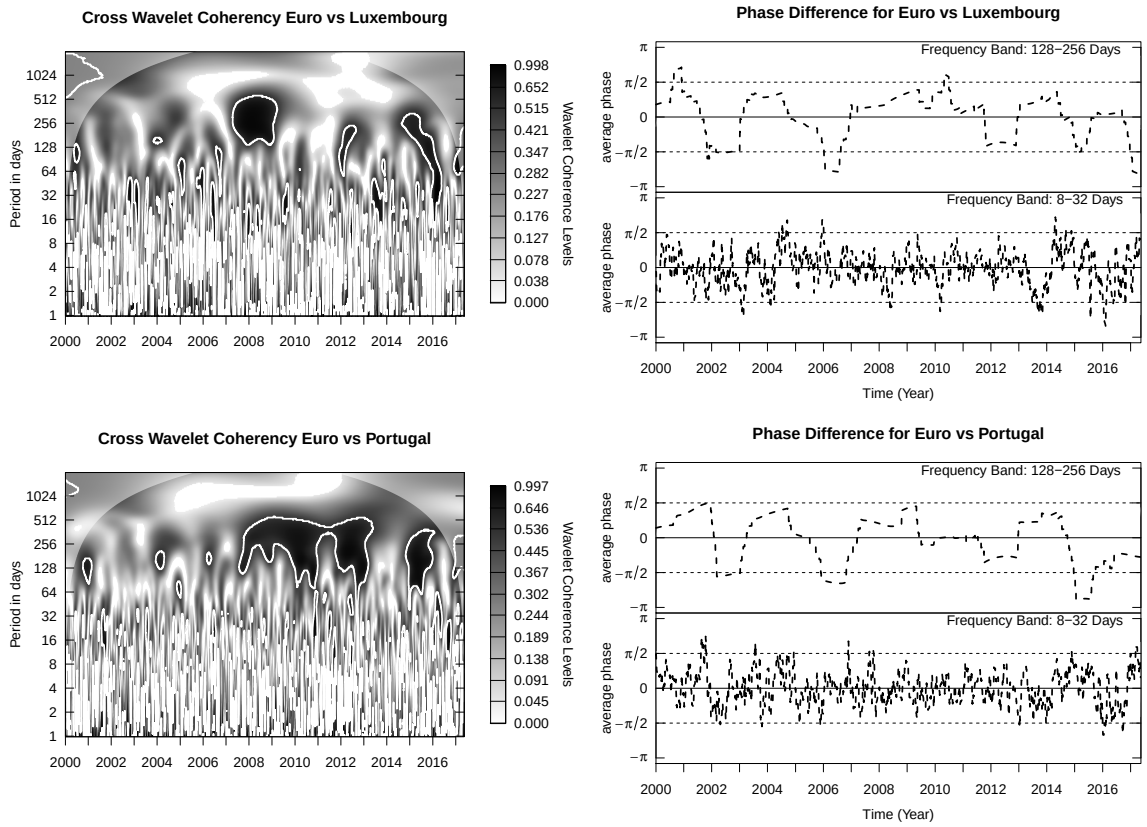


Figure 3.6: On the left we plot the wavelet coherence where the white contour indicates the 5% significance level estimated by Monte Carlo simulations (10,000 trials). The code for coherence ranges from dark grey (high power) to light gray (low power). On the right we plot the phase difference for the frequency bands 256-512 days and 128-256 days.

3.5 Conclusion

We use wavelet analysis to study the causal relationship between the largest European stock markets and the nominal effective Euro exchange rate. Based on the cross-wavelet coherence results, we identify a significant bidirectional causal relationship between 2000 and 2003 as well as between 2007 and 2010 at 128-256 days time-scale. Additionally, we discover a high degree of interdependence during financial turmoil during 2002, 2007-2010 and 2012-2016. Also, we provide evidence for financial contagion between the nominal effective exchange rate and the European stock markets as significant cross-wavelet coherence be observable in every stock market at frequencies lower than 32 days. The phase difference analysis identifies a time and frequency-varying lead-lag relationship. The bidirectional causal relationship between the nominal effective exchange rate and European stock indices continually fluctuates between a cyclical and an anti-cyclical relationship. However, during the dot-com bubble in 2002, the financial crisis in 2007 and the European debt crisis in 2012 we observe homogeneous lead-lag relationships over the entire dataset. Further, the observed phase pattern and thus the implicated non-stable lead-lag relationship favor solely for specific time intervals the Dornbusch model by Dornbusch and Fischer (1980), as well as the portfolio approach by Dominguez and Frankel (1993).

Based on a time-frequency analysis we were not only able to distinguish between contagion and interdependence but also to identify lead-lag relationships. Knowledge of the directional causality is of the essence for developing appropriate policy implications. While we recognize the limits of the bivariate approach, in the future, researchers might apply a multivariate approach to revisit the relation. It might be of interest to assess the nexus by including variables, e.g., interest rates, oil price return, money supply, and other macroeconomic variables as can be justified on theoretical grounds in the particular economic context.

4 Event detection and the impact of surprise.

4.1 Introduction

According to the infamous Efficient Market Hypothesis (EMH) market prices fully and instantaneously incorporate all information and expectations of market participants. Weaker forms of the EMH relax these strong assumptions to a semi-strong and weak form efficient markets in which only publicly available information and historical information is reflected in the prices. Fama (1997) argues market efficiency not to hold at any single point but on average over time and agents implying anomalies to vanish subsequent to their discovery. In light of this theory, agents have stable and well-defined preferences which they apply in maximizing their expected utility when facing risky investment decisions. Empirically there is evidence that such behaviour leads participants to reflect new information accurately and mispricing quickly vanish, e.g. (Malkiel 2003).

With the rise of the internet new sources of information have gained the agent's attention making it more time consuming and thus more expensive to monitor the relevant information stream. Active users on online platforms like Facebook or Twitter produce content about events in real time (Sakaki, Okazaki, and Matsuo 2010) making them a relevant stream of information (Aiello et al. 2013).

While more timely information is beneficial, empirical findings are pointing to behavioural biases such as herding or overreaction to unexpected events (Thaler and Benartzi 1985). The impact of such unexpected events on financial markets is extensively discussed within the context of interest rate, surprise indexes, and earnings (Han 2010; Scotti 2016; Bernard and Thomas 1989; Brown 2003).

We use Twitter to quantify the presence of surprise and the effect on the S&P 500. We develop a three-step algorithm to mimic the agent's behaviour and apply modern techniques in natural language processing to analyse it empirically. Firstly we access for any event if the news triggers a discrepancy from prior beliefs. When an agent

is exposed to new information (tweets), this is only considered as a 'surprise' if the distance from earlier beliefs is sufficiently high and the 'surprise threshold' is crossed. If this happens the agent experiences surprise, leading to an interruption of ongoing mental process and resources such as attention are reallocated towards the unexpected. In a second step, the event is analysed by the verification of the schema discrepancy, the study of the causes of the unexpected event, the evaluation of the unexpected event's significance and the assessment of the event's relevance for ongoing action. The final step incorporates the schema update that involves producing the immediate reactions to the unexpected event such as the update, extension, or revision of the schema or sets of beliefs that gave rise to a discrepancy. Under the EMH we would observe increased volatility that epitomizes the price adjustment as a reaction to the unexpected event.

Empirically we perform content analysis through topic modelling on all information provided by the Twitter accounts of the S&P 500 companies within a 31-month period, and we quantify the surprise level of financial tweets within a time-frame of 24 hours. We integrate the informational content of the Twitter ecosystem into a network topology which allows quantifying the surprise level by gaining awareness of propagating rare events on an aggregate level. Specifically, we investigate whether Twitter covers topics that are relevant to the financial markets and especially volatility modelling. In doing so, we focus on quantifying the propagation of rare events to explain a reaction to the financial markets.

We contribute to the literature by providing a novel approach which allows quantifying the propagation of information within the network on an aggregated level from an agnostic point of view concerning the topic and information domain. By analysing the impact of aggregated informational content on the S&P 500 return and volatility we provide the first empirical assessment of surprise on the market. We condition the return and volatility response in the pre- and post-announcement periods on realized surprise to identify pre-report effects and to capture the evolution of surprise in time.

There are two branches of literature associated with our approach. The first analyses text with sentiment classification algorithms. Sprenger et al. (2014) use a naïve Bayesian algorithm to classify the sentiment of 250,000 stock-related tweets and analyse its association with stock returns and trading volume. Liu et al. (2017) use a polarity model of sentiment to compute a sentiment score of a randomly selected dataset stemming from a financial forum and analyse its predictive information about the stock market.

Gilbert and Karahalios (2010) show the non-financial information sources to provide novel information about future stock movements. The impact of Twitter sentiment has been analysed during financial events such as IPOs, earning announcements or Federal Open Market Committee (FOMC). Liew and Wang (2016) found a positive relationship between tweet sentiment and IPO performance within a cross-sectional analysis. These findings are also in line with Liew, Guo, and Zhang (2016), who shows that in addition to the consensus earnings of crowdsourced information, tweet sentiment before the earnings announcement can predict stock market returns. The second branch of literature focuses on the impact of Twitter users and information diffusion and propagation. Cha et al. (2010) measure the influence of Twitter users based on three measures: in-degree, retweets, and mentions. Based on a large dataset of 54 million Twitter accounts, they found users with a large number of followers (in-degree) cannot be characterized as influential which has been measured based on retweets and mentions. This is in line with Romero et al. (2011) who shows the linear correlation between user popularity concerning followers and influence to be weaker that one might expect. The study also reveals that the majority of Twitter users consumes information as a passive medium and do not forward the information to the network per se. This observation implies that an individual needs to overcome user passivity to influence instead of attracting a large follower base. As a consequence non-popular individuals may have a further reach than the average, regardless of their popularity.

The remainder of this paper continues with the methodology and data processing in section 4.2 and discuss results in section 4.3. The final section concludes.

4.2 Methodology

We link the aggregate level of surprise in the ecosystem of Twitter to real-world events. An event is an arbitrary location of a space-time region which consists of actively participating agents or passive factors. As we are interested in the general surprise level, we do not target the identification of specific events but the quantification of the surprise level in a broader context. The Twitter ecosystem is suitable for our experiment as in a more general definition events can be of any scale and hence influence an arbitrary amount of people inducing them to tweet about it. Furthermore, events can allocate in spatial and temporal regions utilizing metadata which are specific to each tweet. To

quantify the surprise level in real-time, we introduce a hierarchical processing pipeline. The first part of our approach is to pre-process the raw Twitter data. This step is critical as Twitter data is noisy, redundant and unstructured. The informational content within an unfiltered Twitter stream is sparse, making aggressive preprocessing essential. In a second step, we hierarchically cluster the tweets and extract the keywords that are representative of the context embedded in the tweet. In the final step, we construct a network to explain the dynamics of information flow and compute the aggregated surprise level based on a specially developed method. To assess the impact of the proposed surprise index on the daily volatility of the S&P500 index, we follow a two-step econometric methodology proposed by Andersen and Bollerslev (1998). Specifically, we estimate the conditional means of returns and fit a conditional volatility model to the predicted residuals. The fitted conditional volatility is then used in a weighted least squares (WLS) procedure to obtain heteroskedasticity-robust standard errors in the conditional mean equation. In addition, we investigate the predictability of the S&P 500 volatility when the proposed surprise metric is incorporated in the GARCH(1,1) model.

4.2.1 Data Collection

Event detection from a text-based stream can enable agents to react on the news without human involvement, a critical requirement for scaling up in large portfolios. To investigate event detection we construct a dataset consisting of real-world text-based information stream that links to events. We extract 1,568,572 unique tweets from the public API of Twitter which covers the period between 01/01/2015 and 01/06/2017. We capture every tweet published by every verified Twitter account belonging to one of the S&P 500 companies. Restricting the Twitter stream to specific accounts, namely to verified accounts belonging to the S&P 500 companies makes it safe to assume that the observed rising events are linked to the brand and thus to the company holding the Twitter account.

To analyse the impact of surprise on a financial system, we choose the S&P 500 index prices [1] as a representative economic agglomeration. We compute daily returns

[1] We obtained the S&P 500 index data from the Chicago Board Options Exchange.

as follows:

$$R_t = 100 \times ln(P_t/P_{t-1}) \tag{4.1}$$

where P_t is the end-of-day closing price of S&P 500 index. The descriptive statistics for daily returns and absolute daily returns are provided in table 4.1.

Summary statistics of S&P 500 daily returns.

	daily returns	daily absolute returns
Mean	−0.026	0.584
Median.	−0.005	0.377
Std. dev.	0.834	0.596
Interquartile range	0.757	0.684
JB	226.3131***	1066.821 ***
ADF unit root test	−9.408***	−5.691***
LB	0.091	53.646***
Obs.	627	627

Table 4.1: The sample period is January 1, 2015 - June 30, 2017. Interquartile range is the difference between the 75 th and 25 th percentiles. Null hypothesis of the Jarque–Bera test is the normality of the series in investigation. Null hypothesis of the Augmented Dickey–Fuller unit root test is the existence of a unit root. Null hypothesis of the Ljung Box test is no serial correlation. The stars on the test statistic correspond to p-values ≤ 0.01 ***, ≤ 0.05 ** or ≤ 0.1 *.

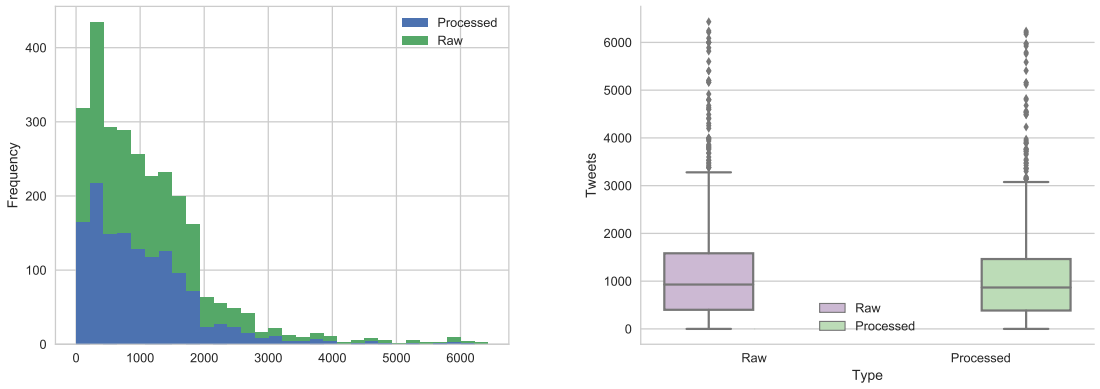
The average daily return is close to zero, while the average absolute return is 0.58%. We reject the hypothesis of normality and unit root according to the results of Jarque-Bera and augmented Dickey-Fuller tests for both returns and absolute returns. The results of the Ljung-Box test indicate the existence of serial correlation in the return series, presumably due to microstructure effects (Andersen et al. 2003).

