공학박사 학위논문

Empirical Research on Financial Investment based on Transfer Entropy and Machine Learning

전이 엔트로피와 기계학습에 기반한 금융투자 실증 연구

2020 년 8 월

서울대학교 대학원

산업공학과

김 선 도

Empirical Research on Financial Investment based on Transfer Entropy and Machine Learning

전이 엔트로피와 기계학습에 기반한 금융투자 실증 연구

지도교수 장 우 진

이 논문을 공학박사 학위논문으로 제출함

2020 년 7 월

서울대학교 대학원

산업공학과

김 선 도

김선도의 공학박사 학위논문을 인준함

2020 년 7 월

위 원 장	조 성 준	(인)
부위원장	장 우 진	(인)
위 원	이 재 욱	(인)
위 원	이 덕 주	(인)
위 원	변 재 현	(인)

Abstract

Empirical Research on Financial Investment based on Transfer Entropy and Machine Learning

Sondo Kim

Department of Industrial Engineering

The Graduate School

Seoul National University

Stock markets have been studied extensively as one of the crucial fields of economy. In particular, research has been actively conducted to analyze and predict the stock market based on relationships among the dynamics of stock prices and returns. In this context, transfer entropy is a non-parametric indicator in analyzing relationships between components of a system, and has a more flexible analytical ability than correlation or Granger-causality. The study of stock price prediction is also being studied from traditional linear models to the latest machine learning models, and research on the optimal asset allocation strategy based on these studies are conducted.

The purpose of this dissertation is to derive ETE based network indicator with a market explanatory power for the US stock market by using effective transfer entropy, which is mainly used in econophysics and information theory. The improvement of the performance of the stock price direction prediction through various machine learning algorithms by ETE based network indicator is also analyzed. Furthermore, we apply the prediction result of the stock price through the machine learning algorithm with ETE based network indicator to optimal portfolio strategy through the Black-Litterman model to study the practical use of the investment strategy.

At first, we explore that the ETE based on 3 and 6 months moving windows can be regarded as the market explanatory variable by analyzing the association between the financial crises and statistical explanatory power among the stocks. We found that 3 and 6 months moving windows ETEs increase in major financial crises, and that the sectors related to the financial crises have a statistical explanatory power to other sectors through the time-varying analysis of the ETE network indicators.

Then, we discover that the prediction performance on the stock price direction can be improved when the ETE driven variable is integrated as a new feature in the logistic regression, multilayer perceptron, random forest, XGBoost, and long short-term memory network. Meanwhile, we suggest utilizing the adjusted accuracy derived from the risk-adjusted return in finance as a prediction performance measure. Notably, we confirm that the multilayer perceptron and long short-term memory network are more suitable for stock price prediction.

Lastly, we examined the possibility for investors to develop an investment strategy that maximizes profits through the Black-Litterman model using ETE and machine learning. The characteristics of the inflow and outflow ETE network indicators with market explanatory power and the stock price direction prediction results using machine learning algorithms are applied to the investor's view of the Black-Literman model. The Black-Litterman portfolio, which applies the results of the stock price direction prediction using machine learning algorithms to the investor's view, provides a better return on risk than the market portfolio and market index, and the Black-Litterman portfolio with the ETE network indicator has the highest yield. The use of ETE and stock price prediction leads to improved return on investment, and improving predictive performance increases the return on investment.

This dissertation is the first study on the optimal portfolio establishment strategy through the Black-Litterman model and stock price direction prediction using machine learning algorithm to apply ETE of information theory to the financial investment field.

Keywords: Information theory, Econophysics, Transfer entropy, Machine learning, Feature engineering, Prediction algorithms, Stock markets, Time series analysis, Black-Litterman model, Optimal asset allocation, Portfolio management
Student Number: 2016-30252

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Stock markets have been studied extensively as one of the crucial fields of economy(Ghysels and Osborn, 2001). In particular, research has been actively conducted to analyze and predict the stock market based on relationships among the dynamics of stock prices and returns. Since the stocks exhibit diverse interactions, many theoretical or empirical studies of such relationships have provided meaningful implications to investors and policy-makers developing appropriate actions regarding the market condition. Specifically, the prediction on stock price and the overall market and constructing a an appropriate portfolio is one of the essential tasks for investors to establish an optimal investment strategy.

Many previous studies have utilized concepts in statistical physics such as complex systems and information theory to quantify the correlations among the entities in an economic or financial system(Mantegna and Stanley, 1999; Noh, 2000; Bonanno et al., 2003; Zunino et al., 2008; Chi et al., 2010). Notably, the correlation analysis is a simple and good indicator for measuring the degree of similarity between two variables. Many studies of correlation-based time series analysis have revealed the characteristics of the system using random matrix theory and network analysis(Plerou et al., 2002; Kim et al., 2011; Kumar and Deo, 2012). Since then, the studies have discovered that a linear model such as the Pearson correlation is not sufficient enough to quantify the relationships among the stocks. More importantly, the causal relationship is not directly linked to the presence of correlation. In this context, the Granger-causality(Granger, 1969) has been introduced to define the causal relationship between time series. However, its function is limited to express the existence of information flow based on a linear relationship rather than measuring the amount of information flow.

To overcome such limitations of a simple linear model of a Granger-causal relationship, the concept of transfer entropy (TE), proposed by Schreiber(Schreiber, 2000), has been suggested instead to measure the amount of information flow. TE is a non-parametric measure of the amount of information transfer from a variable to a variable based on the Shannon entropy(Shannon, 1948). TE has been widely used in researches such as social networks, neuroscience, and financial market analysis due to its ability to capture the asymmetrical interactions within the system and to distinguish the driving and responding elements efficiently(Kim et al., 2016; Faes et al., 2013; Vicente et al., 2011). Despite its advantages, TE could involve a noise due to its requirement on a large amount of data. Thereupon, the effective transfer entropy (ETE)(Marschinski and Kantz, 2002) has been proposed to obtain a more robust quantification of information flow. Since then, many studies have utilized TE and ETE to identify information flows in different financial markets(Sensoy et al., 2014; Kwon and Yang, 2008b,a; Kwon and Oh, 2012; Dimpfl and Peter, 2013; Sandoval, 2014; Chunxia et al., 2016; Lim et al., 2017; Yue et al., 2020).

Based on findings in Granger-causal relationships within the financial system, the prediction on the price or volatility of the stock market has been widely studied(Clements et al., 2004). Especially, the prediction on stock price is a critical issue since an improved prediction performance can guarantee a higher expected return to investors. In this regard, many previous studies have attempted the prediction based on various models(Ola et al., 2014; Moradi et al., 2019) including traditional linear models to machine learning models such as an artificial neural network(Tsaih et al., 1998; Guresen et al., 2011; Rather et al., 2015; Zahedi and Rounaghi, 2015), random forest(Patel et al., 2015; Ballings et al., 2015), support vector machine(Kim, 2003; Kara et al., 2011), XGBoost(Nobre and Neves, 2019; Jiang et al., 2020), and long short-term memory(Chen et al., 2015; Nelson et al., 2017). Commonly, two different approaches are considered in machine learning-based stock price prediction. The first is to improve the existing machine learning models theoretically(Guo et al., 2015; Qiu and Song, 2016), and the second is to integrate created variables such as network indicators, Google trends, and public announcements in addition to a simple set of stock price and return series(Lee et al., 2019; Hu et al., 2018; Feuerriegel and Gordon, 2018; Hagenau et al., 2013; Geva and Zahavi, 2014; Chan and Franklin, 2011; Shynkevich et al., 2016; Ntakaris et al., 2019).

In addition to research on prediction of the stock price and the stock price direction, exploring asset allocation methods that minimize risk and maximize profits is also an important task for investors to develop optimal investment strategies. Markowitz (1952) proposed a mathematical model of efficient asset allocation which is also called mean-variance model and completely transformed the field of portfolio optimization, and its derived contents are still used to construct almost any portfolio. Markowitz introduced the concept of an efficient frontier and suggested a way to construct an optimal asset ratio that maximizes returns under a given risk. Although this methodology is innovative, Markowitz's work presents some major drawbacks in practical applications. The resulting portfolio may not be intuitive, and the available stocks are biased and not sufficiently dispersed. The optimal portfolio is also very sensitive to small changes in input data. One of the well-known modern approaches to asset allocation based on Markowitz's portfolio selection model to improve this problem is the Black-Litterman model(Black and Litterman, 1990, 1992). This model produces a more stable and diverse portfolio than the mean-variance model. The main contribution of this model is to provide the flexibility to combine the capital asset pricing model (CAPM) market equilibrium with the investor's view of asset returns, which results in an intuitive and decentralized portfolio. There are two different approaches are considered in the application of the Black-Litterman model. The first is to improve the models itself such as distributions(Xiao and Valdez, 2015; Fang et al., 2018; Fernandes et al., 2018; Palczewski and Palczewski, 2019), and the second is to apply the investor's view using various prediction models or variables(Beach and Orlov, 2007; Palomba et al., 2008; Creamer, 2015; Silva et al., 2017; Pyo and Lee, 2018; Kara et al., 2019).

Although many studies have devoted to integrating the network-driven indicators, an attempt to utilize the concept of entropy in stock price prediction has not been widely studied. Instead, the previous studies on TE have focused on revealing the statistical explanatory power between various financial variables. Furthermore, the previous studies related to TE and ETE analysis have focused on the inter-market and static-interval analysis, which occurs a limited finding in intra-sector and dynamic-interval. In addition, few studies have dealt with the problem of asset allocation based on interdisciplinary studies of econophysics and machine learning.

In this dissertation, we focus on three main topics: (1) obtaining entropy-driven indicators with market explanatory power in the US market, (2) integration of entropy-driven indicators in machine learning algorithms to improve the performance of prediction in the direction of a stock price, (3) and the practical use of previous studies through the Black-Litterman model.

1.2 Research Objectives

Research focuses on the US financial market, the largest single market in the global financial system, and study 55 companies from 11 different industry sectors based on moving window methods. Indeed, the moving window method allows dynamic and time-varying observations for different crisis and non-crisis periods based on interval analysis. Once we explore that the evolution of ETE can express the dynamics of stock prices, we utilize it as an extra variable for various machine learning algorithms to predict its direction. By comparing the performances of sets with and without the ETE variable, we confirm the usability of ETE in stock price prediction.

In summary, the random forest (RF) of Breiman (2001) has been recognized as a decent classifier that has been heavily used in stock price prediction. Also, the XGBoost (XGB) of Chen and Guestrin (2016) is also a popular boosting-type ensemble classifier used for classification problems. The long short-term memory (LSTM) network of Hochreiter and Schmidhuber (1997), a model that improves the exploding and vanishing gradient problem in the recurrent neural network, has shown decent performances in sequence learning and time series prediction, which eventually leads to studied in financial time series prediction. Thus, in this study, we focus on discovering a useful input variable in predicting the stock price direction by utilizing the ETE-driven network indicator based on five representative machine learning algorithms: logistic regression (LR) of Cox (1958) as a traditional predictive model, multilayer perceptron (MLP) of Rosenblatt (1961) as a back-propagated neural network, RF as a bagging-type ensemble method, XGB as a boosting-type ensemble method, and LSTM as a single classifier model.

Note that we utilize the results of the characteristics of the ETE network indicator that describes the market and those of the prediction of stock price direction through machine learning algorithms using ETE network indicator to the investor's view of the Black-Litterman model. The proposed Black-Litterman model compares the market capitalization portfolio and the market index with performance to verify whether it is practically effective in establishing an optimal asset allocation strategy.

1.3 Organization of the Thesis

The rest of this dissertation is organized as follows. Chapter 2 reviews related previous studies of Granger-causal relationship using TE and ETE in financial markets, studies of stock price prediction using machine learning algorithms, and the studies on the application of the Black-Litterman model to optimal asset allocation problems. Chapter 3 describes the concept of ETE and focuses on the time-varying and cross-section analysis through ETE in the US market. Chapter 4 demonstrates machine learning algorithms for stock price direction prediction and the analysis framework, including different set-ups for prediction. In addition, stock price direction prediction using ETE network indicator is conducted and analyzed in terms of models and sectors, and analysis is performed using defined performance metrics. In Chapter 5, the step-by-step procedures of the Black-Litterman model including reflecting the investor's view is described and a comparative analysis of Black-Litterman portfolio, market capitalization portfolio, and the market index is conducted based on the results of Chapter 3, 4. Finally, the contributions and limitations of this dissertation are provided in Chapter 6 with future work for improvement.