4.2.2 Data Pre-Processing

The social micro-blogging service provided by Twitter is solely restricted by the number of total characters (140 before and 280 since September 2017). As a result of the character restrictions, the informational content of tweets is highly compressed, and users refrain eg. from the utilization of proper punctuation and stop-words. Although Twitter provides almost real-time information and discussion of current events most tweets

appear to be highly fragmented, noisy and contain irrelevant events as well. Traditional natural language processing algorithms are designed to work on carefully written and well-structured texts instead of grammatically, semantically and syntactically messy texts (Liu et al. 2013). To counteract this, we apply an aggressive pre-processing and filtering which is constructed as follows.

First, we normalize each tweet by removing links, user mentions, digits, hashtags, and any excessive punctuation. Next, we tokenize each tweet and perform spell-checking using vector representations for words (Pennington, Socher, and Manning 2014). We filter out tweets that have less than three text tokens. This structure-based filtering identifies noisy tweets which do not carry enough news-like content. Reprocessing reduced the Twitter dataset by 106,009 (-6.76%) tweets leading to 1,462,563 remaining tweets. The effect of preprocessing on the dataset is visualized in figure 4.1a and 4.1b. Each of the 1379 Twitter accounts generated within the considered period of 31 months



(a) **(b)**

Figure 4.1: Activity on Twitter per account covering the period between 01/01/2015 and 01/06/2017.

on average 1137 tweets, preprocessing removed on average 76 tweets per account.

In the second step, we construct a binary tweet-term matrix based on the filtered tweets and compute the term frequency and inverse term frequency (TF-IDF) to identify highly informative features. The frequency based term weighting evaluates the importance of each term to a corpus d by discounting the term frequency within the corpus by the logarithmic inverse term frequency. Given the TF-IDF score, we

hierarchically cluster the tweets based on the pairwise cosine distance. Given the assumption that tweets belonging to the same broad topic do also possess similar semantic structure it is safe to assume that each cluster, which is formed by the pairwise cosine distance of each TF-IDF, represents a topic and at least one sub-topic. We cut the resulting dendrogram at a fixed threshold of $\theta = 0.5$, producing the final set of clusters for the corresponding time slot. This implies that two tweets with a pairwise cosine distance smaller than 0.5 belong to the same cluster. The threshold variable θ controls how loose or tight the clusters will be. A higher threshold would lead to a weaker segmentation which bears the risk of collating different topics in the same cluster. A lower threshold would result in tighter clusters and thus to a stronger segmentation. However, the likelihood that several clusters represent the same topic increases as well.

4.2.3 Topic Extraction

Each cluster forms a set of numerous event instances that are clustered superficially under the purview of their subject. One possible approach to extract specific events at first level is to bypass the semantic clustering. However, given the significant amount of data produced by Twitter, it is consequently not advisable to zero down on specific events without a top-level segmentation. Also, by following the proposed high-level topic segmentation, we can accomplish event identification and track within the logical domain of a topic. This prevents identical entities to be erroneously clustered together in case they belong to unrelated topics. We extract the topics for each cluster using the unsupervised topic modelling technique latent Dirichlet allocation (LDA) which infers topic distributions in documents. The traditional LDA is designed for well-structured documents, such as news articles in which it delivers satisfactory results (Steyvers and Griffiths 2010). Due to the highly compressed, sparse and noisy structure of tweets, the standard LDA fails to infer topic distributions when it is applied to microblog posts. As the standard LDA is built on the assumption that the underlying document represents a mixture of topics it does not incorporate an essential observation concerning Twitter: Due to character limitation, tweets mostly cover a single topic. A possible solution is given by aggregating tweets of a user in a single document. As a result, we receive the topic profile of each user which is not in accordance with our objective of identifying and tracking events in time but serves the profiling of single users. This has been first

facilitated by Zhao et al. (2011) and allows to model topic distributions at a user level. Given that a user selects a topic from his favourite list of topics and embeds it among background words within a tweet. We use a slightly modified version of the Twitter-LDA as proposed in Zhao et al. (2011) to capture this particular behaviour. Our approach allows us to model topic distributions at separate time segments and does not require a user-specific tweet history. Instead of clustering tweets based on users, we aggregate tweets based on the TF-IDF pattern which is derived from the tweet-term matrix to obtain more coherent and logical topic clusters. We introduce a topic distribution θ^c for each cluster c whereas every tweet is generated by picking a topic t from its cluster topic distribution respectively. In the next step, the tweet is populated by additional topic words for the topic t following the distribution ϕ_t and background words following the distribution ϕ_B, forming a bag of words. An arbitrary word belongs to the background class or the topic class according to a Bernoulli distribution π. Some symmetric Dirichlet distribution governs each multinomial distribution. We

Algorithm 1 Twitter-LDA Topic Extraction

draw $\phi^B \sim Dir(\beta), \pi \sim Dir(\gamma)$
for `topic in` $t = 1, ..., T$ **do**
  draw $\phi^t \sim Dir(\beta)$
for `cluster in` $c = 1, ..., C$, **do**
  draw $\theta^c \sim Dir(\alpha)$
  for `tweet in` $s = 1, ..., N_c$ **do**
    draw $z_{c,s} \sim Multi(\theta^c)$
    for `word in` $n = 1, ...N_{c,s}$ **do** draw $y_{c,s,n} \sim Multi(\pi)$
      if $y_{c,s,n} = 0$ **then return** $w_{c,s,n} \sim Mult(\phi^B)$
      else if $y_{c,s,n} = 1$ **then return** $w_{c,s,n} \sim Mult(\phi^B)$

illustrate the modified Twitter-LDA in pseudo-code 1. The parameters α, β, γ specify the Dirichlet distributions. We use collapsed Gibbs sampling to obtain samples of the topic and background words assignment and to estimate the model parameters from these samples.

As an extracted topic can be mentioned in a different context during a time interval, it is not possible to track an event solely by its keywords which we identified by the Twitter-LDA. To counter this, we extract the named entities of an event that are classified as a person, localization, and organization. Due to the specific grammatical

structure of micro-texts, we found that general name-entity recognition (NER) models do not perform well in a noisy environment such as the Twitter stream. For example, the Stanford NER which has been trained on the CoNLL03 shared task dataset achieves 90.8% on that task while drops to 45.% on tweets. That is to the domain mismatch due to a short, mostly informal, often ungrammatical and noise prone structure of tweets that sharply differs from more traditional genres of text (Liu et al. 2013). We use a pre-trained NER model which has been specifically trained for Twitter by Liu et al. (2013). A glance at the results of the NER process reveals that a significant number of tweets does not have any entity mentions. It is highly unlikely that a tweet referring to an important event associated with a high impact does not include a single associated entity. For this reason, we are confident that ignoring these tweets do not distort our overall results but are beneficial to de-noising the Twitter stream.

4.2.4 Event Detection

The Twitter ecosystem can be regarded as a network which offers the potential for studying the spread of information and associated events. A network-based approach can be a useful tool for revealing existing heterogeneity among events according to their level of connectivity. The practical consequence of the graph theory perspective is that we model the Twitter ecosystem as a two-mode network which consists of nodes representing tweets and topic-keywords. A link between a tweet node and a keyword node implies that the keyword has been mentioned in that particular tweet. A link between two keywords implies that both keywords have been mentioned in a discussion which occurs in case a user responds to the initial tweet. Replying to a tweet may also lead to new topics. This is incorporated in the graph by linking the topic-keywords that have been identified in the reply-tweets keywords to the initial tweet's keywords. This allows identifying keywords that are involved in discussions and are thus at some level controversial. Additionally, linking the extracted keywords stemming from the reply-tweet to the initial tweet allows observing whether the specific topics are discussed across multiple tweets. Our graph is fully defined by the set of nodes V and the set of edges E. An edge $e_{i,j}$ connects two single nodes v_i and v_j. By distinguishing between a keyword stemming from a tweet and a keyword stemming from a reply one obtains a directed graph $G = (V, E)$. In an undirected graph, the degree k_i of a node v_i stands for

the number of its neighbouring nodes and is an indicator of the node's connectedness to the network. Highly connected nodes are called hubs and postulate in our case frequently mentioned keywords across multiple tweets and replies. An event is formed by such a locally dense connected group of keywords when each node v_i within the sub-graph C has more links than with the rest of the graph. Specifically, a sub-graph C forms a clique for each node $v_i \in C$ when,

$$k_i^{int}(C) > k_i^{ext}(C) \tag{4.2}$$

where the internal degree k_i^{int} describes the number of edges that connect node i to other nodes in C and the external degree k_i^{ext} describes the number of edges that connect node i to the rest of the network G. For each node i the largest complete subgraph containing node i, given that 4.2 holds, is defined as the maximal clique. We obtain the maximal cliques using the backtracking algorithm implemented by Bron and Kerbosch (1973) and adapted by Tomita, Tanaka, and Takahashi 2006. With progressing time, cliques are affected by the dynamic evolution of the network leading to birth or death, growth or contraction and merging or splitting of cliques. In to order to track changes in the clique structure during the time we define the node with the largest degree within each clique as its identifier. This allows to analyse networks' properties in time as we can observe the evolution of events and quantify their propagation speed in the network. In addition to connectivity measures, we use a strength parameter to further quantify the relevance of each clique to the system. For this purpose, each edge $e_{i,j}$ that connects a tweet to a topic-keyword is assigned a weight w_{ij} which represents the total amount of retweets.

$$s_l = \sum_{C_l} w_{ij}(e_{ij}) \tag{4.3}$$

where w_i the strength of each clique is defined for each topic-keyword node v_j taking into account the size of each topic-cluster i the node is connected too.

4.2.5 Surprise Level

To assess the impact of aggregated events on the financial market we compute an indicator that aggregates for each time frame popularity and speed of propagation

for each topic-keyword. The output of the key indicator is the surprise level for a particular time frame that considers past information within a look-back window θ. The time frame of interest, as well as the look back window, can be chosen arbitrarily. We propose a surprise level method which rewards rare and punishes popular keywords by taking the inverse degree centrality per topic cluster into account:

$$Surprise_t = \frac{1}{V_t} \sum_{j \in G(V_t)} \frac{1}{k^C_{l,t}} \cdot \sigma_{t-\theta,t}(\Delta k^C_{l,t}) \cdot s_{l,t} \tag{4.4}$$

where $k^C_{l,t}$ is the degree centrality of keyword cluster C_l and V the number of nodes and therefore the number of distinguishable objects in the network.at time t. Rare events exhibit a large inverse degree while the value converges to zero for popular events. We take the speed of propagation into account by scaling this factor with the standard deviation of the degrees' first difference between t and $t-1$. The averaged value obtained by 4.4 allows measuring the surprise level of propagating information on an aggregated level. In a second step, we aim to assess its impact on financial markets.

4.2.6 Return Model

We model the daily return, R_t as follows:

$$R_t = \alpha + \sum_{p=1}^{P} \gamma_p R_{t-p} + \sum_{q=1}^{Q} \delta_q \epsilon_{t-q} + \sum_{l} \sum_{j=-1}^{J} \beta_{l,j} S_{t-j} + \epsilon_t \tag{4.5}$$

t=1,...,T (T=912).

l = $\{positive, negative\}$

where R_t is the continuously compounded return as defined in Equation 4.1; $S_{l,t-j}$ is the surprise index; and ϵ_t is the random error term. In the regressions we use the first difference of the surprise index time series. In order to determine the duration of surprise index, we allow the surprise indicator to last for J periods. (1-day intervals from j=-1 to j=J). With this notation, $j = 0$ refers to the time-interval within the Twitter posts are released, while $j = -1$ refers to the interval right before the Twitter post are released, which is included to test for any information leakage effects (Andersen et al. 2003). According to Akaike information criterion, we set $P = 3$, $Q = 0$ and $J = 2$.

We estimate the return model given by equation (4.5) using OLS to obtain the residuals, $\hat{\epsilon}_t$. In the second step, we approximate the time-varying volatility of the residuals $\hat{\epsilon}_t$ by the following linear model using WLS:

$$|\hat{\epsilon}_t| = \mu + \sum_{p=1}^{P'} \gamma'_p |\hat{\epsilon}_{t-p}| + \sum_{q=1}^{Q'} \delta'_q u_{t-q} + \psi \frac{\hat{\sigma}_{d(t)}}{\sqrt{N}} + \sum_{l} \sum_{j=-1}^{J'} \beta'_{l,j} |S_{t-j}| + u_t \tag{4.6}$$

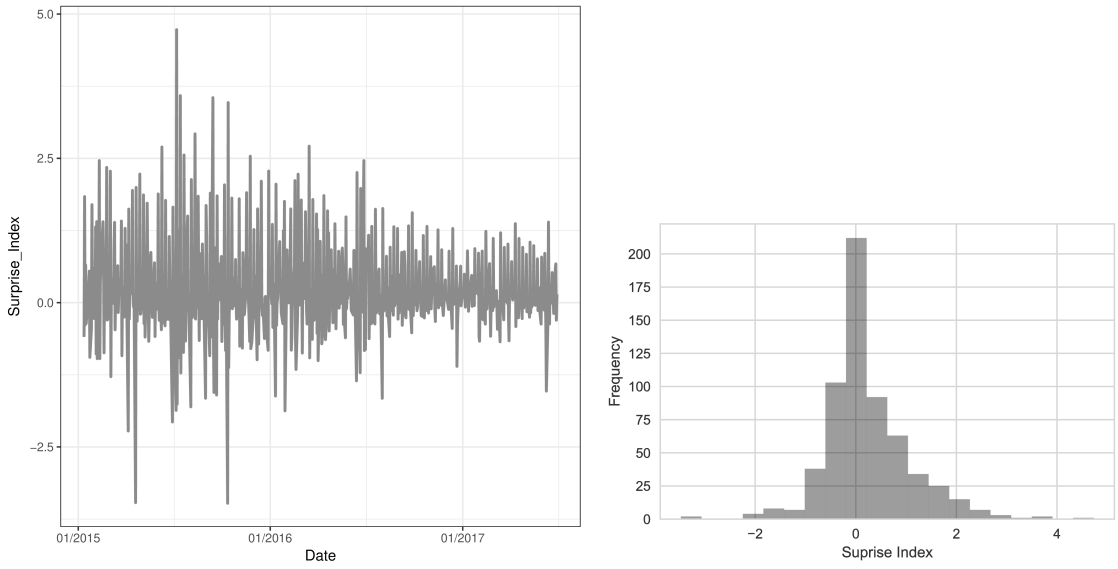
where $t = 1, ..., T(T = 912)$ and $l \in positive, negative$ and $|\hat{\epsilon}_t|$ is the absolute value of estimate residuals from the return model in equation (4.5), $|S_{t-j}|$ the absolute value of the surprise index calculated as in equation (4.4) and u_t is the random error term. Similar to Andersen et al. 2003, the term $\frac{\hat{\sigma}_{d(t)}}{\sqrt{N}}$ describes the average volatility over the trading week w, computed from a Gaussian GARCH(1,1). The first two summation terms represent autoregressive and moving average terms. Again, to determine the impact duration of a surprise shock, we allow the impact to last for J' periods. According to the Akaike information criterion, we set $P' = 4$, $Q' = 1$, $J' = 3$.