Chapter 2

Literature Review

This chapter introduces previous studies on financial market analysis through TE and ETE, as well as previous studies on predicting the stock price and the stock price direction using machine learning algorithms. It also covers previous studies of optimal asset allocation using the Black-Litterman model.

2.1 Analysis of transfer entropy

Since the concept of Granger-causality(Granger, 1969) was discovered, many studies have utilized it to detect the Granger-causal relationship in various financial time-series. For instance, Mok (1993) analyzed the relationship among the daily stock prices, exchange rates, and interest rates of Hong Kong using the autoregressive integrated moving average and Granger-causality test, and confirmed the sporadic unidirectional Granger-causality from stock price to interest rate as well as a weak bidirectional Granger-causality between stock price and exchange rate. Swanson et al. (1999) conducted a multivariate time series analysis on the data revision process for industrial production (IP) and the composite leading indicator (CLI) in the United States. They confirmed a kind of causal feedback of revision processes for IP and CLI by showing that previously available IP revisions are useful for describing

CLI revisions and that IP revisions are predictable from past CLI revisions. Soytas and Sari (2003) conducted a Granger-causality test between energy consumption and GDP in nine emerging market groups and seven advanced countries. However, the Granger-causality is a model-specific approach that assumes a specific underlying dynamics such as the Vector Autoregressive (VAR) and the Vector Error Correction Model (VECM), which focuses on the existence of a Granger-causal relationship rather than the degree of a Granger-causal relationship.

In contrast, the TE(Schreiber, 2000) is a model-free measure coping with the limitation of Granger-causality in quantifying the degree of information transfer without any constraints such as linear dynamics. The previous studies on Granger-causality detection through information flow analysis using TE and ETE in financial markets are as follows. Marschinski and Kantz (2002) analyzed the relationship between the Dow Jones Industrial Average (DJIA) and the German DAX Xetra Stock Index (DAX) using ETE, and discovered asymmetry information flow where more information is transferred from the DJIA to the DAX than the vice versa. Kwon and Yang (2008b) performed a TE analysis using daily data of 25 global financial market indices and discovered that the US stock market has the most significant impact on the global stock market as well as confirmed that the Asia and Pacific market received information the most. Sensoy et al. (2014) considered the direction and intensity of information flow for the exchange rate and stock price of nine emerging countries through ETE. Specifically, a low level of interaction before the 2008 financial crisis was revealed with the dominance of exchange rate over the stock prices in the crisis, whereas a robust bidirectional interaction in and after the crisis with a dominance of stock prices over the exchange rate. However, the studies of Kwon and Yang (2008b) and Sensoy et al. (2014) are limited in showing an overall level of relationships without considering the relationships at industry or individual stock level.

In this context, Kwon and Yang (2008a) conducted a TE analysis of the Dow Jones index, S&P 500 index, and 125 stocks in the US market, and confirmed that individual stock prices are affected by the market index considering the direction of information transfer. Also, Kwon and Oh (2012) performed TE analysis between the market index and stock price for nine emerging or mature stock markets, and confirmed the market index as the primary driving force for determining individual stock prices and higher asymmetric information flow of developed countries than that of emerging countries. Dimpfl and Peter (2013) analyzed the information flow between the credit default swap market and the corporate bond market using the TE of 27 iTraxx companies before and after the financial crisis, and discovered the dominance of CDS market over the corporate bond market, increment of information transmission between markets over time, and the highest importance of the CDS market in the crisis period. Sandoval (2014) applied ETE analysis to 197 global financial companies selected based on the amount of market capitalization. Through assessing influence among companies as well as their network structure, he revealed the most significant impact and importance of banks and insurance companies of European and the US to the global financial markets. Chunxia et al. (2016) analyzed the correlation between the information flow and trading volume through TE among ten sectors of the US market. As the effect of the financial crisis intensified, the information flow between sectors increased, and the financial sector always showed a massive outflow of the TE at all intervals. In particular, it is confirmed that the main information flow from the financial sector changed before and after the financial crisis. Lim et al. (2017) analyzed the information flow of the CDS and stock market in US through the TE into the inter and intra structure aspects, and confirmed a substantial change in the information transfer during the financial crisis as well as the precedence of the sudden change of transfer entropy in the CDS market than that of stock market. Yue et al. (2020) analyzed the information flows among the sectors in the Chinese stock market using the transfer entropy. They used the maximum spanning arborescence to extract information flow and the hierarchical structure of the networks. They identified the information source and sink sectors and observed that the root node sector acts as an information sink of the incoming information flow networks.

Table 2.1 summarizes the methodology of previous studies. In general, the previous studies tend to compress information by choosing three to five bins in the binning process to discretize the stock returns. Furthermore, the static analysis of certain time intervals has studied in the financial network analysis to identify non-linear Granger-causal relationships. Then, the studies have devoted to demonstrating the empirical findings rather than its practical application. Therefore, in this study, we set 22 bins(Sandoval, 2014) to reflect as much information in stock returns as possible. Then, we utilize the moving window method to analyze the time-varying dynamics of ETE and apply its values in predicting the direction of stock prices.

2.2 Stock price prediction based on machine learning

In recent years, as the use of machine learning algorithms has attracted attention in academia, many studies have attempted and discovered the utilization of various

Table 2.1: Outline of previous literature on TE and ETE

Reference	**Method**	**Dataset**	**moving window**	**binning method**	**Prediction**
Marschinski and Kantz (2002)	ETE	SMI	no	symbolic (2 to 5)	no
Kwon and Yang (2008b)	TE	SMI	no	symbolic (3)	no
Kwon and Yang (2008a)	TE	SMI / SP	no	symbolic (3)	no
Kwon and Oh (2012)	TE	SMI / SP	no	symbolic (2)	no
Dimpfl and Peter (2013)	ETE	CDS / COB	cross-section	symbolic(3)	no
Sensoy et al. (2014)	ETE	SMI / ER	cross-section	symbolic(3)	no
Sandoval (2014)	ETE	SP	no	binning (24)	no
Chunxia et al. (2016)	TE	SI	580 days	symbolic (3)	no
Lim et al. (2017)	TE	CDS / SP	550 days	symbolic (4)	no
Yue et al. (2020)	TE	SI	no	binning (15)	no
The proposed approach	ETE	SP	20/60/120/240 days	binning (22)	yes

*Note: TE and ETE in the method are the abbreviations of the transfer entropy, effective transfer entropy, respectively. SMI, SP, COB, ER, and SI in the dataset are the abbreviations of the stock market index, stock price, corporate bond, exchange rate, and sector index, respectively. The cross-section in the moving window includes the analyses in the pre-crisis, crisis, and post-crisis.

machine learning algorithms and their adequacies in price prediction in financial markets. Note that the problem of predicting the direction of cumulative returns of stock in the future can be considered as a classification problem.

For instance, Ballings et al. (2015) compared the predictive performance of European stock prices by AUC among the RF, AdaBoost, and Kernel Factory, Neural Networks, Logistic Regression, Support Vector Machines, and K-Nearest Neighbor where the RF showed the highest performance. Patel et al. (2015) predicted the stock price direction of the Indian stock market using the neural network, support vector machine, RF, and naive-Bayes classifier. At first, 10 technical parameters from the stock transaction data is applied as input parameters. Then, these parameters are converted into trend deterministic data. As a result, the RF showed the best prediction performance. Fischer and Krauss (2018) applied the LSTM networks to predict the direction of the stock price for the constituent stocks of the S&P 500 from 1992 until 2015, where the LSTM networks outperformed the memory-free classification methods such as an RF, deep neural net, and logistic regression classifier. Bao et al. (2017) proposed an ensemble approach consisting of the wavelet transforms, stacked autoencoders (SAEs), and LSTM to predict six stock price indices, namely CSI 300, Nifty 50, Hang Seng index, Nikkei 225, S&P 500, and DJIA. Specifically, the wavelet transforms is utilized to decompose the time series of stock prices to remove the noise; SAEs are used to generate deep high-level features for forecasting; LSTM is applied to predict the next day's stock price based on generated features. Briefly, the primary purposes of previous researches on stock price prediction using machine learning algorithms are twofold: advance in the algorithm and utilization of various variables to improve the prediction performance.

In the perspective of advancing the machine learning algorithms in price prediction, Guo et al. (2015) proposed the advanced neural network model by incorporating the principal component analysis and radial basis function. In contrast, Qiu and Song (2016) proposed a genetic algorithm-based neural network in price prediction. Tsaih et al. (1998) proposed a rule-based system trading strategies to predict the direction of the S&P 500 stock index futures using the reasoning neural network. The results confirmed that the proposed model outperformed the backpropagation network and perceptron neural network. Tsai and Hsiao (2010) proposed a multiple feature selection model combining principal component analysis, genetic algorithm, and decision trees. Then, a back-propagation neural network is applied as a prediction model, which yields the best performance in predicting the price direction of the electronic corporations in Taiwan stock exchange. Kao et al. (2013) proposed a stock price prediction model by incorporating the wavelet transform, multivariate adaptive regression splines, and support vector regression for various emerging and mature markets. The model solved the wavelet sub-series selection problem and confirmed that the prediction accuracy outperforms the other five competing approaches. In addition, a sub-series selected by the model can identify which points in the past stock price data have had a significant impact on the predictive model configuration. Shin et al. (2013) has applied the semi-supervised learning, the model has been used for non-time-series type data, to the time series to predict the direction of crude oil prices. In order to apply the existing model to time series prediction, the semi-supervised learning is modified by considering the similarity between different sets of time series data, labels in the stock direction, technical indicator transformation, and feature selection. Ola et al. (2014) confirmed that daily returns of Tehran Stock

Exchange stocks are a chaotic process and that prediction can be improved by applying time-series tests of the local polynomial approximation model. Zahedi and Rounaghi (2015) applied artificial neural network (ANN) and principal component analysis (PCA) to the 20 accounting variables of stocks listed on the Tehran Stock Exchange to confirm the superiority of stock price forecasting through ANN and the effective factor through the PCA method. Moradi et al. (2019) analyzed and predicted stock returns by applying the lags coevolving with radial basis function networks (L-Co-R) algorithm for the Tehran Stock Exchange and London Stock Exchange, and uses the Box-Jenkins methods to analyze the Fractal market hypothesis (FMH) between the two exchanges. As a result, they discovered the fact that L-Co-R algorithm is more applicable for long-term time series, that the Box-Jenkins method performs better in short-term time series, and that the FMH is accepted in Tehran Stock Exchange but rejected for London Stock Exchange.

In the perspective of utilizing new input variable, many variables have been proposed including the price-driven measure(Pyo et al., 2017; Ntakaris et al., 2019), volatility-driven measure(Lee et al., 2019; Ntakaris et al., 2019), and Google trends(Hu et al., 2018), which extended to the text mining-based variables from the financial news (Feuerriegel and Gordon, 2018; Hagenau et al., 2013; Geva and Zahavi, 2014; Chan and Franklin, 2011; Shynkevich et al., 2016). Specifically, Khansa and Liginlal (2011) analyzed the relationship between the stock returns of information security firms and the intensity of malicious attacks using artificial neural networks and vector autoregression analysis. The results confirmed that the malicious intensity has a one-month-lagged positive effect on the stock price return of information security firms, and the time-delayed artificial neural network acts as a complementary approach to the existing VAR analysis. In this milieu, the time-delayed artificial neural network showed 95% accuracy while the regression counterpart only showed 85%. Nam and Seong (2019) performed the prediction on stock price direction using various machine learning algorithms by incorporating the Granger-causality of the financial news data in the Korean market. Note that the TE is applied in Granger-causality analysis and multiple kernel learning to combine features of target firms and Granger-causal firms. The results confirmed that the proposed model outperformed the benchmarks and verified that the direction of stock price could be predicted based on the news of the Granger-causal firms even when the target firms had no news.

2.3 The Black-Litterman model

Black and Litterman's original paper(Black and Litterman, 1990, 1992) only described the key aspects of the idea, and others described in detail how to apply this model.(He and Litterman, 2002; Satchell and Scowcroft, 2000; Idzorek, 2007) Briefly, two different approaches are considered in the applications of the Black-Litterman model. One is to change the model such as the distribution of asset returns(Xiao and Valdez, 2015), and the other is to apply a variety of methods to the investor's view from the traditional linear model(Beach and Orlov, 2007; Palomba et al., 2008) to the machine learning model.