4.3 Results

We calculate the proposed metric of surprise for the information published via the Twitter network. The descriptive statistics for the proposed metric of surprise are shown in table 4.2. The distribution of the surprise metric is further illustrated in Figure 4.2a and 4.2b. The ADF test reveals that surprise in the informational content does not follow a unit root process. This means that the surprise metric is not building up over time and reverts to its mean which is line on how we expect an agent to experience surprise as described in chapter 4.1. Accordingly, the Ljung-Box test statistic does not imply any serial correlation. This makes it safe to assume that surprising events do not appear in clusters and form independent entities. While most of the probability mass is surrounded around the mean, we observe rare realizations with a multiple of the standard deviation beyond the lower and upper quartile. In view of the summary statistics in table 4.2 and the Figures 4.2a and 4.2b, the proposed surprised metric behaves very similar to a stock market index return time series.

Summary statistics of the surprise metric	
Mean	0.233
Median	0.0826
Std dev	0.854
Min	−3.476
Max	4.731
Interquartile range	0.800 83
JB	326.15 ***
ADF	−11.414 ***
Ljung Box	76.968 ***
Obs	627

Table 4.2: The sample period is January 1, 2015 - June 30, 2017. Interquartile range is the difference between the 75 th and 25 th percentiles. Null hypothesis of the Jarque–Bera test is the normality of the series in investigation. Null hypothesis of the Augmented Dickey–Fuller unit root test is the existence of a unit root. Null hypothesis of the Ljung Box test is no serial correlation. The stars on the test statistic correspond to p-values ≤ 0.01 ***, ≤ 0.05 ** or ≤ 0.1 *.

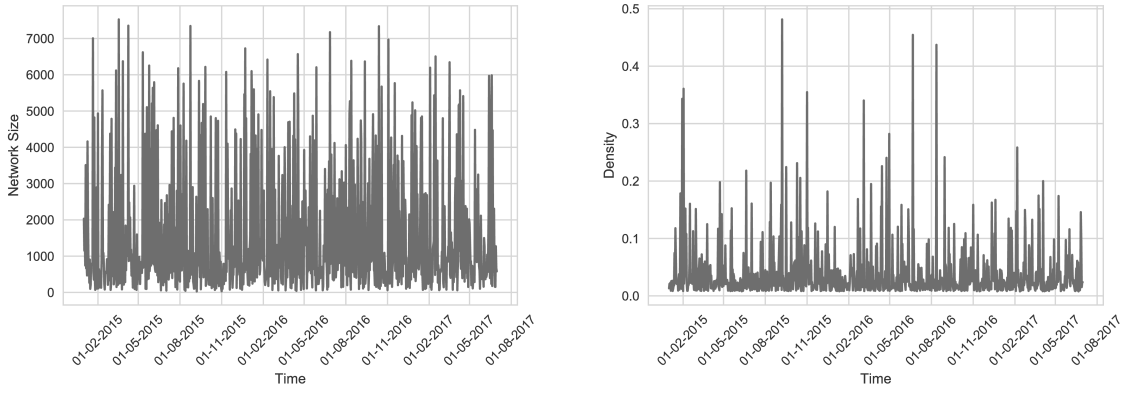


(a) **(b)**

Figure 4.2: (a) Time series plot and (b) histogram for the surprise index based on the activity of the S&P 500 companies on Twitter covering the period between 01/01/2015 and 01/06/2017.

4.3.1 Network Characteristics

A fundamental characteristic of Twitter is its order V in the form of the number of topic keywords included in the network. In the considered period, we observe strong fluctuating number of topic-keywords nodes in the graph (Figure 4.3a). The observed fluctuation in network size implies that the spectrum covered by the identified topics is highly dynamic and heterogeneous across time. A quantity reflecting the interdependence within the discussed topics is the density in the undirected topic partition of the network. Calculated density values in between $\rho_{min} = 0.01$ and $\rho_{max} = 0.1$ indicate a comparatively low linkage among discussed topics. A second



(a) **(b)**

Figure 4.3: (a) Size of the network for each time interval. (b) Average network density for each time interval.

quantity reflecting the connectivity across the network is the average degree $\langle k \rangle$. The degree of a node k_i denotes the number of its direct neighbouring nodes in the graph. In respect of Twitter, a node with a large degree represents a highly connected topic with many subtopics. The average degree is strongly correlated to the size and the density of the network. The cumulative distribution of the degree values in the undirected version of the Twitter topic network is presented in figure 4.4. When a node in the topic network reaches a critical number of edges, new edges cannot connect to it. This indicates that for small k the degree distribution follows a power-law corresponding to an exponent $\gamma = 2.6$, for large k an exponential cut-off develops. In this respect, the Twitter topic network cannot be regarded as a scale-free network.

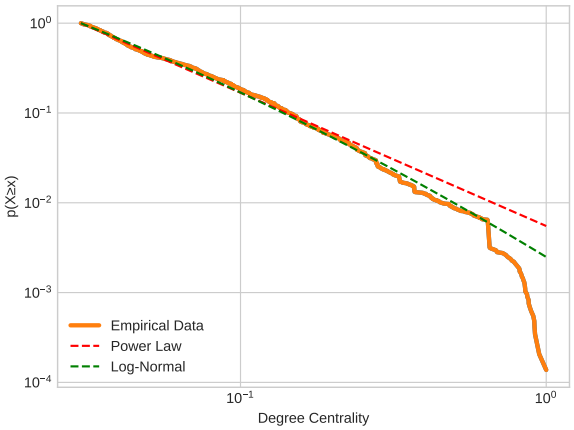


Figure 4.4: Cumulative degree centrality distribution in comparison to a power law and log-normal distribution.

4.3.2 Volatility Modelling

We present the results along with the diagnostics in Tables 4.3. Starting with the results for the return model given by Equation (4.5), the estimated coefficients on the autoregressive terms indicate that daily returns exhibit negative and statistical significant autocorrelation. For the surprise index, the significant estimates corresponding to the interval [-1,0] suggest that there is a pre-report effect for events published via Twitter posts. The coefficients during the first day following the event realization remain significant while the coefficients become non-significant for the successive interval. Positive (negative) surprise index first differences have a negative (positive) effect on returns. The impact of the [0,1] interval is larger compared to the effect during the subsequent interval. The observed decreasing impact of surprising events on returns is in line with the results presented in (Malkiel 2003) as pricing readjusts to new relevant information quickly.

In table 4.3 the coefficients of the volatility model for the autoregressive terms are mostly significant except for the third lag. The surprise index coefficients show that an increasing level of surprise leads to an increased volatility. Similarly to the results of the return model, we observe a pre-report effect which postulates that the volatility rises before the event is reported on Twitter. The coefficients for the interval greater than 1 day are not statistical significant, claiming that the effect of surprise level differences wear off within the first time interval.

Estimation results for the return model given by Equation (4.5) and the volatility model given by Equation (4.6)

Variable	Estimate	
	Return Model	Volatility Model
Intercept	0.000	0.003 **
R_{t-1}	−0.028 ***	
R_{t-2}	−0.019 ***	
R_{t-3}	−0.011 **	
$\lvert\hat{\epsilon}_{t-1}\rvert$		1.021 ***
$\lvert\hat{\epsilon}_{t-2}\rvert$		−0.037 ***
$\lvert\hat{\epsilon}_{t-3}\rvert$		0.012
$\lvert\hat{\epsilon}_{t-4}\rvert$		−0.028 ***
u_{t-1}		0.874 ***
Surprise Index		
$S_+[-1,0]$	−0.004 **	0.003
$S_-[-1,0]$	0.003 **	0.026 **
$S_+[0,1]$	−0.003 **	0.010 *
$S_-[0,1]$	0.002 *	−0.013 *
$S_+[1,2]$	−0.031	−0.019
$S_-[1,2]$	0.024	−0.012
Joint F-tests		
$\forall\gamma_{l,j} = 0$	43.445 ***	
$\forall\beta_{l,j} = 0$	3.211 **	
$\forall\beta_{l,j} = \forall\gamma_{l,j} = 0$	41.587	
Likelihood ratio tests		
$\forall\gamma'_{l,j} = 0$		3018 ***
$\forall\delta'_1 = 0$		745 ***
$\forall\beta'_{l,j} = 0$		1473 ***

Table 4.3: The estimates for the return model (Equation 4.5) are obtained using WLS. The estimates for the volatility model (Equation 4.6) are obtained using maximum likelihood estimation. The Null hypothesis of the F-test and of the Likelihood ratio test is that the estimated parameters are jointly zero. *,**,*** indicates the statistical significance level 10%, 5%, and 1% respectively.

4.3.3 Volatility Forecasting

We additionally evaluate the practical value for financial forecasting by including the surprise metric in a classical GARCH(1,1) model. Therefore, the volatility of a return series is modelled as an autoregressive process:

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma s_{t-1} \tag{4.7}$$

We add to the initial model 4.7 an additional lagged regressor s_{t-1} to investigate if the aggregated surprise value provides an explanatory value to the volatility. We fit the volatility model using the first 250 observations and forecast 1-step ahead. We re-estimate the model parameters every ten observations to capture the impact of structural data changes. The out-of-sample consists of 350 observations, while the in-sample of 250. Table 4.4 reports the results for the baseline GARCH model as well as

Estimation results for the GARCH model given by Equation (4.7)

Variable	Estimate	
	GARCH(1, 1)	GARCH(1, 1) + S
ω	0.076 **	0.094
α	0.201 **	0.205
β	0.699 ***	0.695
γ		0.065*
Diagnostics		
LogLikelihood	−613.713	−616.393
AIC	2.402	2.400
$Q_{r^2}(5)$	18.885	18.671
JB	173.443	138.662
Forecasting		
MSE	0.463	0.457
MAE	0.464	0.456

Table 4.4: Regression results for the first 250 observations; 01/01/2015 - 31/12/2015. T-values are in parentheses. AIC denotes the Akaike information criterion. $Q_{r^2}(5)$ denotes the Ljung-Box statistics for the first five lags of squared residuals. JB is the value of the Jarque and Bera Normality test statistics (p-values are reported within the brackets) on standardized residuals. MSE denotes mean squared error. MAE denotes mean absolute error.

including the proposed surprise index. Estimating a GARCH model with the proposed surprise index shows that there exist a significant linear relationship between volatility and surprise in the Twitter network. The positive sign of γ implies a positive correlation between volatility and surprise, as expected. The Ljung-Box test strongly rejects serial autocorrelation among the standardized squared residuals for both model implications. The Jarque-Bera normality test rejects the normality of standardized residuals even after controlling for volatility clustering. Judging by the Akaike Information Criterion (AIC) the GARCH model which includes the surprise index would be confirmed to outperform the plain GARCH model regarding goodness-of-fit.

A simple assumption about the effectiveness of the surprise index is that the quantified surprise level does not provide additional information if the stock market volatility can be predicted well by the historical returns only. This implies that if the explanatory power of the baseline model is high, events manifested by the surprise index are already expected and therefore considered in the pricing. In such cases, only the baseline model is sufficient to predict volatility and integration of surprise level does not improve the explanatory power much. On the other hand, if the prediction power of the baseline model is low, the stock markets seem to expect any events that will cause a surprise.

4.4 Conclusion

Based on the review of recent works on analysing social media data with a focus on mining the public opinion concerning publicly traded companies, we provide a framework to handle a vital research problem with satisfactory experimental results via the proposed methodology. More specifically, while mostly the overall sentiment in user-generated content is considered in the previous research, this study proposes a method for quantifying the surprise level of events despite their content or sentiment. The contribution of this study can be summarized as follows. First, we propose a method that quantifies the heterogeneity of the informational content within the Twitter ecosystem. Second, we can measure the level of surprise by taking the diffusion speed of rare or unforeseen events by modelling the information flow in Twitter as a network. Third, we investigate how the S&P 500 returns and volatility react to changes inherent in the daily aggregated surprise index. We find that an increase in the surprise index has an inverse effect on S&P 500 returns. A positive increase of the uprise level has

a positive impact on the daily volatility. Moreover, the results indicate for both the return and volatility models a pre-report effect. Overall, our results support the EMH in that prices adjust to unexpected events, however, the adjustment process may take up to 2 days. Although this study provides novel results, the following limitations need to be noted. The first limitation concerns the choice of our time frame. Although we deemed a two and a half years period sufficient for our study, as the size of the Twitter network still increases exponentially it is yet unknown whether we may consider the observed conditions can as stable. Moreover, endogeneity cannot be completely ruled out in our study as we cannot answer the question of whether market volatility determines the level of Twitter activity or vice versa.

5 Conclusion

In the first part of this thesis, we study the informative value of time-frequency decompositions based on wavelets. In Chapter 2 we address the challenge of forecasting major stock indices. For the binary classification task, we can identify three wavelet filter out of fifteen, namely $D4$, $D6$ and $CF6$, which outperform the naive approach. However, this does not hold for the task of point forecasting task as our approach fails to outperform the naive strategy. We exemplify the performance of the forecasting algorithm by simulating a trading application in which the wavelet-based approach outperforms the naive strategy by a factor of about two in terms of Sharpe ratio.
In Chapter 3 we study the transmission processes between the European stock markets and the nominal European exchange rate. To effectively distinguish between contagion and interdependence we estimate the spectral characteristics of the time series as a function of time and frequency using a continuous wavelet transform. This approach also allows us to reveal linkages and co-movement patterns, including lead-lag relationships, between periodic components. The empirical results support both cyclical and anti-cyclical relationships between the series while existing casual and reverse casual relation vary across scale and time. We strongly emphasize the limitations of a bivariate approach and encourage the conduction of a multivariate analysis to revisit our findings.
In the second part of the thesis, namely Chapter 4, we aim to quantify the level of surprise in events published on Twitter. The proposed methodology maps the informational content of every Twitter account belonging to at least one S&P500 constituent into a network. This approach allows to observe the formation of events and to quantify its propagation on an aggregate level. In a regression analysis, we investigate whether the proposed surprise metric has any explanatory power for the S&P500 logarithmic returns and volatility. We find that an increase in the surprise index has an inverse effect on daily returns and a positive impact on the daily volatility. Moreover, the results support the hypothesis of a pre-report effect for both the return and volatility models.

Future work might also explore the limits of a multivariate approach to analyze directional interdependence and contagion patterns at different time scales. Considering deep neural network embeddings to reduce the dimensionality of frequency decompositions is one of the next steps in this research.

A Extended plots for the forecasting performance of the WL-RFE-SVR

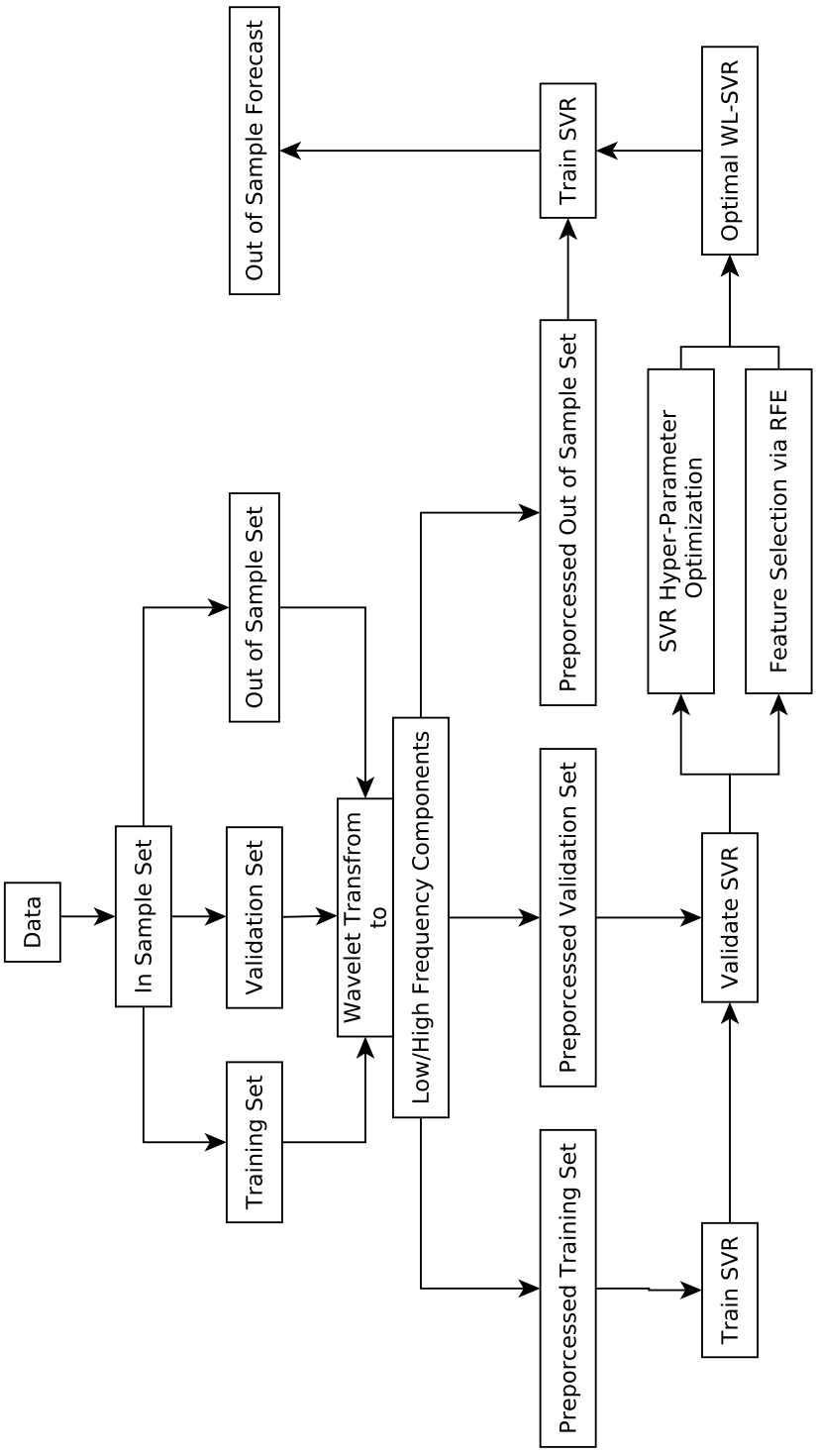


Figure A.1: Flowchart of the proposed WL-RFE-SVR algorithm.

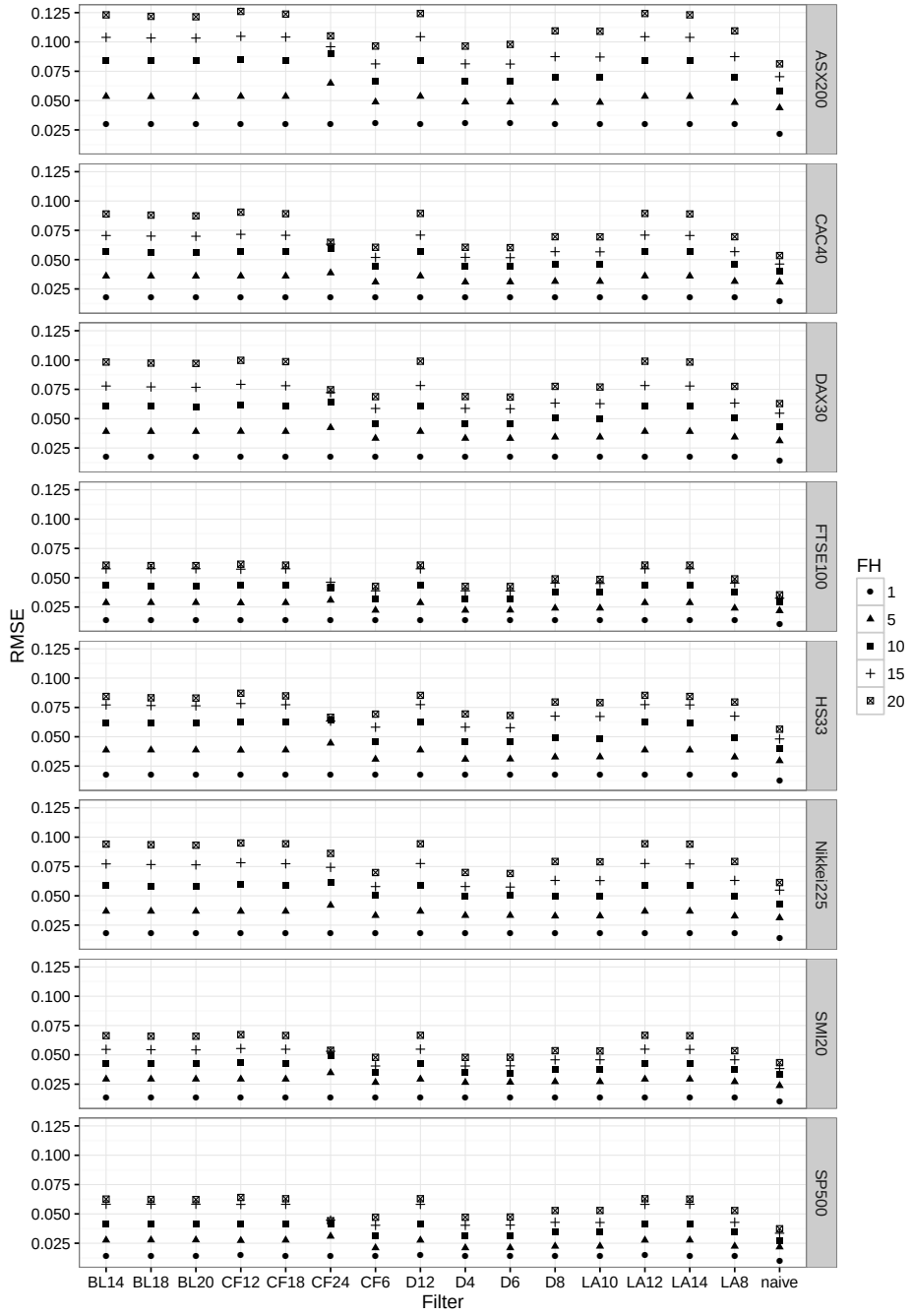


Figure A.2: Point forecast performance of each wavelet filter including the naive method based on root mean squared error (RMSE), defined by Equation (2.41), for 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method the time series has not been preprocessed by a wavelet filter. As a result the input vector consists of 20 lagged returns. The considered stock indices are from Germany (DAX30), United Kingdom (FTSE100) ,US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

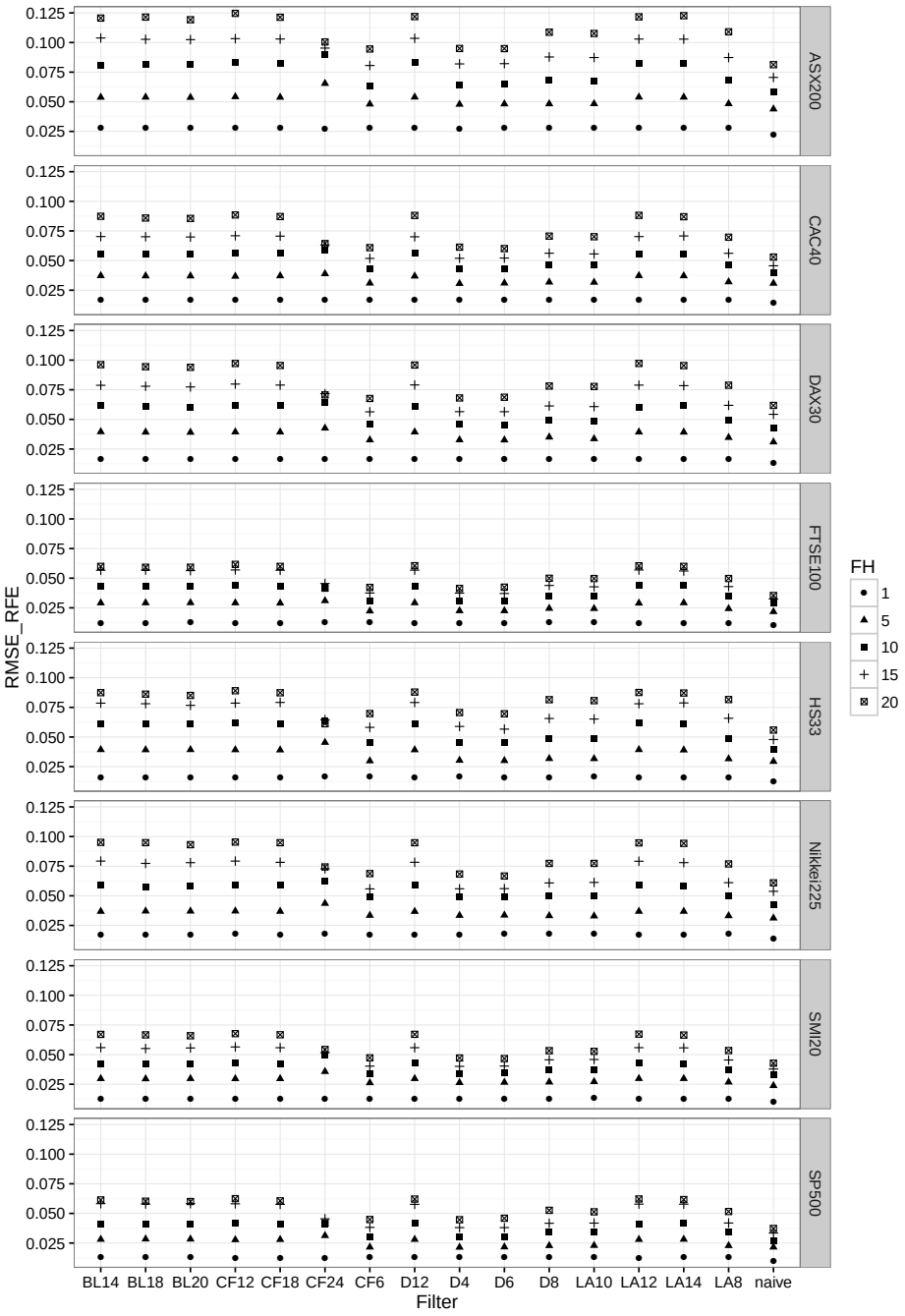


Figure A.3: Point forecast performance of each wavelet filter including the naive method based on the root mean squared error (RMSE), defined by Equation (2.41), for 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The feature set has been selected based on the results from the Recursive Filter Elimination algorithm, which is described in Section 2.2.2. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymmetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method, the time series has not been preprocessed by a wavelet filter. As a result, the input vector consists of 20 lagged returns. The considered stock indices are from Germany (DAX30), United Kingdom (FTSE100), US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