From the perspective of improving the models, Palczewski and Palczewski (2019) extended the existing Black-Litterman model into general continuous distributions and deviation measures of risk. They show that the Black-Litterman model can be extended to the distribution of the investor's view and the arbitrary prior equilibrium distribution and that the choice of the prior distribution and the distribution of investor's view has a significant influence on the posterior distribution, optimal portfolio weights, and their performance. Also, they confirmed that the prior mean returns can be robustly calculated for a number of deviation risk measures and that the importance sampling procedure of the proposed model significantly improves computations of the estimation of the prior mean returns and portfolio CVaR. Fang et al. (2018) described the investor's views of Black-Litterman model as fuzzy sets and proposed two Black-Litterman models using fuzzy views and fuzzy random views respectively. In the CSI 300 market, compared to the mean-variance model, the market capitalization portfolio, and the Black-Litterman model, the proposed models outperformed the existing model in that it effectively covers the information of the view and describes the uncertainty of the view. Fernandes et al. (2018) proposed an investment strategy based on Black-Litterman model with conditional information in Brazilian stock market. He proposed a method for determining 1-step ahead returns using price-earnings ratio and the past returns in consideration of investors' risk profiles. As a result, it was confirmed that the proposed method updates the conditional probability distribution of asset returns and alleviates the asset allocation instability due to estimation error, and the Black-litterman portfolio constructed by the proposed method outperforms the mean-variance portfolio.

From the perspective of applying investor's view in various ways, Creamer (2015) proposed a method based on either the news sentiment using high-frequency data or on a combination of accounting variables to apply the investor's view of the BL model, and showed superiority compared to the market portfolio and market index. Silva et al. (2017) proposed a new way of creating an investor's view using Verbal Decision Analysis (VDA) in the Brazilian stock market. The proposed method mitigates the impact of poor view estimation and allows investors to create a personal risk-return balanced portfolio without the help of experts. Pyo and Lee (2018) applied the low-risk anomaly characteristics to the Black-Litterman model for KOSPI200 in the Korean stock market. Artificial neural network (ANN), support vector regression (SVR), gaussian process regression (GPR), and GARCH are used to predict the market volatility of assets and ANN performed best. In these results, they identified high- and low-volatility stocks by volatility predicted from ANN and applied a relative view that low volatility stocks would yield better than high volatility stocks, confirming that the low-risk portfolio outperformed than high-risk portfolio. Kara et al. (2019) applied GARCH and SVR to Black-Litterman model for the Istanbul Stock Exchange and Dow Jones Index of the US stock market. GARCH modeling is used to predict eight econometric indicators, and these prediction results are again used to predict return through SVR and finally reflect this result to the investor's view in the Black-Litterman model. For both markets and different holding periods, the proposed model outperformed index returns and random generated portfolios for returns and Sharpe ratio.

Chapter 3

Effective transfer entropy analysis for the US market

3.1 Effective transfer entropy

TE(Schreiber, 2000) is a non-parametric indicator to measure the statistical explanatory power between two processes. Note that the TE can detect the non-linear Granger-causal relationship. It is widely used in fields such as social networks, neuroscience, and financial market analysis due to its efficient detection of the asymmetric interaction in the system.

When two variables interact with each other, a time series of one variable Y can affect the future time point of another time series of variable X. Let a time series X is a Markov process of degrees k; then it refers that the state x_{n+1} of X is affected by k previous states of the same variable, which can be expressed as,

$$\begin{aligned} &p(x_{n+1}|x_n, x_{n-1}, \ldots, x_0) \\ &= p(x_{n+1}|x_n, x_{n-1}, \ldots, x_{n-k+1}), x_i \in X \end{aligned} \tag{3.1}$$

where $p(A|B)$ represents a conditional probability of A given B, $p(A|B) = p(A, B)/p(B)$.

Furthermore, let the state x_{n+1} of X is dependent on the l previous states of Y, then the TE from a variable Y to a variable X can be defined as the average information included in Y excluding the information reflected by the past state of X for the next state information of X. Therefore, if X and Y denote the amount of information measured by the Shannon entropy($= -\sum_i p_i \log_2 p_i$), and the variable x_{n+1} of X is affected by k previous states of X and l previous states of Y, the TE from the variable Y to the variable X can be defined as follows.

$$\begin{aligned} TE_{Y \to X}(k,l) &= \sum_{x_{n+1}, x_n^{(k)}, y^{(l)}} p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 p(x_{n+1}|x_n^{(k)}, y_n^{(l)}) \\ &- \sum_{x_{n+1}, x_n^{(k)}, y^{(l)}} p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 p(x_{n+1}|x_n^{(k)}) \\ &= \sum_{x_{n+1}, x_n^{(k)}, y^{(l)}} p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 \frac{p(x_{n+1}|x_n^{(k)}, y_n^{(l)})}{p(x_{n+1}|x_n^{(k)})} \end{aligned} \tag{3.2}$$

where $x_n^{(k)} = (x_n, x_{n-1}, \ldots, x_{n-k+1})$, $y_n^{(l)} = (y_n, y_{n-1}, \ldots, y_{n-l+1})$, and $p(A,B)$ is the joint probability of A and B. The definition of TE assumes that events at some point are affected by events of k and l previous states. Based on the previous research(Sandoval Jr, 2014) showing the low memory in the log-returns of stock prices, we computed the TE under the conditions $k = l = 1$, which expresses the weak form of efficient market hypothesis stating that the current price reflects all past information. Hence, the (3.2) can be re-defined as follows.

$$TE_{Y\to X} = \sum_{x_{n+1},x_n,y_n} p(x_{n+1},x_n,y_n)\log_2\frac{p(x_{n+1}|x_n,y_n)}{p(x_{n+1}|x_n)} \\ = \sum_{x_{n+1},x_n,y_n} p(x_{n+1},x_n,y_n)\log_2\frac{p(x_{n+1},x_n,y_n)p(x_n)}{p(x_{n+1},x_n)p(x_n,y_n)} \tag{3.3}$$

In summary, $TE_{Y\to X}$ is the difference between the information regarding the future value of X_i obtained from X_i and Y_i and the information regarding the future value of X_i obtained only from X_i. So, the positive $TE_{Y\to X}$ indicates that the variable Y affects the future value of the variable X, which can be interpreted as the degree of uncertainty that decreases when Y is considered. In the same context, $TE_{Y\to X}$ means the degree to which the dynamics of Y affects the transition probability of X, which can be seen as the amount of information flow from Y to X. Therefore, a large TE value refers to more significant information flow. Also, the TE can measure the amount of in-flow information coming from Y to X through $TE_{Y\to X}$ and the amount of out-flow information going from X to Y through $TE_{X\to Y}$, respectively, based on its asymmetric property.

TE is a proper measure to estimate a statistical dependency regardless of the data type. However, a relatively large amount of data is required to derive the transfer entropy. Also, a critical disadvantage of TE is its inclusion of noise due to the finite sample effects and non-stationarity of data. In this context, ETE(Marschinski and Kantz, 2002) is proposed to solve the disadvantage of TE. At first, we randomly shuffle the elements of each time series to break the statistical explanatory power between variables, yet keeping the individual probability distributions of each time series. Then, we obtain transfer entropy from this time series, called the randomized

TE (RTE). Finally, we obtain ETE by subtracting the RTE from the original TE to eliminate the noise.

$$ETE_{Y \to X} = TE_{Y \to X} - RTE_{Y \to X} \tag{3.4}$$

3.2 Data and experiment set-ups

3.2.1 Data

In this study, the classification of industry and its constituent stocks are based on *The MSCI(Morgan Stanley Capital International) USA IMI(Investable Market Index) Sector Indexes* as of December 31, 2018. This index covers approximately 99% of the 2,400 large-, mid-, and small-cap stocks in the US stock market and divides stocks into 11 sectors based on the Global Industry Classification Standard (GICS®).

Based on the stocks in which the company continuously exists within the period of the experiment, we utilize a total of 55 stocks, the top five stocks by market capitalization in each sector. Note that all the stocks are listed on the NASDAQ and NYSE. The data period is from January 3, 2000 to December 31, 2018, which yield 4,779 daily adjusted stock prices of 55 stocks. The data is extracted from Thomson Reuters Datastream. Table 3.1 shows the 11 sectors used in the study and the corresponding stocks and the descriptive statistics of the log-returns of entire data are summarized in Table 3.2.

Table 3.1: Description of sectors and the corresponding stocks

SECTOR	COMPANY	TICKER	SECTOR	COMPANY	TICKER
Telecommunication	Verizon Communications	VZ	**Financials**	Jpmorgan Chase & Co	JPM
Services	Disney (Walt)	DIS	**(FINC)**	Berkshire Hathaway B	BRK-B
(TELE)	At&T	T		Bank Of America Corp	BAC
	Comcast Corp A	CMCSA		Wells Fargo & Co	WFC
	21St Century Fox A	FOXA		Citigroup	C
Consumer	Amazon.Com	AMZN	**Consumer**	Procter & Gamble Co	PG
Discretionary	Home Depot	HD	**Staples**	Coca Cola (The)	KO
(COND)	Mcdonald'S Corp	MCD	**(CONS)**	Pepsico	PEP
	Nike B	NKE		Walmart	WMT
	Starbucks Corp	SBUX		Altria Group	MO
Energy	Exxon Mobil Corp	XOM	**Information**	Apple	AAPL
(ENRG)	Chevron Corp	CVX	**Technology**	Microsoft Corp	MSFT
	Conocophillips	COP	**(INFT)**	Intel Corp	INTC
	Eog Resources	EOG		Cisco Systems	CSCO
	Schlumberger	SLB		Oracle Corp	ORCL
Health Care	Johnson & Johnson	JNJ	**Utilities**	Nextera Energy	NEE
(HLCA)	Pfizer	PFE	**(UTIL)**	Duke Energy Corp	DUK
	Unitedhealth Group	UNH		Dominion Energy	D
	Merck & Co	MRK		Southern Company (The)	SO
	Abbott Laboratories	ABT		Exelon Corp	EXC
Industrials	Boeing Co	BA	**Real Estate**	American Tower Corp	AMT
(INDS)	3M Co	MMM	**(REES)**	Simon Property Group	SPG
	Union Pacific Corp	UNP		Crown Castle Intl Corp	CCI
	Honeywell International	HON		Prologis	PLD
	United Technologies Corp	UTX		Public Storage	PSA
Materials	Dowdupont	DWDP			
(MTRS)	Ecolab	ECL			
	Air Products & Chemicals	APD			
	Sherwin-Williams Co	SHW			
	Newmont Mining Corp	NEM			