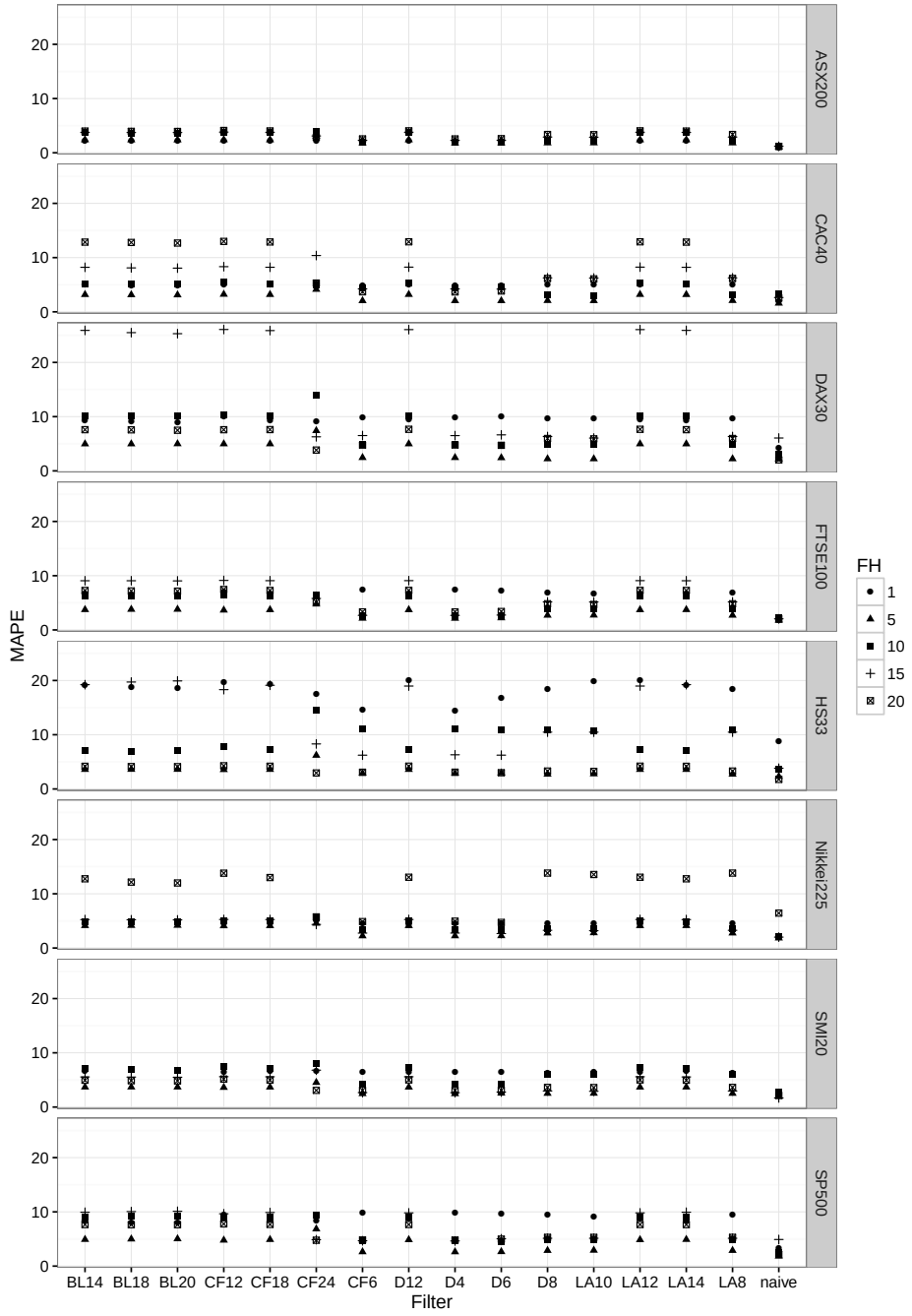


Figure A.4: Point forecast performance of each wavelet filter including the naive method based on mean absolute percentage error (MAPE), defined by Equation (2.42), for 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymmetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method, the time series has not been preprocessed by a wavelet filter. As a result, the input vector consists of 20 lagged returns. The considered stock indices are from Germany (DAX30), United Kingdom (FTSE100), US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

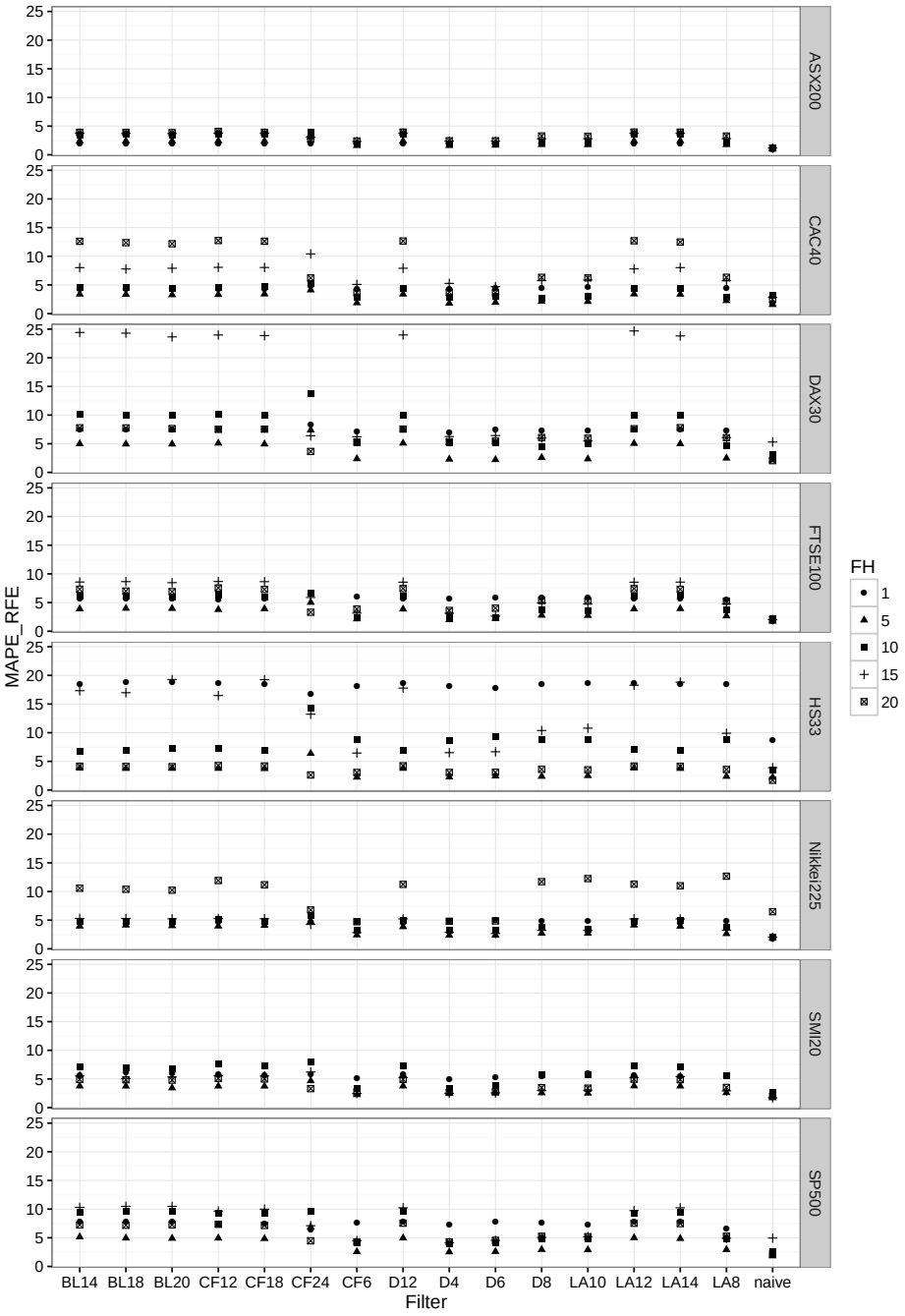


Figure A.5: Point forecast performance of each wavelet filter including the naive method based on mean absolute percentage error (MAPE), defined by Equation (2.42), for 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The feature set has been selected based on the results from the Recursive Filter Elimination algorithm, which is described in Section 2.2.2. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymmetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method, the time series has not been preprocessed by a wavelet filter. As a result, the input vector consists of 20 lagged returns. The considered stock indices are from Germany (DAX30), United Kingdom (FTSE100), US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

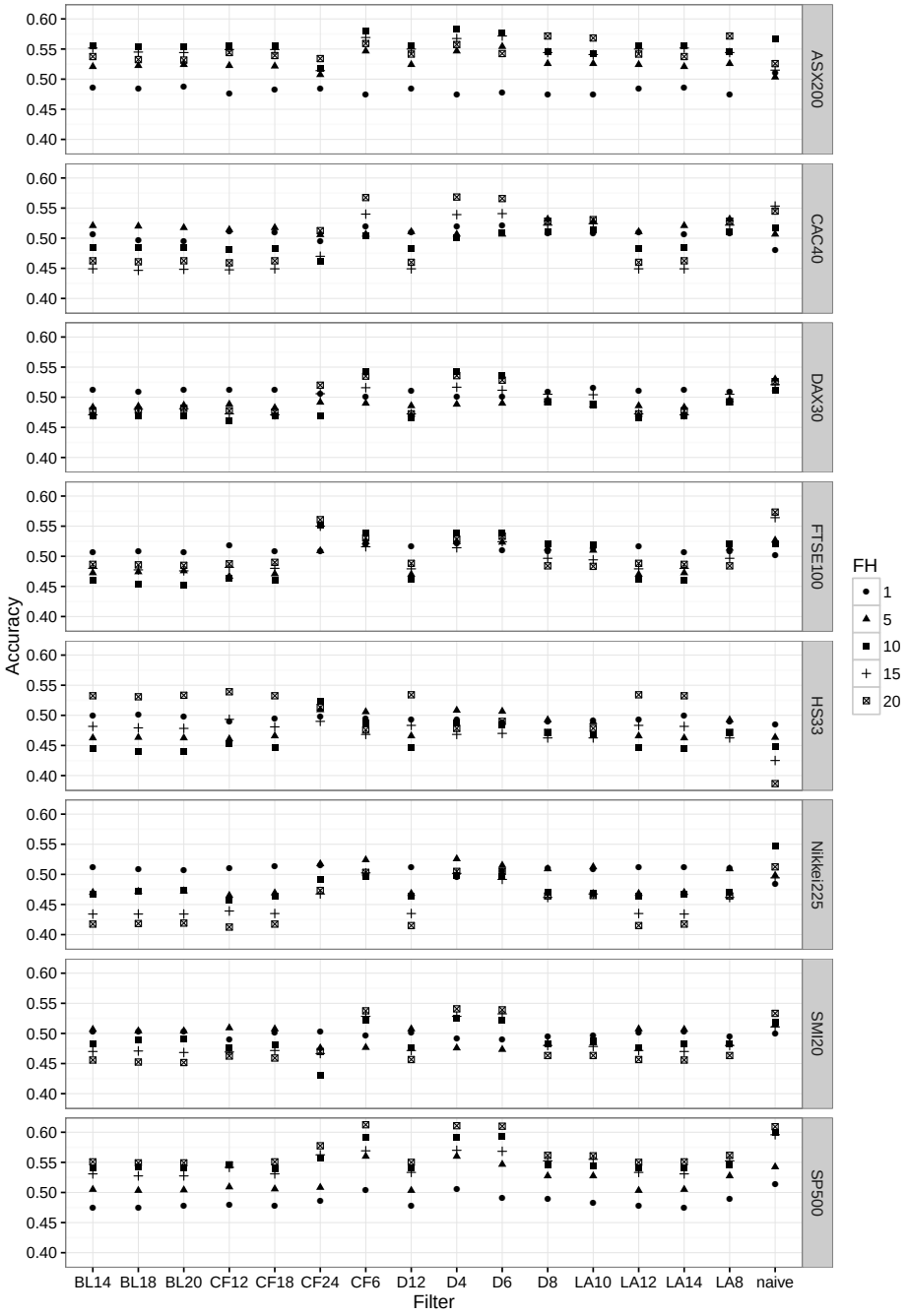


Figure A.6: Classification forecast performance of each wavelet filter including the naive method based on Accuracy, defined by Equation (2.43), for 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method the time series has not been preprocessed by a wavelet filter. As a result the input vector consists of 20 lagged returns. The considered stock indices are from Germany (DAX30), United Kingdom (FTSE100) ,US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

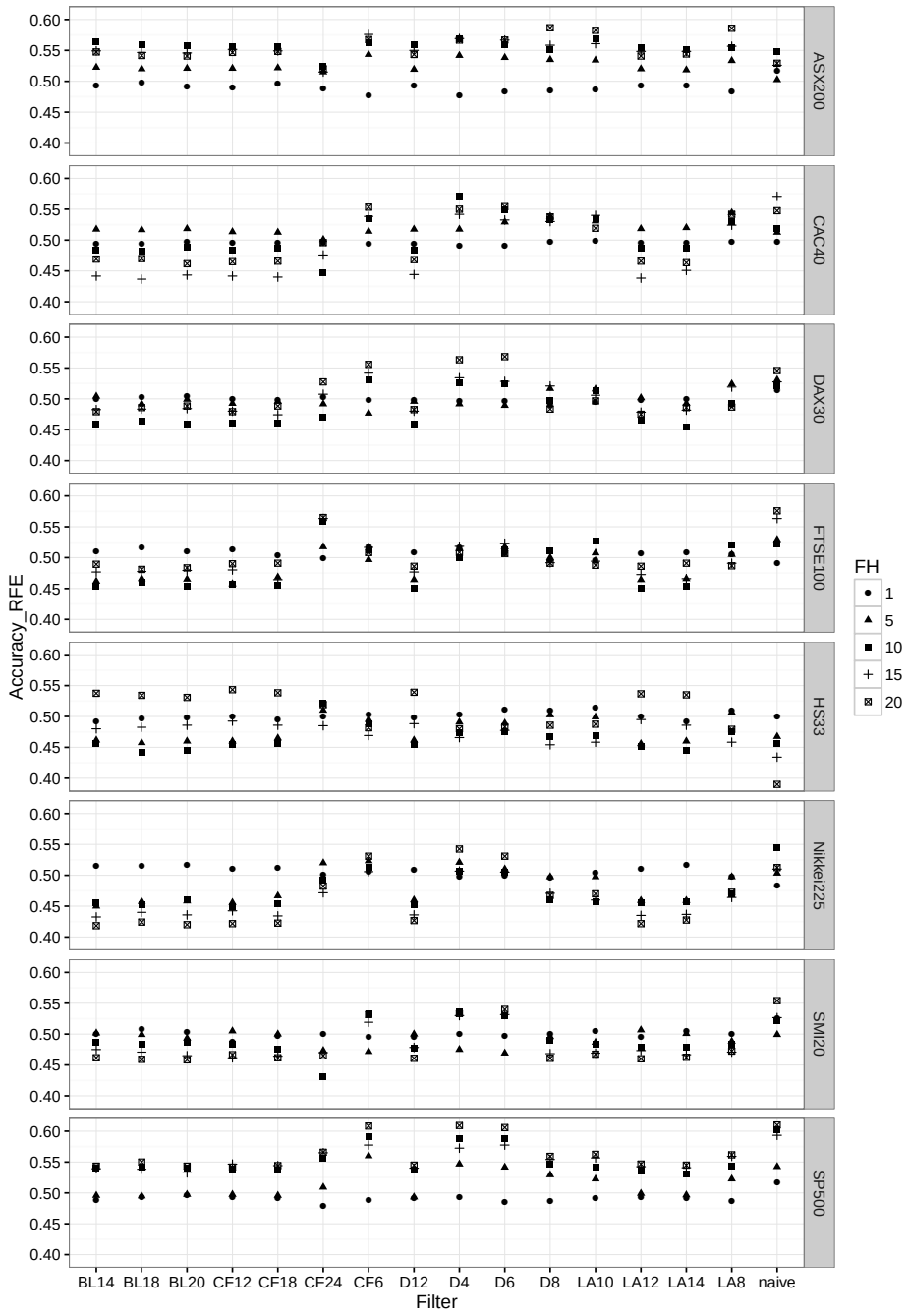


Figure A.7: Classification forecast performance of each wavelet filter including the naive method based on Accuracy, defined by Equation (2.43), for 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The feature set has been selected based on the results from the Recursive Filter Elimination algorithm, which is described in Section 2.2.2. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method the time series has not been preprocessed by a wavelet filter. As a result the input vector consists of 20 lagged returns. The considered stock indices are from Germany (DAX30), United Kingdom (FTSE100) ,US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

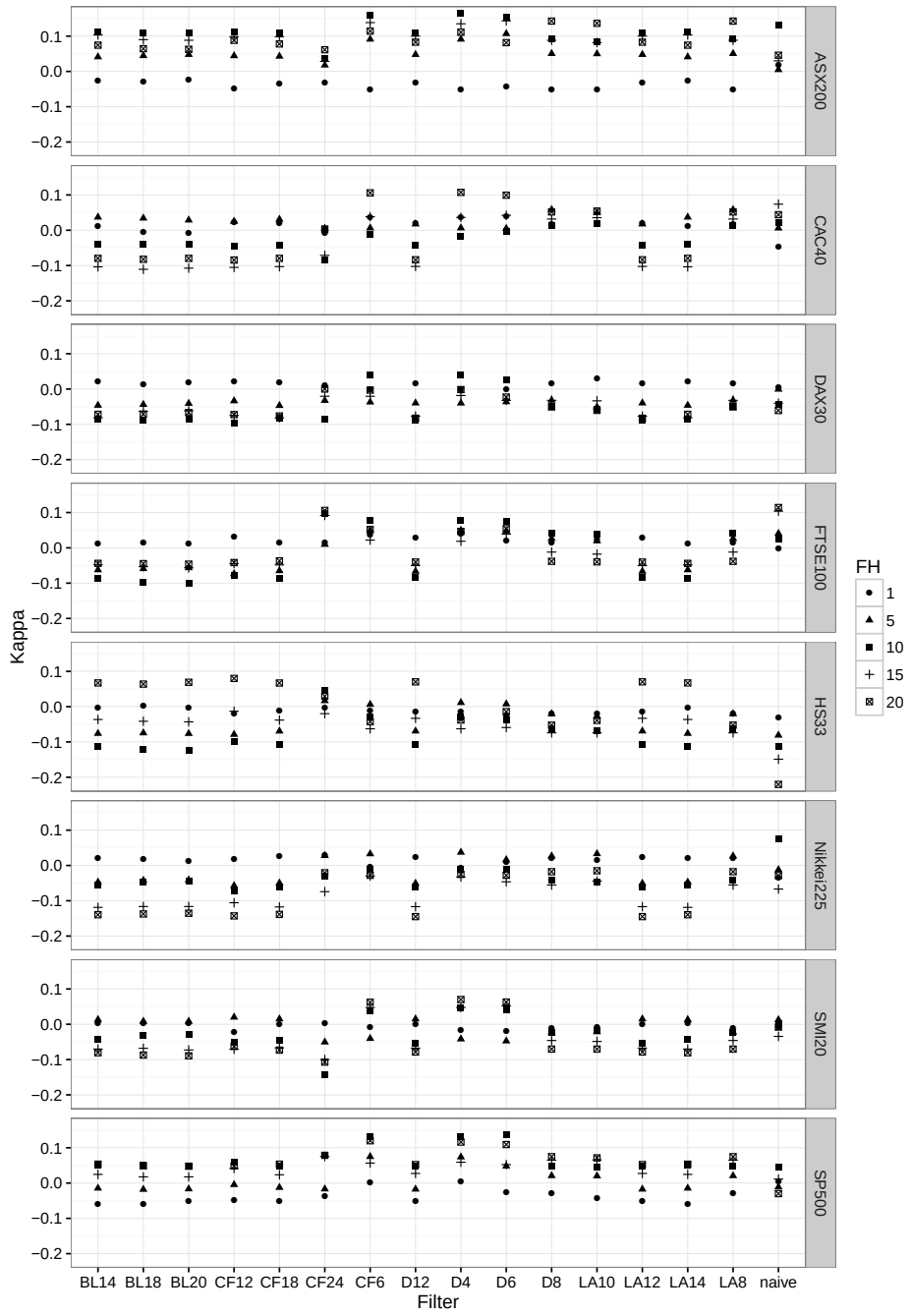


Figure A.8: Classification forecast performance of each wavelet filter including the naive method based on Kappa, defined by Equation (2.44), for 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method the time series has not been preprocessed by a wavelet filter. As a result the input vector consists of 20 lagged returns. The considered stock indices are from Germany (DAX30), United Kingdom (FTSE100) ,US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

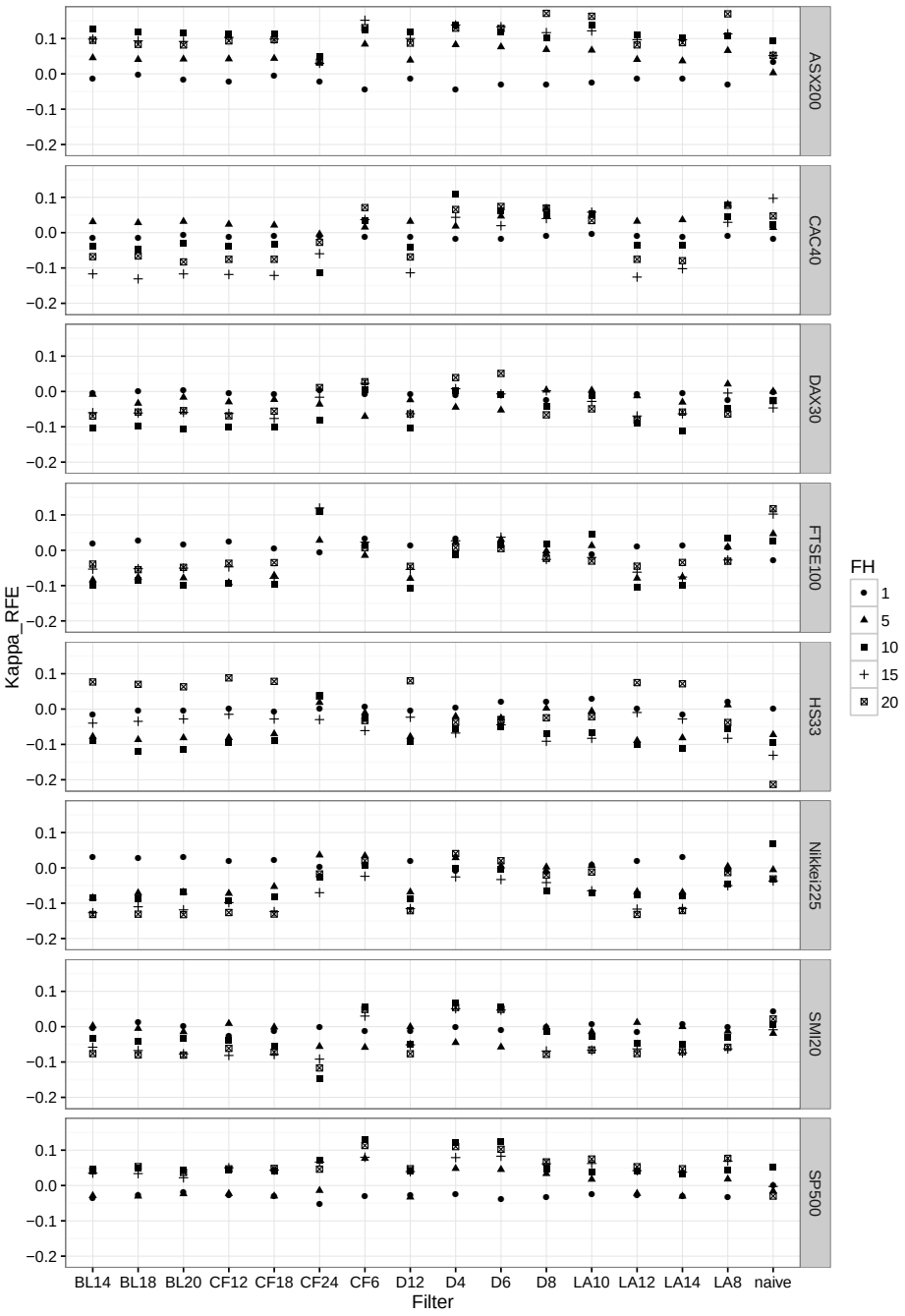


Figure A.9: Classification forecast performance of each wavelet filter including the naive method based on Kappa, defined by Equation (2.44), for 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The feature set has been selected based on the results from the Recursive Filter Elimination algorithm, which is described in Section 2.2.2. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method the time series has not been preprocessed by a wavelet filter. As a result the input vector consists of 20 lagged returns. The considered stock indices are from Germany (DAX30), United Kingdom (FTSE100) ,US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

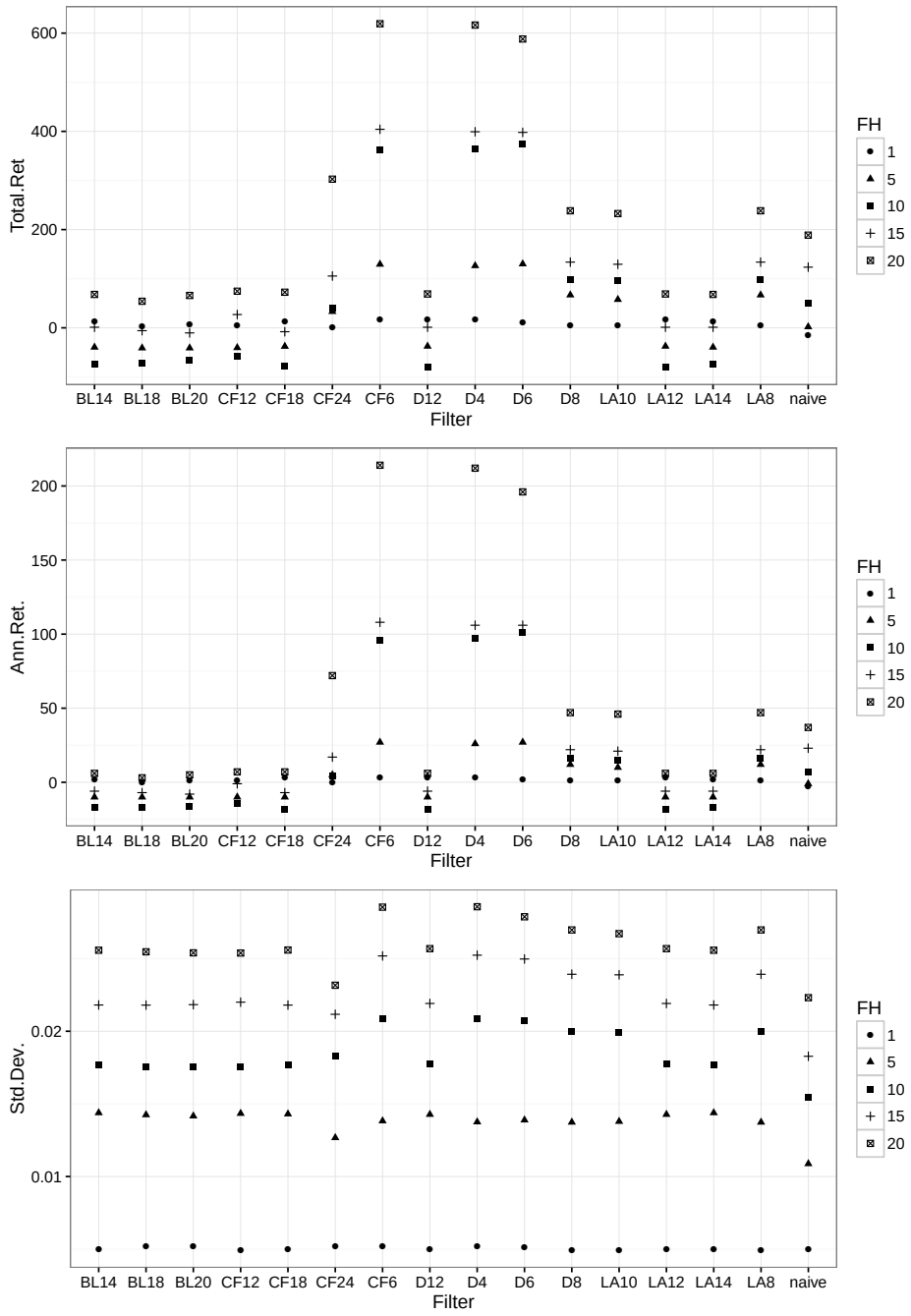


Figure A.10: Trading performance in terms of Total Return (Equation (2.45)), annual return (Equation (2.46)) and standard deviation (Equation (2.48) for the equally weighted portfolio consisting of the major stock indices from Germany (DAX30), United Kingdom (FTSE100) ,US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). The results are based on 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method the time series has not been preprocessed by a wavelet filter. As a result the input vector consists of 20 lagged returns. The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