Table 3.2: Descriptive statistics for representative stocks in S&P 500

TICKER	Mean	Std	Skew	Kurt	J-B	ADF
VZ	0.0002	0.0156	0.16	9.88	9445***	-71.1***
DIS	0.0003	0.0189	-0.09	12.85	19330***	-71.5***
T	0.0001	0.0162	0.03	10.47	11101***	-69.4***
CMCSA	0.0002	0.0209	0.00	11.35	13883***	-74.5***
FOXA	0.0003	0.0214	0.25	11.79	15435***	-70***
AMZN	0.0006	0.0331	0.45	15.15	29545***	-68.9***
HD	0.0003	0.0196	-0.98	26.37	109504***	-68.6***
MCD	0.0004	0.0145	-0.14	9.71	8987***	-70.7***
NKE	0.0008	0.0190	-0.27	14.82	27867***	-70***
SBUX	0.0007	0.0213	0.15	10.36	10798***	-74.5***
XOM	0.0002	0.0152	0.02	13.50	21943***	-76.1***
CVX	0.0004	0.0160	0.05	13.66	22628***	-73.7***
COP	0.0005	0.0189	-0.31	8.48	6055***	-71.6***
EOG	0.0007	0.0239	-0.04	7.70	4390***	-70.6***
SLB	0.0001	0.0223	-0.34	8.93	7085***	-71.4***
JNJ	0.0003	0.0121	-0.62	19.15	52207***	-69.6***
PFE	0.0002	0.0157	-0.24	8.56	6201***	-70.7***
UNH	0.0008	0.0197	0.26	23.56	84202***	-70***
MRK	0.0002	0.0173	-1.43	32.35	173114***	-68.9***
ABT	0.0005	0.0149	-0.48	12.01	16361***	-68.8***
BA	0.0005	0.0189	-0.28	8.82	6803***	-69.4***
MMM	0.0004	0.0145	-0.02	8.44	5884***	-71.9***
UNP	0.0007	0.0178	-0.25	6.96	3164***	-71.2***
HON	0.0003	0.0196	-0.24	17.02	39193***	-69.1***
UTX	0.0004	0.0168	-1.23	29.76	143791***	-73.7***
DWDP	0.0002	0.0220	-0.21	9.83	9335***	-71.8***
ECL	0.0005	0.0147	-0.08	9.69	8911***	-75.4***
APD	0.0005	0.0173	-0.16	8.72	6543***	-70.3***
SHW	0.0007	0.0177	-0.53	17.93	44571***	-75***
NEM	0.0001	0.0255	0.19	7.69	4405***	-74.1***
JPM	0.0003	0.0246	0.26	17.05	39350***	-74.7***
BRK-B	0.0004	0.0140	0.74	15.44	31264***	-69***
BAC	0.0001	0.0292	-0.35	30.17	147036***	-69.6***
WFC	0.0003	0.0238	0.87	31.06	157326***	-76.1***
C	-0.0004	0.0309	-0.54	45.06	352428***	-65.1***
PG	0.0002	0.0135	-4.10	113.51	2444856***	-71.9***
KO	0.0003	0.0130	0.03	12.06	16343***	-69.8***
PEP	0.0003	0.0123	0.03	14.64	26955***	-72.6***
WMT	0.0001	0.0150	0.10	10.07	9962***	-71.4***
MO	0.0007	0.0153	-0.04	16.26	35012***	-70.5***
AAPL	0.0009	0.0269	-4.36	121.20	2796451***	-71.1***
MSFT	0.0002	0.0193	-0.14	12.66	18580***	-72.4***
INTC	0.0001	0.0235	-0.47	12.08	16576***	-72.1***
CSCO	0.0000	0.0245	0.14	12.15	16669***	-73.3***
ORCL	0.0001	0.0248	-0.02	11.52	14453***	-74.6***
NEE	0.0006	0.0140	0.16	12.70	18740***	-71.9***
DUK	0.0003	0.0153	-0.24	14.98	28599***	-72.1***
D	0.0005	0.0133	-0.61	13.01	20251***	-70.2***
SO	0.0005	0.0119	0.23	9.55	8584***	-73.6***
EXC	0.0004	0.0161	-0.05	11.14	13203***	-72.6***
AMT	0.0004	0.0302	-0.51	27.77	122317***	-64.1***
SPG	0.0006	0.0212	0.24	22.49	75696***	-83***
CCI	0.0003	0.0299	-0.30	28.10	125513***	-65.1***
PLD	0.0004	0.0250	-1.00	35.39	209679***	-80.1***
PSA	0.0006	0.0187	0.12	19.07	51452***	-82.1***

*Note: Std, Skew, Kurt, J-B, ADF are the abbreviations of the standard deviation, skewness, kurtosis, Jarque-Bera tests, augmented Dickey-Fuller tests, respectively. Also, the star superscripts *, **, *** refer to 5%, 1%, and 0.1% statistical significances, respectively.

3.2.2 Experiment set-ups

At first, we select the six major events to observe changes of ETE during the financial crisis a follows.

1. **Dot-com bubble** (1999/01/02 - 2002/09/30) : The Dot-com bubble occurred in the US market when the Internet began to spread as a new business model. Therefore, the crisis period is set from the beginning of 1999, when large-cap stocks including Qualcomm in the Nasdaq market rose sharply, to Sep 30, 2002, when most dot-com companies went bankrupt, and Nasdaq reached its lowest point.

2. **Subprime mortgage crisis** (2007/04/01 - 2009/06/30) : The subprime mortgage crisis is set from Apr 1, 2007, when New Century, the US's second-largest subprime mortgage lender, requested filing for bankruptcy protection, to the first half of 2009 (Jun 30, 2009), when US Congress announced the American Recovery and Reinvestment Act of 2009.

3. **European crisis** (2010/04/23 - 2010/12/31) : The European crisis is set from Apr 23, 2010, when the Greek government requested financial assistance to the EU and the IMF, to the end of 2010 (Dec 31, 2010).

4. **US debt-ceiling crisis** (2011/04/18 - 2012/01/31) : The US debt-ceiling crisis is set from Apr 18, 2011, when the S&P, a renowned credit rating company, announced its first negative view in history on the US AAA sovereign-debt rating to Jan 31, 2012, when the US set an extremely low-interest rate plan to deal with the fall of its rating.

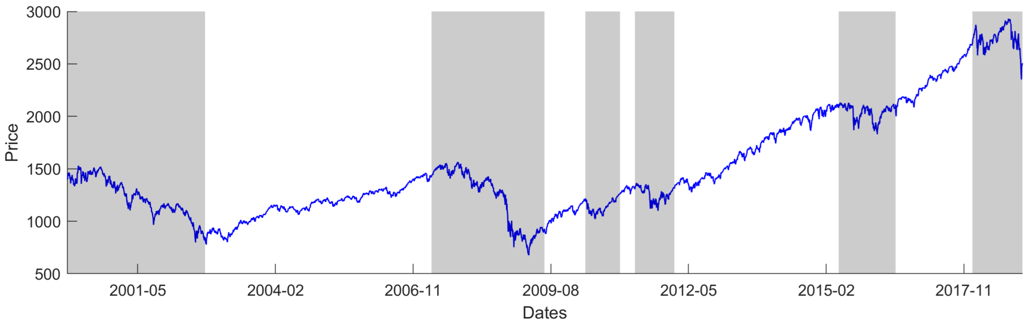
5. **Brexit** (2015/05/07 - 2016/06/23) : The Brexit is set from May 7, 2015, the day of the 2015 UK election, where the conservative party which insisted a hold on the referendum on Brexit won the election, to Jun 23, 2016, the day of the Brexit referendum.

6. **US-China trade war** (2018/01/02 - 2018/12/31) : the US-China trade war is set from January 2018, when the US began to impose sanctions on Chinese companies, to the end of the experimental period.

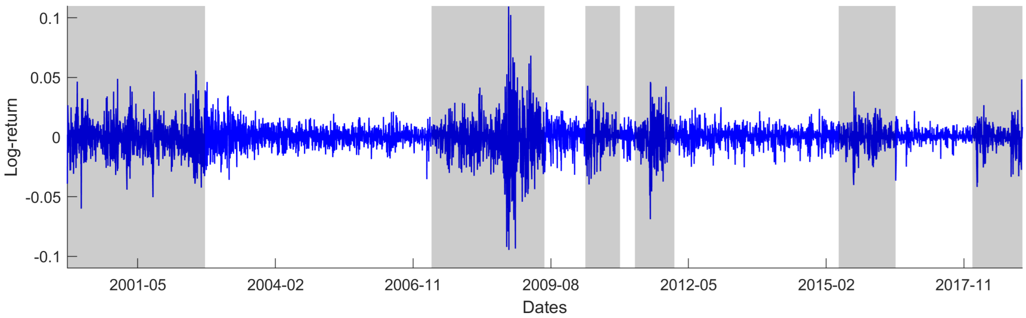
From the perspective of stocks and events, we check whether the 55 stocks employed in this study could represent the US market. Then, we choose the S&P 500 as a representative indicator of the entire US market in order to examine the actual relationship between the selected financial crisis and the US market. The S&P 500 price, log-returns of S&P 500, and the average log-returns of 55 stocks are shown in Figure 3.1.

The result shows that the average log-returns are similar to that of the S&P 500. Furthermore, we set the selected events as gray boxes. In the gray boxes, the volatility of log-returns in the US market is relatively high, indicating that the selected events represent the financial crises in the US market.

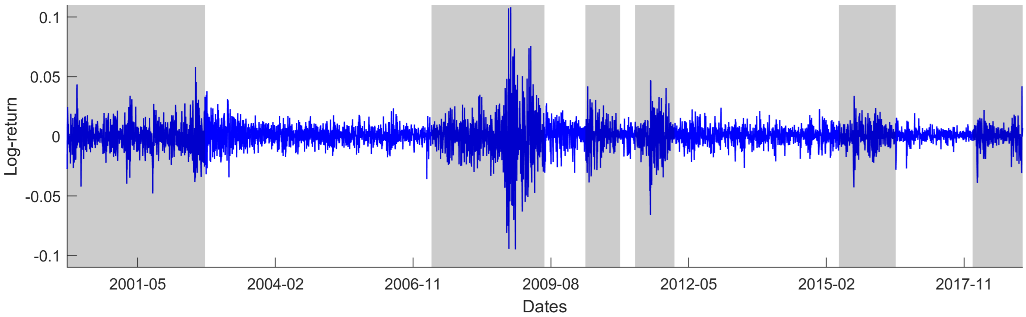
In this study, we choose to use an average of 25 RTE simulations. Also, we set 22 bins to construct the discrete probability distribution for the log-returns of the stock prices. In order to reduce the influence of outliers of log-returns, we integrated the log-returns less than -6% and greater than $+6\%$ into bin #1 and #22, respectively. Then, the interval between -6% and $+6\%$ is divided by 0.6% (from bin #2 to #21). The histogram of the bin for the dataset is shown in Figure 3.2.



(a) S&P 500 price



(b) Log-returns of S&P 500 index



(c) The average log-returns of 55 stocks

Figure 3.1: Price and log-returns of S&P 500, and the average log-returns of 55 stocks

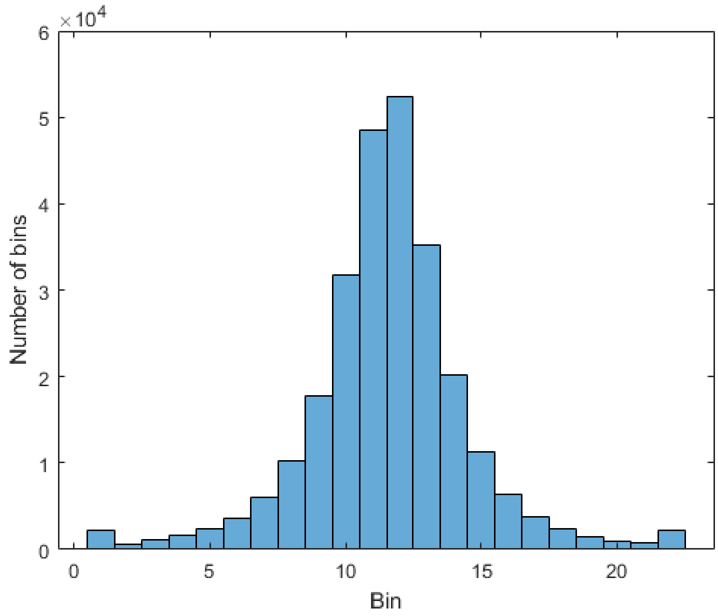


Figure 3.2: Histogram of the bins of all stocks

To analyze the time-varying property of ETE, we use the moving window method and set four different sizes: 1 month (1M, 20 days), 3 months (3M, 60 days), 6 months (6M, 120 days), and 1 year (1Y, 240 days). In this regard, we can obtain 55×55 ETE matrix for 55 stocks for each time point and derive a time series of the total mean value of ETE at each time point. Then, we examine the evolutions of ETE with the major US financial crises to explore appropriate sizes of moving windows size.

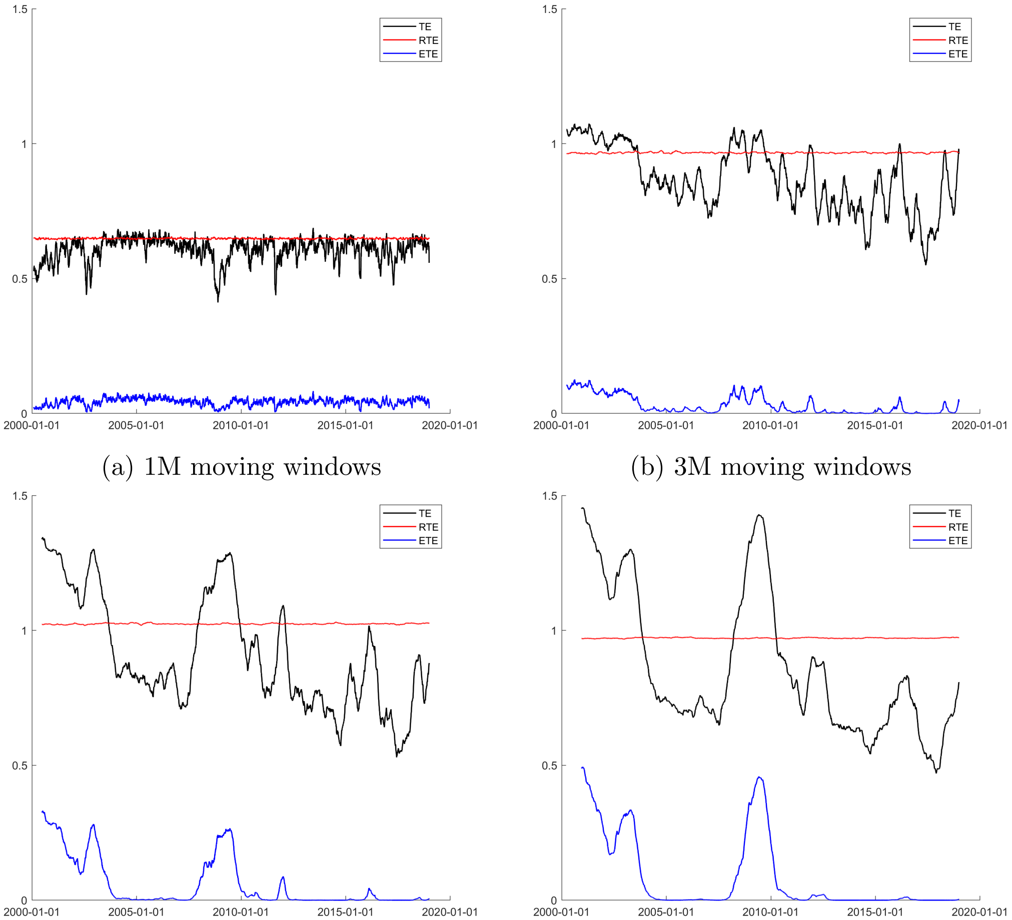
For the selected appropriate moving window-based ETE, we derive the average of inflow ($ETE_{Y \to X}$) and outflow ($ETE_{X \to Y}$) values at each time point and define these two indicators as the ETE network indicator. Furthermore, we perform time series analysis and interval analysis of information flow by sector through the time series of ETE network indicators.