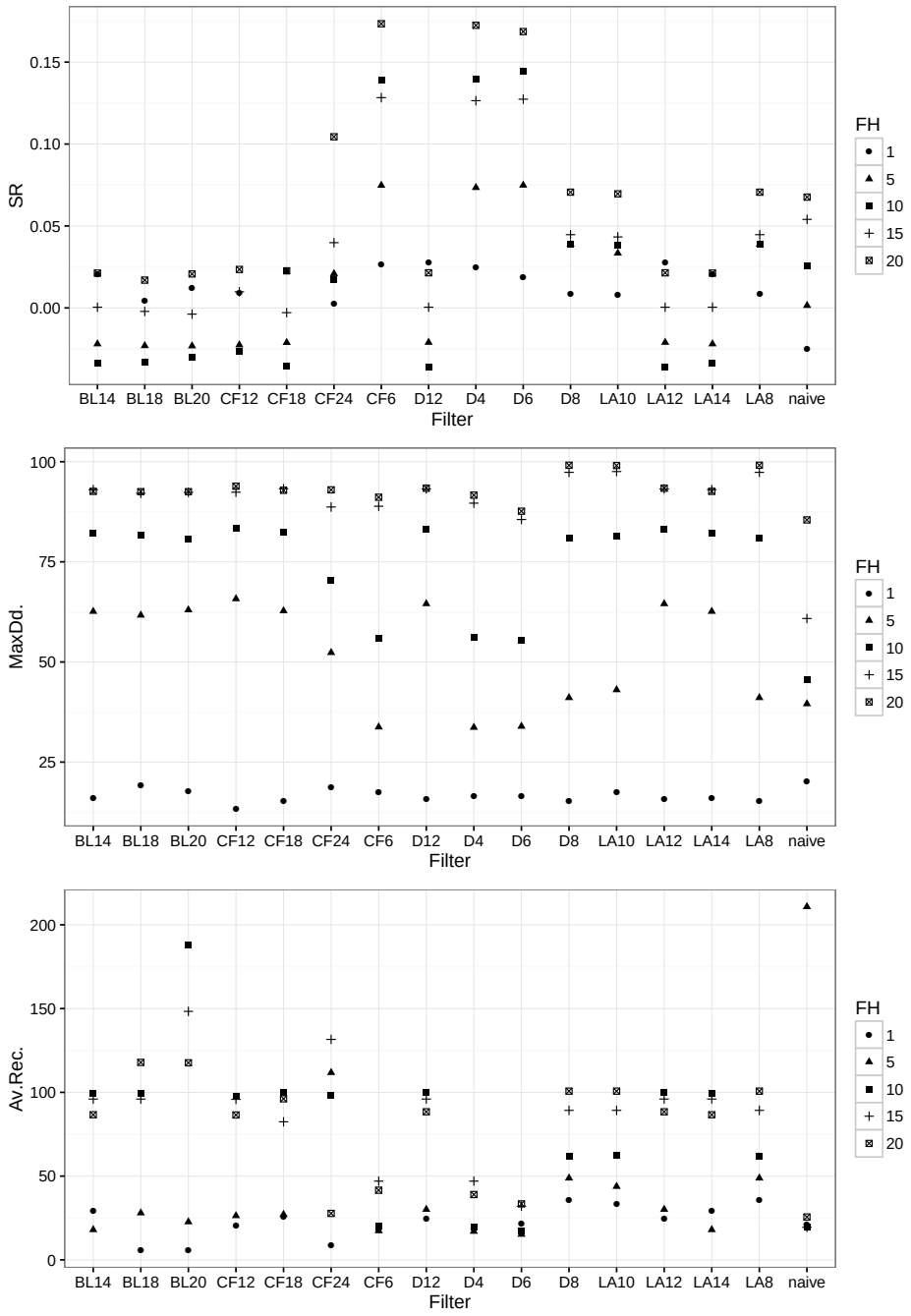


Figure A.11: Trading performance in terms of Sharpe Ratio (Equation (2.47)), maximum drawdown (Equation (2.49)) and average recovery for the equally weighted portfolio consisting of the major stock indices from Germany (DAX30), United Kingdom (FTSE100) ,US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). Average Recovery is defined as the mean duration of an equity drawdown period. The results are based on 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymmetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method, the time series has not been preprocessed by a wavelet filter. As a result, the input vector consists of 20 lagged returns. The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

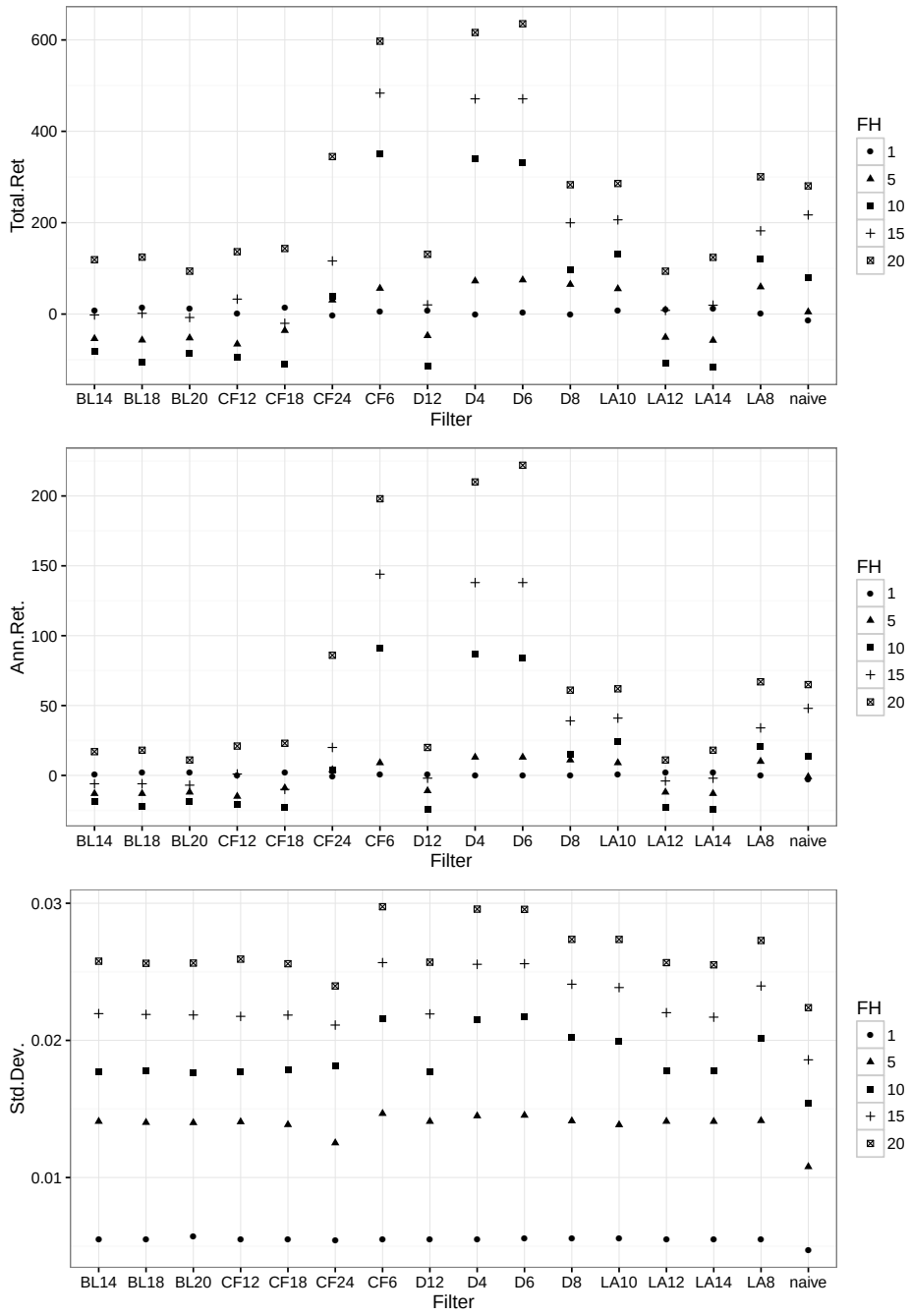


Figure A.12: Trading performance in terms of total return (Equation (2.45)), annual return (Equation (2.46)) and standard deviation (Equation 2.48) for the equally weighted portfolio consisting of the major stock indices from Germany (DAX30), United Kingdom (FTSE100) ,US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). The results are based on 1200 out-of-sample forecasts for the period from 29.20.2010 to 26.06.2015. The feature set has been selected based on the results from the Recursive Filter Elimination algorithm, which is described in Section 2.2.2. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method the time series has not been preprocessed by a wavelet filter. As a result the input vector consists of 20 lagged returns. The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

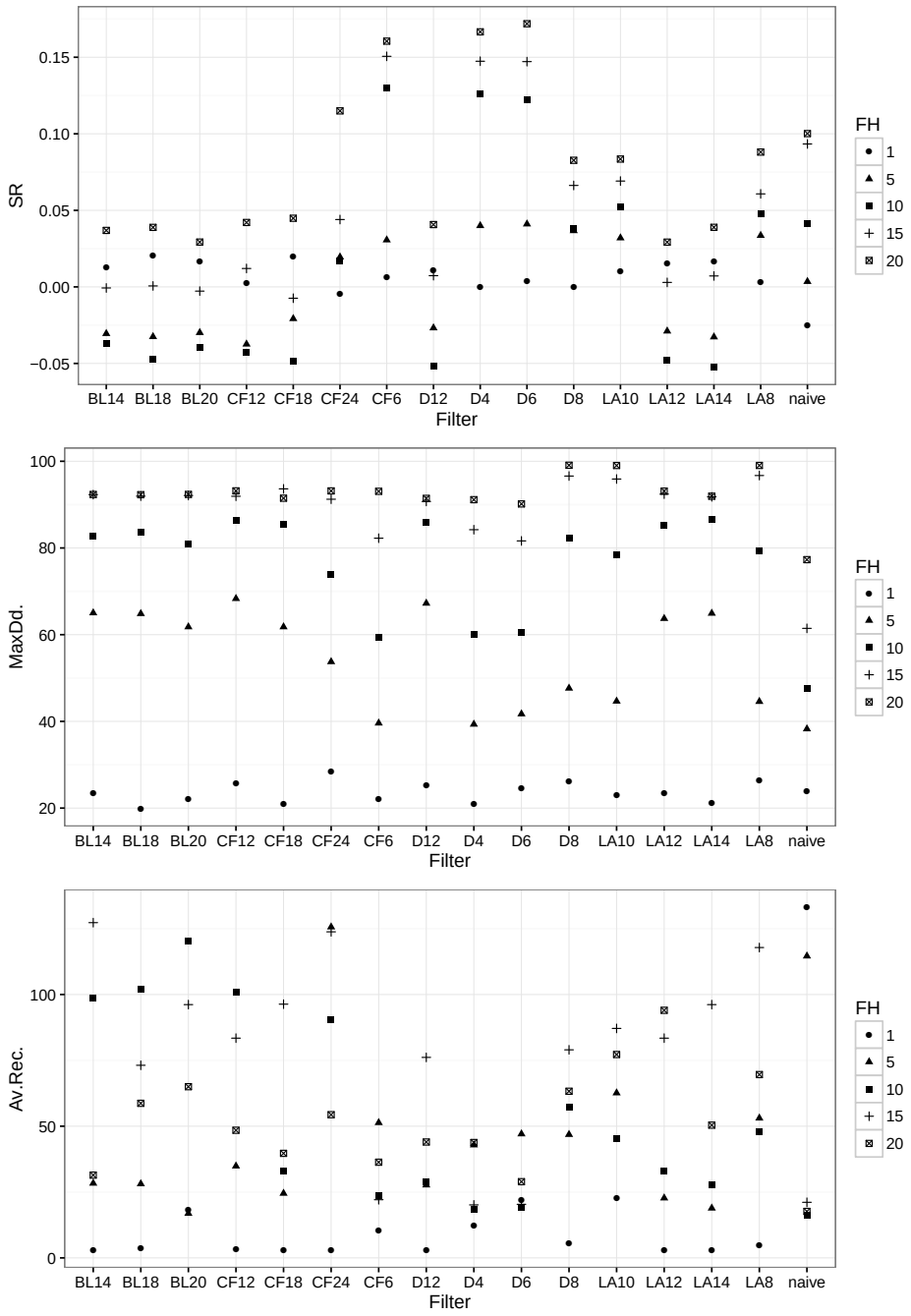


Figure A.13: Trading performance in terms of Sharpe Ratio (Equation (2.47)), maximum drawdown (Equation (2.49)) and average recovery for the equally weighted portfolio consisting of the major stock indices from Germany (DAX30), United Kingdom (FTSE100) ,US (SP500), France (CAC40), Switzerland (SMI20), Australia (ASX200), Japan (Nikkei225) and Hong Kong(HS33). Average Recovery is defined as the mean duration of an equity drawdown period. The feature set has been selected based on the results from the Recursive Filter Elimination algorithm, which is described in Section 2.2.2. The wavelet filters are denoted as Daubechies (D) (4,6,8,12), Least Asymmetric (LA) (8,10,12,14), Best Localized (BL) (14,18,20) and Coiflet (CF) (6, 12, 18, 24), where the integers in brackets indicate the wavelet filter length. For the naive method, the time series has not been preprocessed by a wavelet filter. As a result, the input vector consists of 20 lagged returns. The forecasting horizon is denoted as FH and varies between +1, +5, +10, +15 and +20 days.

Bibliography

Abdalla, Issam SA, and Victor Murinde. 1997. "Exchange rate and stock price interactions in emerging financial markets: evidence on India, Korea, Pakistan and the Philippines". *Applied financial economics* 7 (1): 25–35.

Aguiar-Conraria, L., and M.J. Soares. 2011. "The Continuous Wavelet Transform: A Primer". *NIPE Working Paper* 16 (16): 1–43.

Aguiar-Conraria, Luís, and Maria Joana Soares. 2014. "The continuous wavelet transform: moving beyond uni-and bivariate analysis". *Journal of Economic Surveys* 28 (2): 344–375.

Aiello, Luca Maria, et al. 2013. "Sensing trending topics in twitter". *IEEE Transactions on Multimedia* 15, no. 6 (): 1268–1282.

Ajayi, Richard A., Joseph Friedman, and Seyed M. Mehdian. 1998. "On the relationship between stock returns and exchange rates: Tests of granger causality". *Global Finance Journal* 9 (2): 241–251.

Ajayi, Richard A, and Mbodja Mougoué. 1996. "On the dynamic relation between stock prices and exchange rates". *Journal of Financial Research* 19 (2): 193–207.