3.3 Results

3.3.1 Overall analysis of Effective transfer entropy

Figure 3.3 shows the effective transfer entropy of 55 major stocks in the US market for different sizes of moving windows. For each moving window, we plot the mean of TE, RTE, and ETE matrix that consists of the entire 55 stocks. In all the moving windows, we confirm that the ETE, a noise-reduced TE through RTE, tends to be more stable than a simple TE. Specifically, the ETE of the 1M moving window shows a noise-shaped oscillation, which fails to detect any particular changes in financial time-series. In contrast, the ETE of the 1Y moving window is too smooth to provide the information for the entire period, where roughly 50% is filled with zero. Therefore, we focus on the analysis of ETEs of 3M and 6M moving windows.

Figure 3.4 shows the average ETEs of 3M and 6M moving windows by presenting the six financial crisis as gray regions. The average ETE is relatively high in the gray regions for both cases, which suggests the implication of financial crisis based on the increased ETE. Note that the average ETE of 3M moving windows is usually smaller than that of 6M moving window, whereas the average ETE of 3M shows more volatile movement with more detailed information than that of 6M moving windows ETE. Since the ETEs of 3M and 6M seem to possess meaningful information regarding the financial markets, we further investigate the characteristics of ETE in terms of the detection of its evolution pattern and cross-sectional analysis on the financial crisis in the level of financial sectors.



(a) 1M moving windows

(b) 3M moving windows

(c) 6M moving windows

(d) 1Y moving windows

Figure 3.3: Evolution of TE, RTE, and ETE for different moving windows

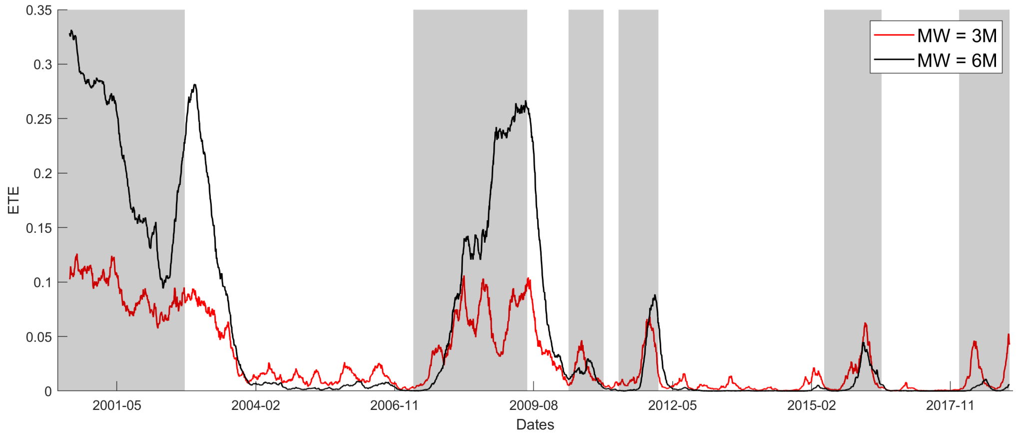
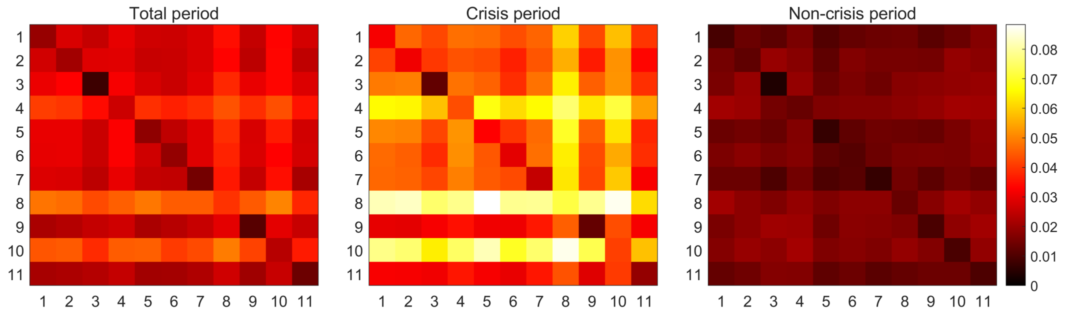


Figure 3.4: Evolution of average ETE in 3M and 6M moving windows

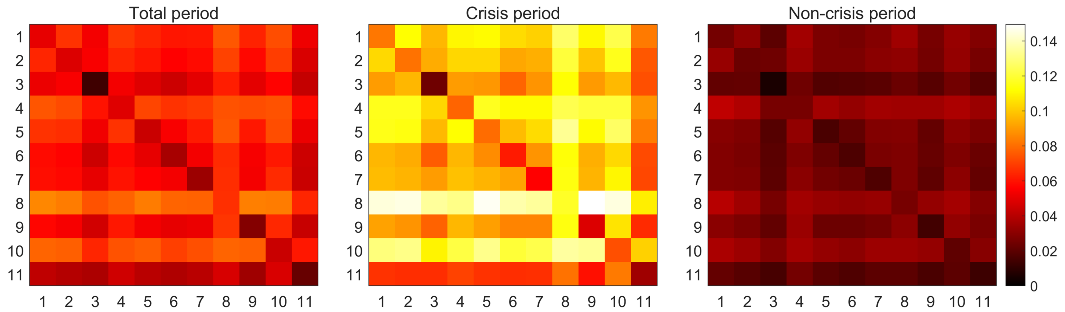
3.3.2 Sector analysis of Effective transfer entropy

Furthermore, we plot the average ETE of a sector for each financial crisis period as well as the total, crisis and non-crisis periods in Figure 3.5 to Figure 3.7 as a heatmap to investigate the information flow between individual sectors. Note that the sector ETE is defined as the average of ETEs of stocks associated in the sector. Comparing the heatmaps of the total, crisis, and non-crisis periods in the top of Figure 3.5a and Figure 3.5b, the information flow between sectors in the crisis period is significantly higher than that of non-crisis period. Note that the brighter the color, the stronger the information transfer.

In Figure 3.6 and Figure 3.7, from the perspective of the heatmap of each crisis and non-crisis period, the average of ETEs in each crisis periods are higher than that of each non-crisis periods, same as the characteristics from the previous heatmap analysis for overall crisis and non-crisis periods. In addition, the sectors that have a statistical explanatory power on other sectors are identified for each period. In the case of heatmaps of 3M moving windows in Figure 3.6, CONS and



(a) 3M moving windows ETE



(b) 6M moving windows ETE

Figure 3.5: Heatmaps of the average ETE in overall periods

*Note: Numbers from 1 to 11 sequentially represent the TELE, COND, ENRG, HLCA, INDS, MTRS, FINC, CONS, INFT, UTIL, REES, respectively.

UTIL show a large information flow in Dot-com bubble, Subprime mortgage crisis and US-China trade war. Also, FINC and REES in European crisis and US debt-ceiling crisis and ENRG in Brexit show relatively large information flow than other sectors. The information flow of non-crisis periods is small except for the non-crisis period between the Subprime mortgage crisis and the European crisis, but still exist the statistical explanatory power between sectors can be confirmed. The heatmaps of 6M moving windows ETE in Figure 3.7 also shows the characteristics of the heatmaps of 3M moving windows ETE, and the difference in statistical explanatory power between sectors can be more clearly distinguished due to the large moving window.

In the Dot-com bubble and Subprime mortgage crisis, which had a huge market impact, there is a lot of information exchange between all sectors. Like other correlation measures, the ETE has a limitation that cannot fully distinguish the correlation that increases due to the strong impact on the market as a whole, especially the market impact caused by exogenous variables. However, within this high correlation, the ETE has a difference that can confirm the statistical explanatory power of sectors by checking the direction of information flow between sectors.

Besides, we further investigate the evolutions of inflow and outflow ETE of different sectors for 3M and 6M moving windows in Figure 3.8 and 3.9. Note that Total indicates the average of all sectors, which equates the evolutions of inflow and outflow. In both 3M and 6M moving windows ETE, the inflow and outflow ETE of the financial crisis period were observed to be larger in all sectors than the non-financial crisis period. Although ETEs of two moving windows generally show the same trend, the difference between inflow and outflow ETE is prominent in 3M moving windows. However, a more stable rise and fall of ETE is observed in 6M than 3M moving windows.

Also, in order to check the strength of the information flow of each sector, we plot the values of outflow minus inflow in Figure 3.10 and 3.11. Note that a positive value indicates a more substantial outgoing information transfer from a specific sector to other sectors that incoming information transfer and vice versa. At first, the difference between inflow and outflow for total is zero as shown in Figure 3.8 and 3.9. In the case of 3M moving windows in Figure 3.10, TELE, COND, and INFT show the most positive values in the dot-com bubble located in the first gray

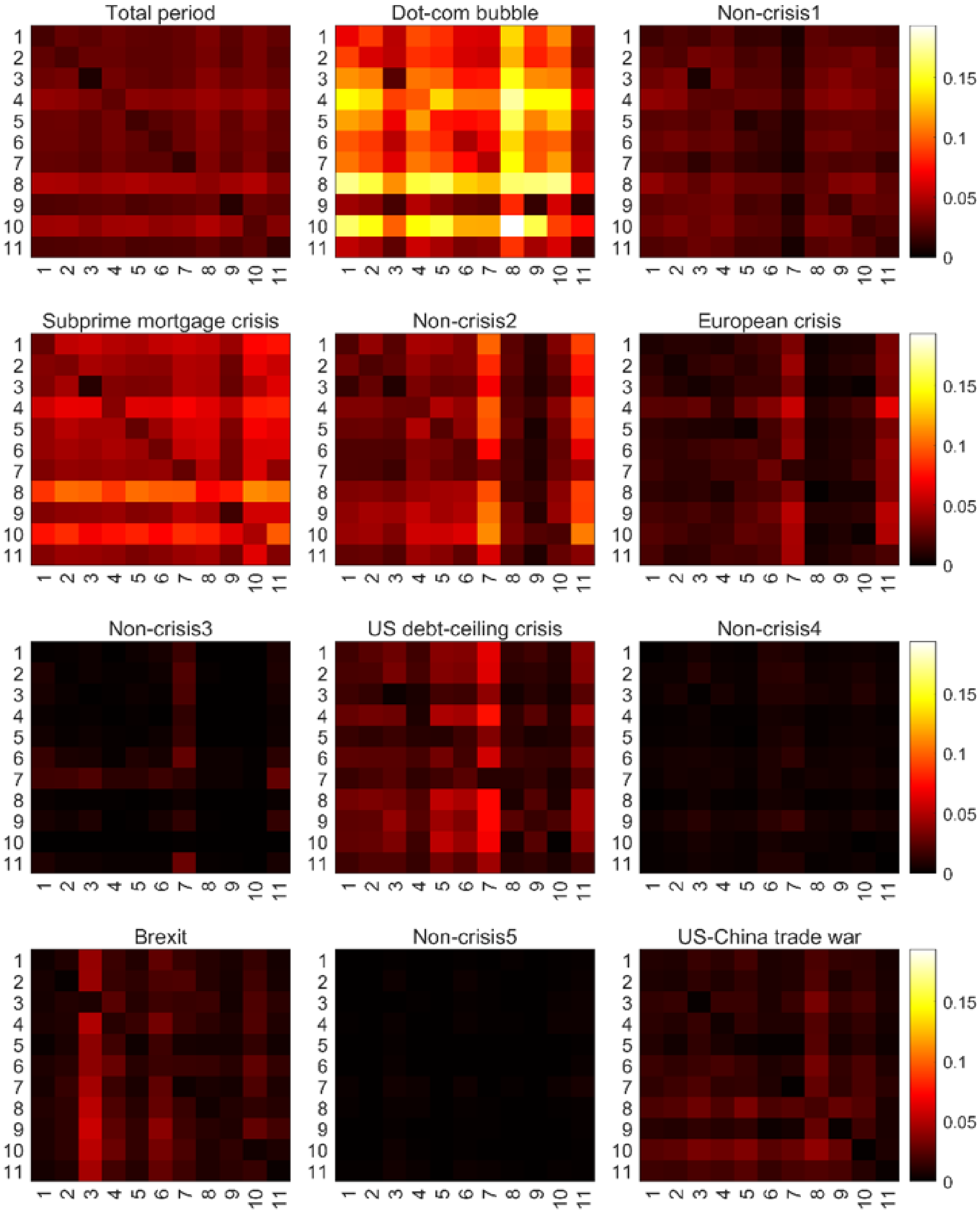


Figure 3.6: Heatmaps of the average ETE in different periods (3M)

*Note: Numbers from 1 to 11 sequentially represent the TELE, COND, ENRG, HLCA, INDS, MTRS, FINC, CONS, INFT, UTIL, REES, respectively.

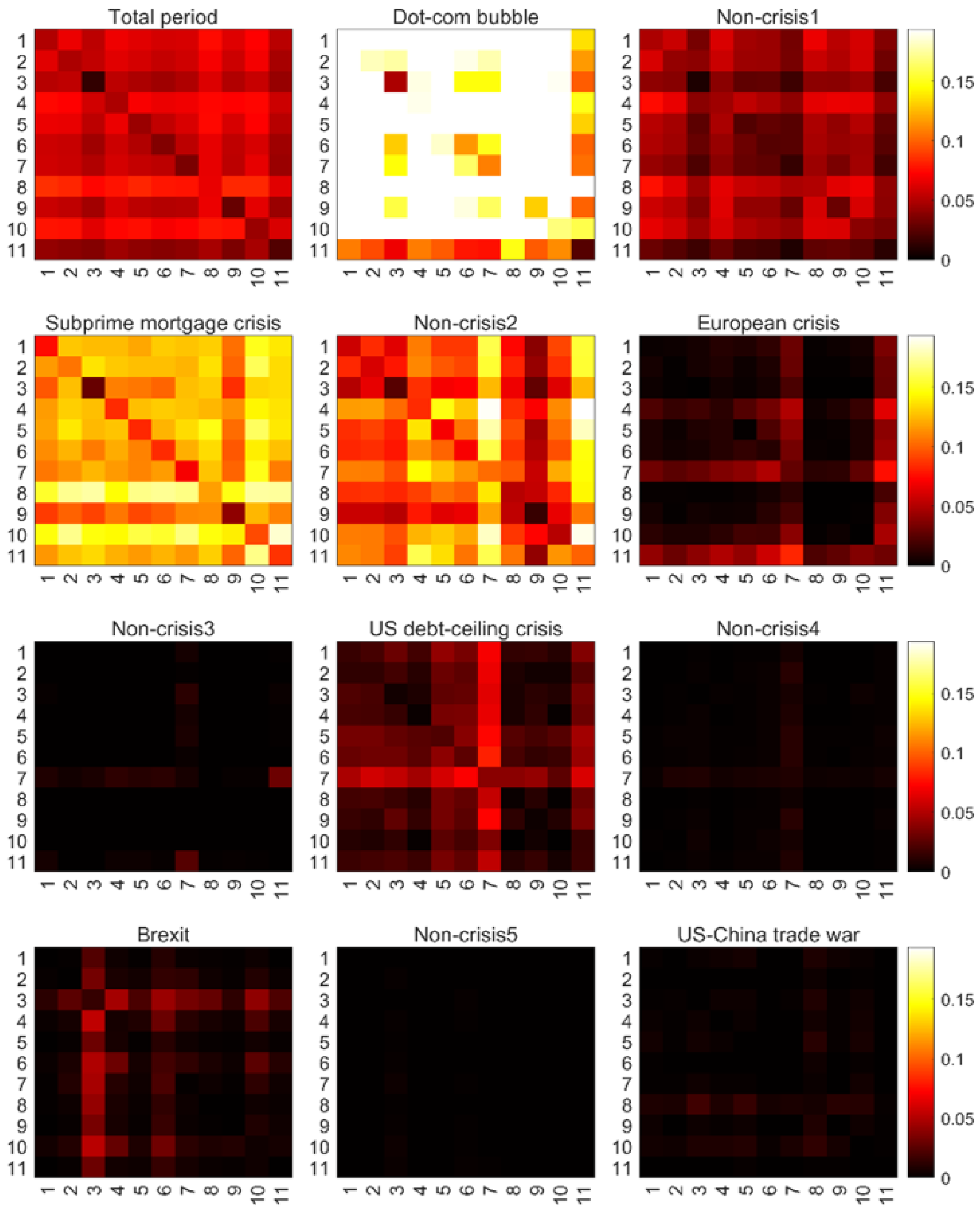


Figure 3.7: Heatmaps of the average ETE in different periods (6M)

*Note: Numbers from 1 to 11 sequentially represent the TELE, COND, ENRG, HLCA, INDS, MTRS, FINC, CONS, INFT, UTIL, REES, respectively.

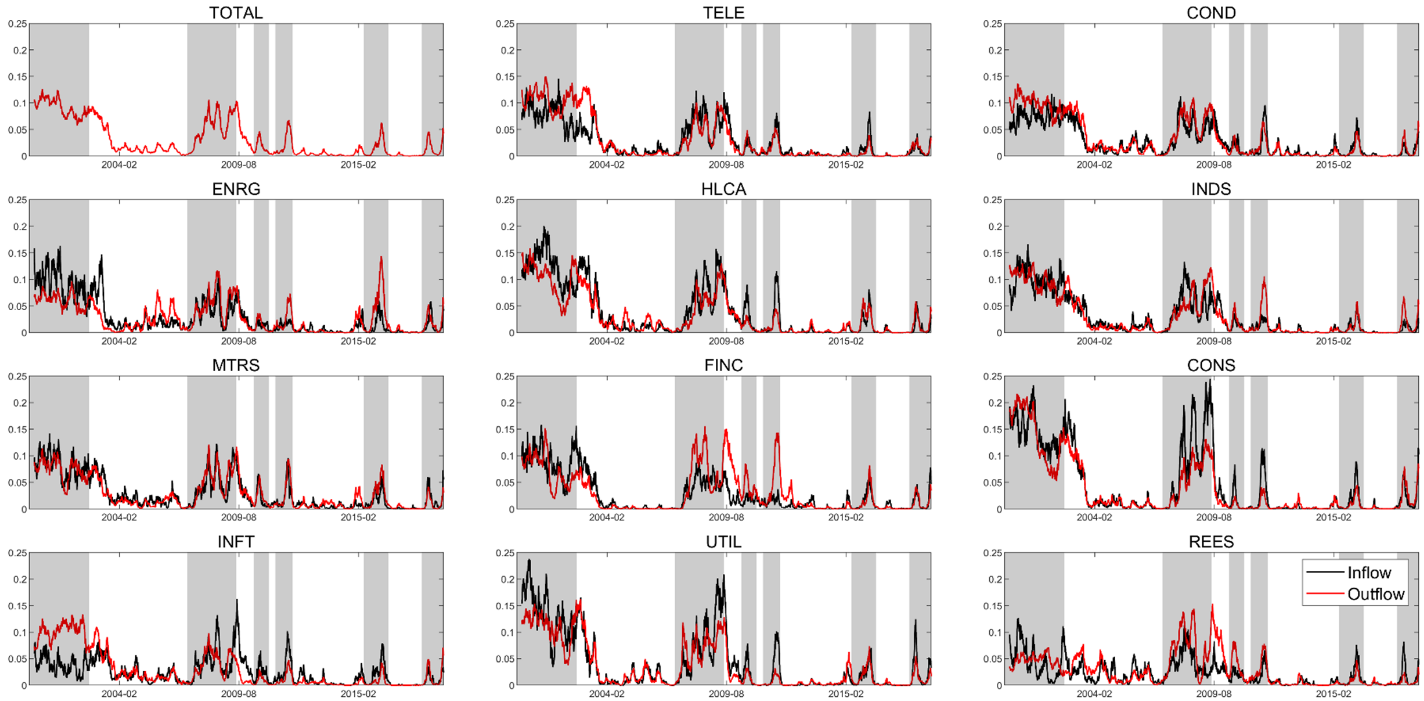


Figure 3.8: Evolution of inflow and outflow ETE in different sectors (3M)

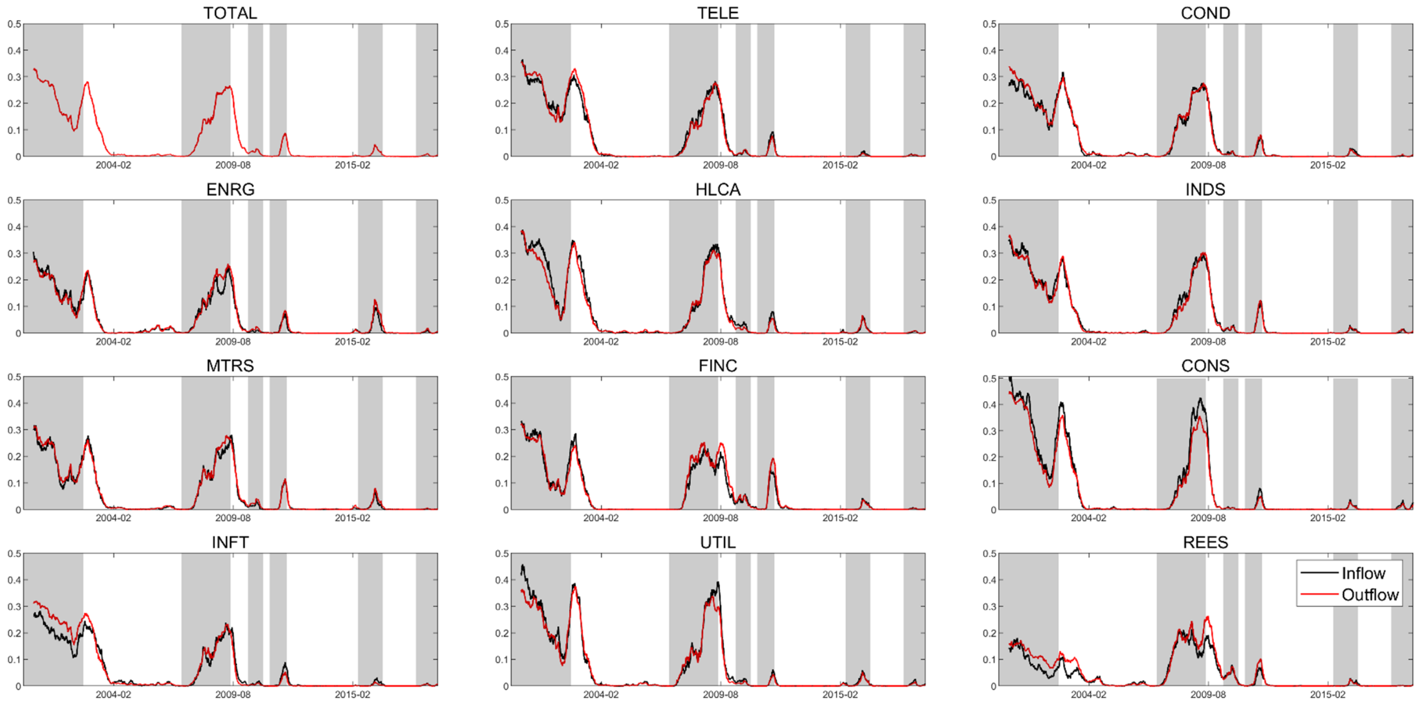


Figure 3.9: Evolution of inflow and outflow ETE in different sectors (6M)

region from the left. In particular, INFT shows the most persistent and large positive values. On the contrary, in the same dot-com bubble period, ENRG, HLCA, and UTIL exhibit negative values most time.