Ambroise, C., and G. J. McLachlan. 2002. "Selection bias in gene extraction on the basis of microarray gene-expression data". *Proceedings of the National Academy of Sciences* 99 (10): 6562–6566.

Andersen, Torben G., and Tim Bollerslev. 1998. "Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies". *Journal of Finance* 53 (1): 219–265.

Andersen, Torben G, et al. 2003. "Micro effects of macro announcements: Real-time price discovery in foreign exchange". *American Economic Review* 93 (1): 38–62.

Atsalakis, George, and Kimon Valavanis. 2009. "Surveying stock market forecasting techniques – Part II : Soft computing methods". *Expert Systems With Applications* 36 (3): 5932–5941.

Baele, Lieven, and Koen Inghelbrecht. 2010. "Time-varying integration, interdependence and contagion". *Journal of International Money and Finance* 29 (5): 791–818.

Bernard, Victor L., and Jacob K. Thomas. 1989. "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?" *Journal of Accounting Research* 27:1.

Billio, M, and M Caporin. 2006. "Market Linkages , Variance Spillover and Correlation Stability : Empirical Evidences of Financial Contagion". *Computational Statistics & Data Analysis* 54 (March): 1–19.

Bodart, Vincent, and Bertrand Candelon. 2009. "Evidence of interdependence and contagion using a frequency domain framework". *Emerging Markets Review* 10 (2): 140–150.

Bron, Coen, and Joep Kerbosch. 1973. "Algorithm 457: finding all cliques of an undirected graph". *Communications of the ACM* 16 (9): 575–577.

Brown, Lawrence D. 2003. "Small negative surprises: Frequency and consequence". *International Journal of Forecasting* 19 (1): 149–159.

Caporale, Guglielmo Maria, Andrea Cipollini, and Nicola Spagnolo. 2005. "Testing for contagion: A conditional correlation analysis". *Journal of Empirical Finance* 12 (3): 476–489.

Cha, Meeyoung, et al. 2010. "Measuring User Influence in Twitter : The Million Follower Fallacy". *International AAAI Conference on Weblogs and Social Media* 10 (10-17): 10–17.

Chang, Pei Chann, and Chin Yuan Fan. 2008. "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): 802–815.

Chang, Shing I., and Srikanth Yadama. 2010. "Statistical process control for monitoring non-linear profiles using wavelet filtering and B-Spline approximation". *International Journal of Production Research* 48 (4): 1049–1068.

Cherkassky, Vladimir, and Yunqian Ma. 2004. “Practical selection of SVM parameters and noise estimation for SVM regression”. *Neural Networks* 17 (1): 113–126.

Chernick, Michael R. 2001. *Wavelet Methods for Time Series Analysis*, 43:491–491. 4. Cambridge University Press.

Corsetti, C., M. Pericoli, and M. Sbracia. 2005. “’Some contagion, some interdependence’:more pittfalls in tests of financial contagion”. *Journal of International Money and Finance* 24 (8): 1177–1199.

Dacorogna, Michel M, et al. 2001. *An Introduction to High-Frequency Finance*. Academic press.

Daubechies, Ingrid. 1992. *Ten lectures on wavelets*. Vol. 61. Siam.

DeLurgio, Stephen A. 1998. *Forecasting principles and applications*. 802. Boston: Irwin/McGraw-Hill.

Dominguez, Kathryn M, and Jeffrey A Frankel. 1993. “Does Foreign-Exchange Intervention Matter? The Portfolio Effect”. *American Economic Review* 83 (5): 1356.

Dornbusch, Rudiger, and Stanley Fischer. 1980. “Exchange Rates and the Current Account”. *American Economic Review* 70 (5): 960–971.

Doroslovacki, Milos I. 1998. “On the least asymmetric wavelets”. *IEEE Transactions on Signal Processing* 46 (4): 1125–1130.

Fama, Eugene F. 1997. “Market Efficiency, Long-Term Returns, and Behavioral Finance”. *Journal of Financial Economics* 49 (3): 283–306.

Forbes, Kristin, and Roberto Rigobon. 2002. “No Contagion, Only Independence: Measuring Stock Market Comovements”. *The Journal of Finance* 57 (5): 2223–2261.

Foufoula-Georgiou, Efi, and Praveen Kumar. 1994. *Wavelets in Geophysics, Volume 4*. 4:373. Academic Press.

Franck, Peter, and Allan Young. 1972. “Stock Price Reaction of Multinational Firms to Exchange Realignments”. *Financial Management* 1 (3): 66.

Funashima, Yoshito. 2017. “Time-varying leads and lags across frequencies using a continuous wavelet transform approach”. *Economic Modelling* 60:24–28.

Gebka, Bartosz, and Michail Karoglou. 2013. "Is there life in the old dogs yet? Making break-tests work on financial contagion". *Review of Quantitative Finance and Accounting* 40 (3): 485–507.

Gilbert, Eric, and Karrie Karahalios. 2010. "Widespread Worry and the Stock Market". *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*: 58–65.

Gonghui, Zheng, et al. 1999. "Multiscale transforms for filtering financial data streams". *Journal of Computational Intelligence in Finance* 7 (18-35): 18–35.

Granger, Clive W.J, Bwo-Nung Huangb, and Chin-Wei Yang. 2000. "A bivariate causality between stock prices and exchange rates: evidence from recent Asianflu". *The Quarterly Review of Economics and Finance* 40 (3): 337–354.

Grinsted, A., J. C. Moore, and S. Jevrejeva. 2004. "Application of the cross wavelet transform and wavelet coherence to geophysical time series". *Nonlinear Processes in Geophysics* 11 (5/6): 561–566.

Han, Young Wook. 2010. "The effects of US macroeconomic surprises on the intraday movements of foreign exchange rates: Cases of USD-EUR and USD-JPY exchange rates". *International Economic Journal* 24 (3): 375–396.

Heisenberg, W. 1927. "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik". *Zeitschrift für Physik* 43 (3): 172–198.

Huang, Shian Chang, and Tung Kuang Wu. 2008. "Combining wavelet-based feature extractions with relevance vector machines for stock index forecasting". *Expert Systems* 25 (2): 133–149.

Hudgins, Lonnie, Carl A Friehe, and Meinhard E Mayer. 1993. "Wavelet transforms and atmopsheric turbulence". *Physical Review Letters* 71 (20): 3279.

Ince, Huseyin, and Theodore B Trafalis. 2008. "Short term forecasting with support vector machines and application to stock price prediction". *International Journal of General Systems* 37 (6): 677–687.

John, George H., Ron Kohavi, and Karl Pfleger. 1994. "Irrelevant Features and the Subset Selection Problem". In *Machine Learning Proceedings 1994*, 121–129.

Kallberg, Jarl, and Paolo Pasquariello. 2008. “Time-series and cross-sectional excess comovement in stock indexes”. *Journal of Empirical Finance* 15 (3): 481–502.

Kao, Ling Jing, et al. 2013. “A hybrid approach by integrating wavelet-based feature extraction with MARS and SVR for stock index forecasting”. *Decision Support Systems* 54 (3): 1228–1244.

Karoui, Aymen. 2006. “The Correlation between Fx Rate Volatility and Stock Exchange Returns Volatility: An Emerging Markets Overview”. *SSRN*.

Kim, Kyoung-jae, and Ingoo Han. 2000. “Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index”. *Expert Systems with Applications* 19 (2): 125–132.

Li, Tao, et al. 2002. “A survey on wavelet applications in data mining”. *ACM SIGKDD Explorations Newsletter* 4 (2): 49–68.

Liew, Jim Kyung-Soo, Shenghan Guo, and Tongli Zhang. 2016. “Tweet Sentiments and Crowd-Sourced EarningsEstimates as Valuable Sources of Information around Earnings Releases”. *The Journal of Alternative Investments* 19 (3): 7–26.

Liew, Jim Kyung-Soo, and Garrett Zhengyuan Wang. 2016. “Twitter Sentiment and IPO Performance: A Cross-Sectional Examination”. *The Journal of Portfolio Management* 42 (4): 129–135.

Liu, Xiaohua, et al. 2013. “Named entity recognition for tweets”. *ACM Transactions on Intelligent Systems and Technology* 4 (1): 1–15.

Liu, Yifan, et al. 2017. “Stock Volatility Prediction Using Recurrent Neural Networks with Sentiment Analysis”. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 192–201.

Lu, Chi-Jie, Tian-Shyug Lee, and Chih-Chou Chiu. 2009. “Financial time series forecasting using independent component analysis and support vector regression”. *Decision Support Systems* 47 (2): 115–125.

Ma, Christopher K., and Wenchi G. Kao. 1990. “On exchange rate changes and stock price reactions”. *Journal of Business Finance & Accounting* 17 (November 1986): 441–450.

Malkiel, Burton G. 2003. "The Efficient Market Hypothesis and Its Critics". *Journal of Economic Perspective* 17 (1): 59–82.

Maraun, D., and J. Kurths. 2004. "Cross wavelet analysis: significance testing and pitfalls". *Nonlinear Processes in Geophysics* 11 (4): 505–514.

Mattera, Davide, and Simon Haykin. 1999. "Support Vector Machines for Dynamic Reconstruction of a Chaotic System". In *Advances in Kernel Methods: Support Vector Learning*, 209–241. MIT Press.

Mcclelland, Robert, and Mary F. Kokoski. 1994. "Econometric issues in the analysis of charitable giving". *Public Finance Review* 22 (4): 498–517.

Murinde, Victor, and Sunil S. Poshakwale. 2004. "Exchange Rate and Stock Price Interactions in European Emerging Financial Markets Before and after the Euro". In *SSRN*.

Oppenheim, Alan V. 1999. *Discrete-time signal processing*. Pearson Education India.

Paraskevopoulos, Timotheos, and Peter N Posch. 2016. "A hybrid forecasting algorithm based on SVR and Wavelet Decomposition". *Quantitative Finance and Economics* 2 (QFE-02-03-525): 525.

Paraskevopoulos, Timotheos, and Peter N. Posch. 2017. "Time-Frequency Linkages and Co-Movements between the Euro and European Stock Market: A Continuous Wavelet Analysis".

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. "Glove: Global Vectors for Word Representation". In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Phylaktis, Kate, and Fabiola Ravazzolo. 2005. "Stock prices and exchange rate dynamics". *Journal of International Money and Finance* 24 (7): 1031–1053.

Ramsey, James B. 2002. "Wavelets in economics and finance: Past and future". *Studies in Nonlinear Dynamics & Econometrics* 6 (3).

Ramsey, James, and Camille Lampart. 1997. "The decomposition of economic relationships by time scale using wavelets". *Economic Researh Reports, CV Starr Center For Applied Economics* 3 (1).

Romero, Daniel, et al. 2011. “Influence and Passivity in Social Media Machine Learning and Knowledge Discovery in Databases”. In *Machine Learning and Knowledge Discovery in Databases*, ed. by Dimitrios Gunopulos et al., 6913:18–33. Berlin, Heidelberg: Springer Berlin Heidelberg.

Saeys, Yvan, Iñaki Inza, and Pedro Larrañaga. 2007. “A review of feature selection techniques in bioinformatics”. *Bioinformatics* 23 (19): 2507–2517.

Saiti, Buerhan, Obiyathulla Ismath Bacha, and Mansur Masih. 2016. “Testing the Conventional and Islamic Financial Market Contagion: Evidence from Wavelet Analysis”. *Emerging Markets Finance and Trade* 52 (8): 1832–1849.

Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. 2010. “Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors”. In *Proceedings of the 19th International Conference on World Wide Web*, 851–860. ACM.

Schölkopf, B, and A J Smola. 2002. “Support Vector Machines and Kernel Algorithms”. In *The Handbook of Brain Theory and Neural Networks*, 1119–1125. John Wiley Sons.

Scotti, Chiara. 2016. “Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises”. *Journal of Monetary Economics* 82:1–19.

Seo, Youngmin, et al. 2015. “Daily water level forecasting using wavelet decomposition and artificial intelligence techniques”. *Journal of Hydrology* 520:224–243.

Smith, C. E. 1992. “Stock markets and the exchange rate: A multi-country approach”. *Journal of Macroeconomics* 14 (4): 607–629.

Soenen, Luc A, and Elizabeth S Hennigar. 1988. “An analysis of exchange-rates and stock-prices-the united-states experience between 1980 and 1986”. *Akron Business and Economic Review* 19 (4): 7–16.

Solnik, Bruno. 1987. “Using Financial Prices to Test Exchange Rate Models: A Note”. *The Journal of Finance* 42 (1): 141–149.

Sprenger, Timm O, et al. 2014. “Tweets and trades: The information content of stock microblogs”. *European Financial Management* 20 (5): 926–957.

Stavarek, Daniel. 2005. “Stock Prices and Exchange Rates in the EU and the United States: Evidence on their Mutual Interactions”. *Czech Journal of Economics and Finance (Finance a uver)* 55 (3-4): 141–161.

Steyvers, Mark, and Tom Griffiths. 2010. "Latent dirichlet allocation". *MIS Quarterly* 3 (3): 993–1022.

Svetnik, Vladimir, et al. 2004. "Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules". In *International Workshop on Multiple Classifier Systems*, 334–343. Springer.

Thaler, Richard, and Shlomo Benartzi. 1985. "Does the Stock Market Overreact?" *The Journal of Finance* 40 (3): 793–805.

Tomita, Etsuji, Akira Tanaka, and Haruhisa Takahashi. 2006. "The worst-case time complexity for generating all maximal cliques and computational experiments". *Theoretical Computer Science* 363 (1): 28–42.

Torrence, Christopher, and Gilbert P Compo. 1998. "A practical guide to wavelet analysis". *Bulletin of the American Meteorological society* 79 (1): 61–78.

Vapnik, Vladimir N. 1999. "An Overview of Statistical Learning Theory". *IEEE transactions on neural networks* 10 (5): 988–999.

Vapnik, Vladimir, Steven E. Golowich, and Alex J. Smola. 1997. "Support Vector Method for Function Approximation, Regression Estimation and Signal Processing". In *Advances in Neural Information Processing Systems 9*, ed. by M. C. Mozer, M. I. Jordan, and T. Petsche, 281–287. MIT Press.

Zhang, Bai Ling, et al. 2001. "Multiresolution forecasting for futures trading using wavelet decompositions". *IEEE Transactions on Neural Networks* 12 (4): 765–775.

Zhao, Wayne Xin, et al. 2011. "Comparing twitter and traditional media using topic models". In *European conference on information retrieval*, 338–349. Springer.