Given that the period is a Dot-com bubble, the strong outgoing information transfer from the INFT is because the period is closely related to the bubble economy of information technology companies. In the Subprime mortgage and European crises, FINC and REES show the most positive values, whereas CONS, INFT, and UTIL show the most negative values. Again, FINC and REES are directly related to the subprime mortgage crisis, and the FINC is the sector most affected by the European crisis.

Similar to 3M moving windows ETE, 6M moving windows ETE in Figure 3.11 shows different directions and intensity of information transfer for each sector. In conclusion, it is confirmed that outflow and inflow ETEs show the characteristics of each sector according to a different time, and such information could be useful in predicting the direction of future stock prices.

3.4 Summary and Discussion

In this chapter, the US market is analyzed based on ETE. A total of 55 stocks in the US market, the top five stocks by market capitalization in each sector are utilized. We define six major events of the US market during the data period and derive ETE using log-returns of stocks, and then conduct time-varying analysis and cross-sectional analysis to verify the information flow between sectors.

At first, the appropriate moving windows ETE describing the US market is

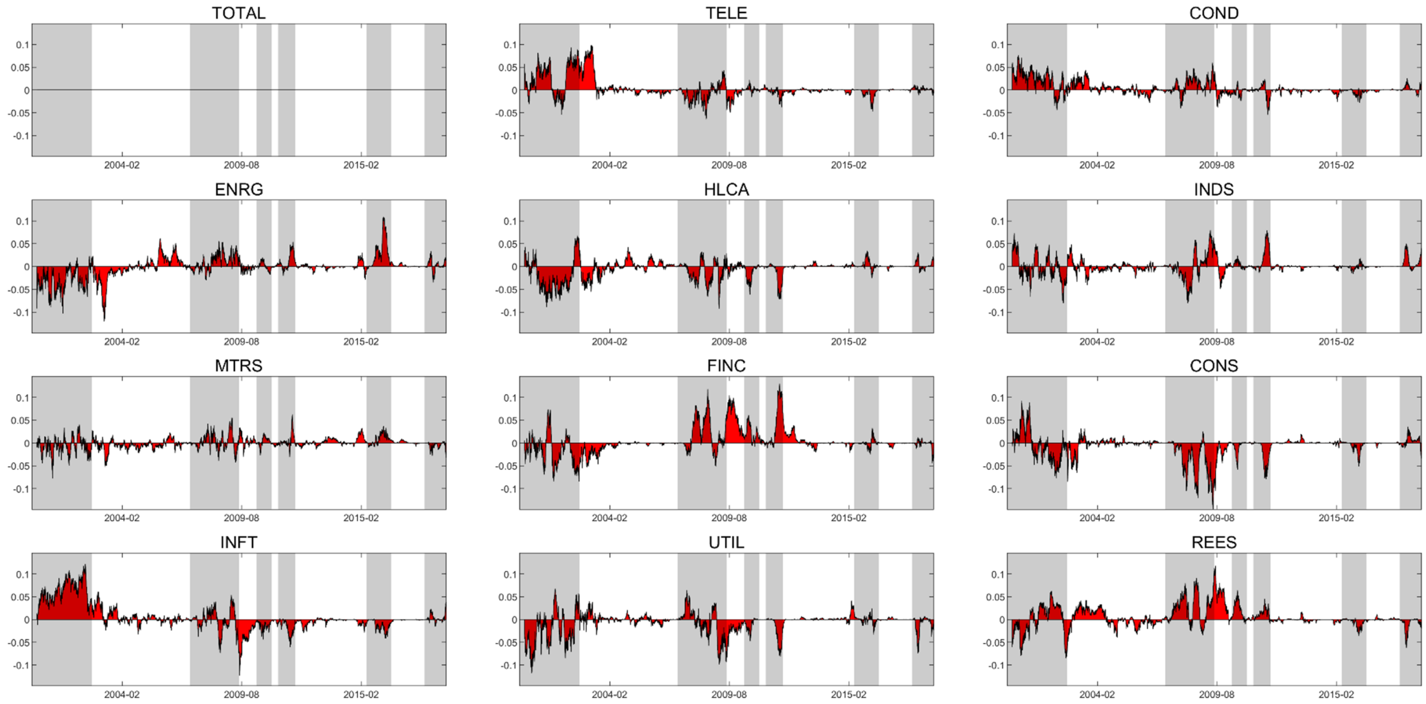


Figure 3.10: Evolution of ‘outflow - inflow’ ETE in different sectors (3M)

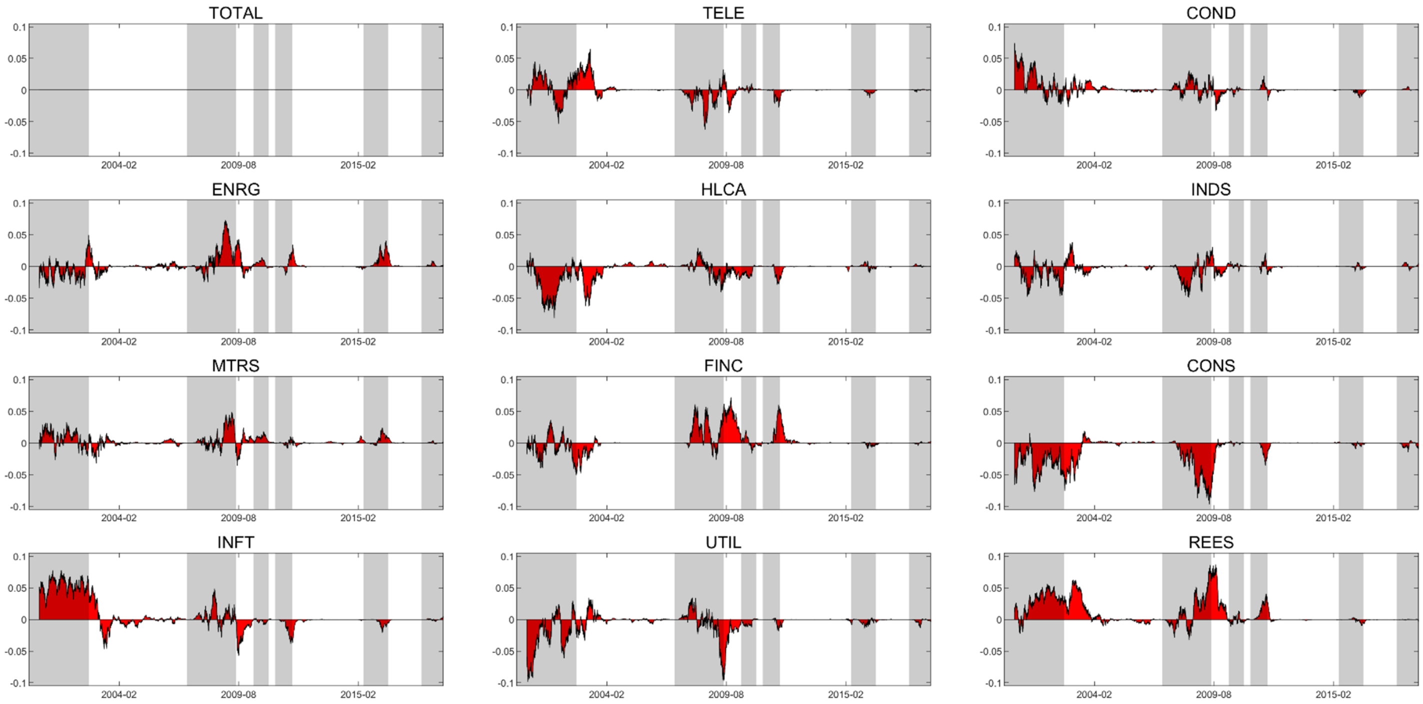


Figure 3.11: Evolution of ‘outflow - inflow’ ETE in different sectors (6M)

searched. For 1M, 3M, 6M, and 1Y moving windows ETE, 3M and 6M moving windows ETE show a tendency to rise during the financial crisis event and confirm that these ETEs have market explanatory power.

Secondly, cross-section analysis of ETE is conducted. For 3M and 6M moving windows ETE, through the heatmap of the average ETE of a sector, the information flow between sectors in the crisis period is significantly higher than that of non-crisis period. The ETE has a limitation that cannot fully distinguish the increased correlation due to strong influence on the market, but it is possible to distinguish information flow between sectors even within such high correlation.

Lastly, a time-varying analysis of ETE is conducted. In both 3M and 6M moving windows ETE, the inflow and outflow ETE of the crisis period are observed to be larger in all sectors than the non-crisis period. In addition, the analysis of the value of outflow minus inflow ETE analysis for verifying the strength of the information flow of each sector is conducted and confirms that the sectors associated with the crisis show the most positive values. It can be interpreted that the sectors related to the crisis in the crisis period export information and have a statistical explanatory power to other sectors.

In conclusion, it is confirmed that outflow and inflow ETEs show the characteristics of each sector according to a different time, and such information could be useful in predicting the direction of future stock prices. After all, the inflow and outflow ETEs are defined as ETE network indicators, the application of the ETE network indicators as variables to the stock price direction prediction using machine learning algorithms will be discussed in Chapter 4.

Chapter 4

Predicting the direction of US stock prices using ETE and machine learning techniques

4.1 Machine learning algorithms

4.1.1 Logistic regression

LR(Cox, 1958) is a multivariate analysis model used to predict the likelihood of an event using a linear combination of independent variables. In general, the LR is useful when the dependent variable is a binary or a multinomial categorical variable and is mainly used in the financial sector to predict the direction of stock prices or analyze the classification among companies. In this study, the LR was applied as a model representing a linear model, and the stock price direction of the cumulative log-returns was classified based on a cut-off value of 0.5.

4.1.2 Multi-layer perceptron

MLP(Rosenblatt, 1961) is a kind of feed-forward artificial neural network and consists of an input layer, a hidden layer, and an output layer. Each node, except the input nodes, is a neuron that uses a nonlinear activation function and is the most common structure of a neural network learning through back-propagation. In this study, we set the each hidden layers' neurons as learning parameters of the MLP. Note that the number of hidden layers is composed of four. Then, the structures of MLP were defined as follows. The number of neurons for each layer is $\{256, 128, 64, 32\}$, $\{128, 64, 32, 16\}$, $\{64, 32, 16, 8\}$, $\{32, 16, 8, 4\}$, $\{16, 8, 4, 2\}$, which yields five sets of hidden layers. In summary, a total of five parameter sets (5 neurons) are used and the ReLU was applied to the activation function.

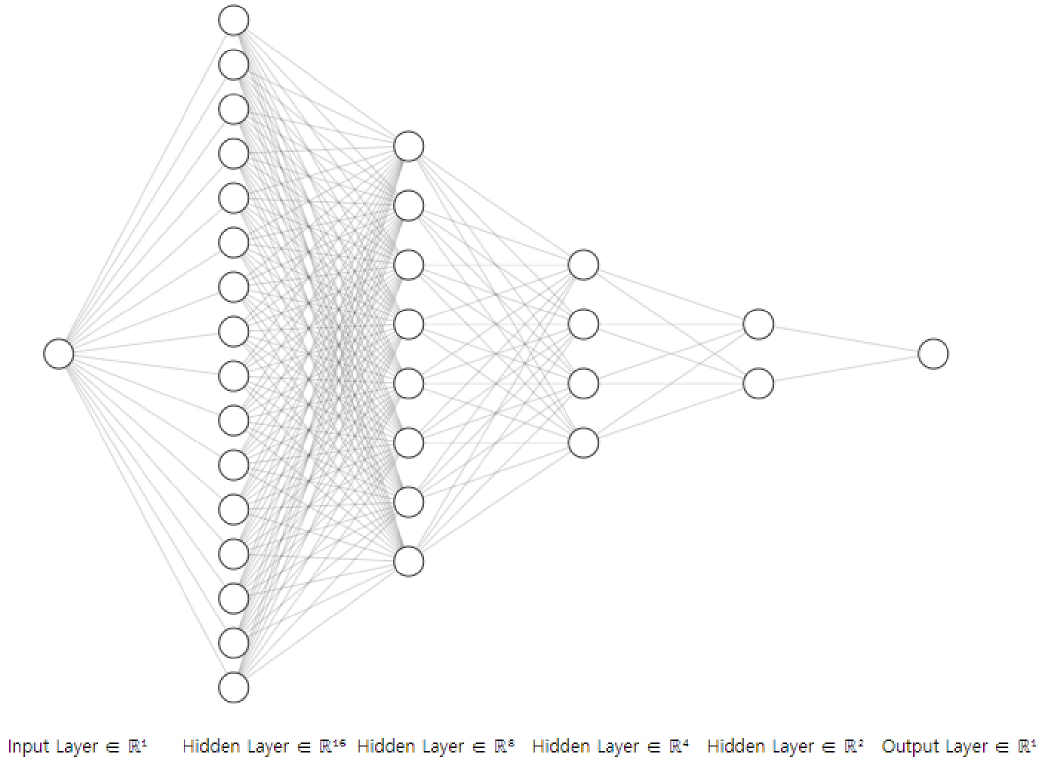


Figure 4.1: Structure of multilayer perceptron (4-hidden layers for $\{16, 8, 4, 2\}$)

4.1.3 Random forest

RF(Breiman, 2001), a kind of ensemble learning method used for classification and regression analysis, is a method to obtain a classifier with high accuracy and stability by generating a multitude of decision trees and combining each predictive value.

Since the RF generates multiple decision trees through bagging and learns by selecting variables randomly for each tree, it is robust against noise and outliers as well as overcomes the disadvantage that the existing single decision tree tends to incur over-fitting according to the training set. Although the training and test time increases as the number of trees increases, a large forest exhibits a relatively continuous result and better generalization ability than a small forest, which increases the stability of the model. In this study, we set the maximum depth of the trees as learning parameter of the RF where the numbers of trees are 200, and the depth is set to 12 to 24 by three intervals. In summary, a total of five parameter sets (5 depths) are used.

4.1.4 Extreme gradient boosting

XGBoost(XGB)(Chen and Guestrin, 2016) is a gradient boosting algorithm that emphasizes parallel processing and optimization. The gradient boosting machine (GBM) method can be defined as follows. After constructing weak learners, the training set and consistency are evaluated to construct new weak learners using the gradient descent method as the explanatory variable. By repeating such a process, several predictive models are ensembled to construct a strong learner. The GBMs have excellent prediction performance, but there might be an overfitting problem if the proper number of splits is not obtained due to no fitting limit of the number of splits in the decision tree. The XGB is an algorithm that modified the structure of GBM to prevent the overfitting problem through regularization and enabling of parallel computation. In this study, we set the maximum depth of the trees as learning parameter of the XGB. The number of trees is set to 200 same as RF, whereas the depth is set to 3 to 15 by three intervals. Therefore, a total of five parameter sets (5 depths) are utilized. The learning rate was set to 0.1.

4.1.5 Long short-term memory network

LSTM(Hochreiter and Schmidhuber, 1997), a kind of Recurrent Neural Network (RNN)(Rumelhart et al., 1986), is a suitable method for classification or prediction based on time series data with unknown gaps between essential events. A typical LSTM unit consists of a cell, input gate, output gate, and forget gate. The cell maintains dependencies between the elements of the input sequence and selectively adjusts the information flow through the gates. The back-propagated error values at the output layer of the unit remain in the LSTM unit's cell, and the error values are continuously supplied to the LSTM unit's gate to learn the cut-off value. Through this process, LSTM networks can deal with vanishing gradient problems that can occur when learning through traditional RNNs. The LSTM network architecture used in this dissertation is defined as follows. The input size is the numbers of features of the input data. The output is binary (up/down) data based on the cumulative log-return of stock during the prediction period. The layers consist of the LSTM layer first, followed by a fully connected layer of size 4, and a sigmoid layer. Specifically, we select the number of hidden units of LSTM layer as the learning parameter of LSTM networks. The number of hidden units are set to 8, 16, 32, 64, and 128, which yields a total of five parameter sets (5 hidden units).

The parameter pairs used in machine learnings are summarized in Table 4.1.

Table 4.1: Parameter set-ups for machine learning algorithms

Model	**Parameters**	**Levels**
MLP	Neuron of first layer	256, 128, 64, 32,16
RF	Depth of trees	12, 15, 18, 21, 24
XGB	Depth of trees	3, 6, 9, 12, 15
LSTM	Hidden unit	8, 16, 32, 64, 128

4.1.6 Adjusted accuracy

In finance, the performance of a portfolio is evaluated based on the risk-adjusted return, commonly known as the Sharpe ratio(Sharpe, 1994). Therefore, we also measure the prediction performance of each algorithm based on a concept similar to the Sharpe ratio. The Sharpe ratio measures the excess return per unit of deviation, usually referred to as risk, and is defined as,

$$S_a = \frac{E[R_a - R_f]}{\sigma_a} = \frac{E[R_a - R_f]}{\sqrt{var[R_a - R_f]}} \tag{4.1}$$

where R_a is the asset return, R_f is the risk-free return, and σ_a is the standard deviation of the asset excess return.

In this study, we propose an adjusted accuracy as a prediction performance measure on the direction of the stock price. Note that the previous studies mainly compare the absolute value of accuracy in evaluating the results. In case of adjusted accuracy, we define R_a, R_f and $\sqrt{var[R_a - R_f]}$ as the prediction accuracy of a single stock, the benchmark accuracy 0.5 from the expectation of binary prediction, and the standard deviation of $R_a - R_f$, respectively.

Through the adjusted accuracy of 55 stocks for each machine learning model and those of 5 stocks for each sector, the performance consistencies of model and sector are confirmed by considering performance distortion caused by outlier performances of specific stocks.

4.2 Data and experiment set-ups

4.2.1 Data

The data used in this chapter are the same as in Chapter 3.

4.2.2 Experiment set-ups

We construct the training dataset based on a various mixture of input and target variables. The input variable(x) at each time point for each stock consist of three cases in two variables: ETE network indicators for inflow and outflow and lags for 3 months (3M, 60 days), 6 months (6M, 120 days), and 1 year (1Y, 240 days) of time series of each stock's past log-returns. Then, the target variable(y) is the cumulative log-returns in different prediction periods for 1 week (1W, 5 days), 1 month (1M, 20 days), and 3 months (3M, 60 days).

Note that the categorized (positive & negative) cumulative log-returns are used for algorithms. This process is then repeated for the different ETEs obtained from different sizes of moving windows. Figure 4.2 summarizes the dataset structure used in this study.

For each dataset, we compare the prediction performances of the mixture of log-returns and ETE network indicator against that of plain log-returns. In this regard, we aim to check the following statements in terms of prediction performance: (1) the validity of ETE network indicators and (2) the most applicable machine learning algorithm. Specifically, the learning is performed as follows. At first, the entire data is divided into 50% of training and 50% of test sets. Then, we normalize the stock log-returns and inflow and outflow ETEs based on their means and standard deviations of each training set. The associated periods of training set for 3M and 6M moving

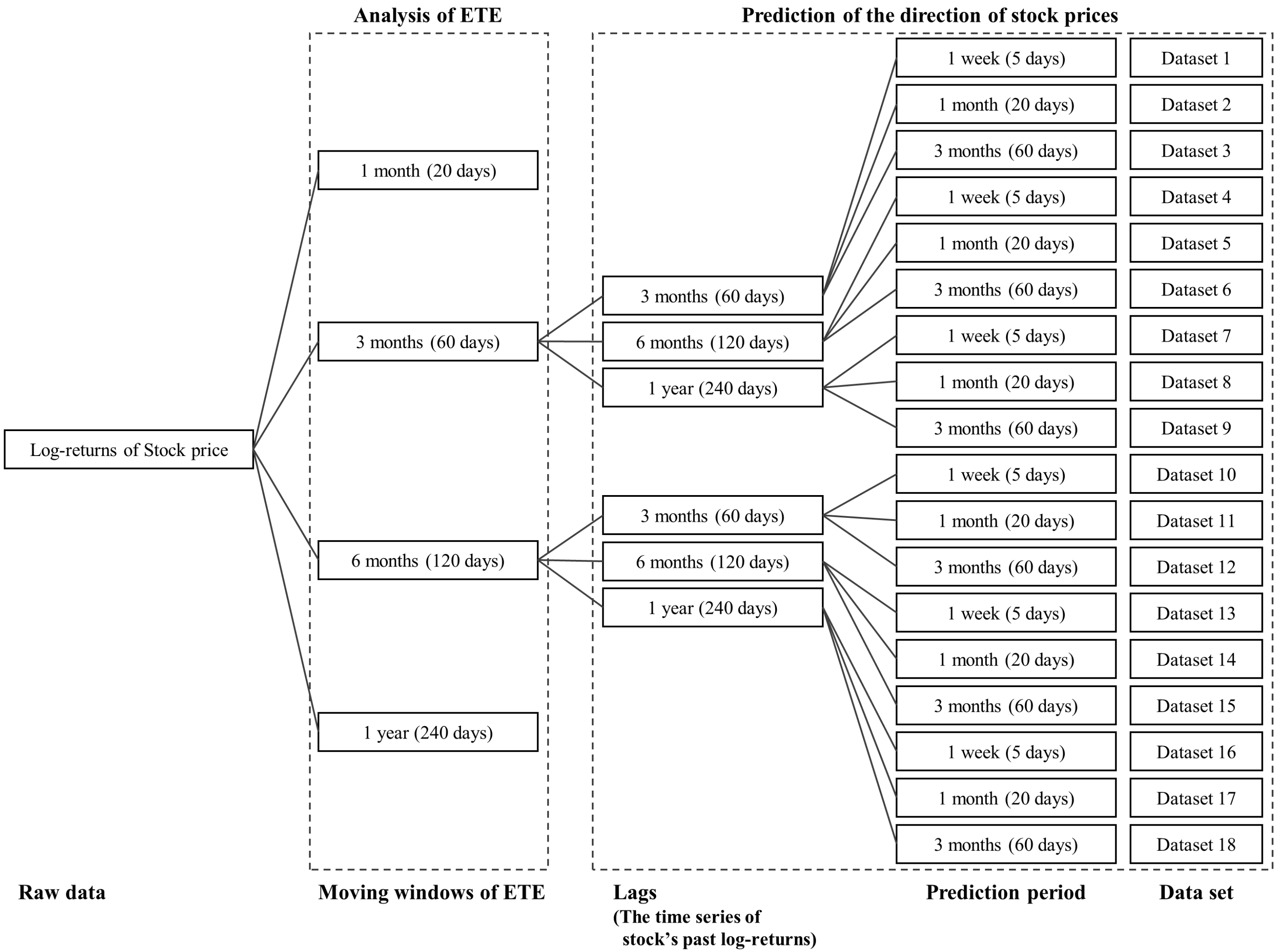
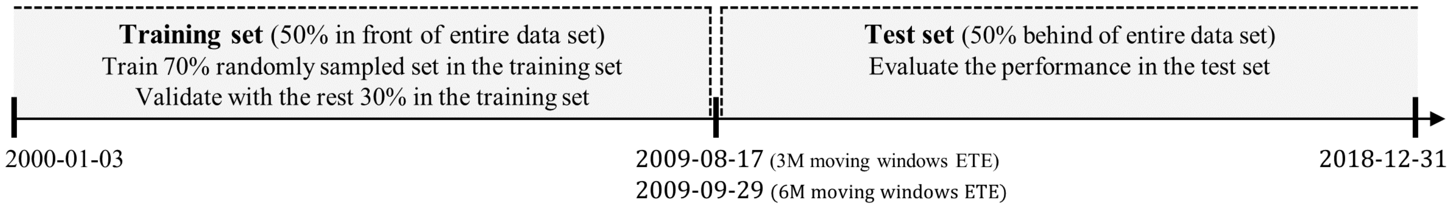


Figure 4.2: Dataset structure used to compute the ETE and to perform the prediction

windows are from 2000-01-03 to 2009-08-17 and from 2000-01-03 to 2009-09-29, respectively. Note that we only train 70% randomly sampled set from the training set that maintains a dataset consisting of the lags of time series of each stock's past log-returns and the ETE network indicators at that time to obtain the generalized training results. In addition, the hyperparameter tuning of each model is performed through the validation process for the remaining 30% of the training set. Lastly, we compare the prediction performance of test set for 3M and 6M, whose periods from 2009-08-18 to 2018-12-31 and from 2009-09-30 to 2018-12-31, respectively. Figure 4.3 shows an example of the proposed machine learning framework.

We define a prediction accuracy to reflect a proportion of correctly classified direction of the cumulative log-returns in the test set. Based on the prediction accuracy, we can define the prediction performance of each machine learning algorithm. For LR, the performance of the algorithm is the same as the prediction accuracy since the LR does not possess any model parameters. In contrast, for MLP, RF, XGB, and LSTM, we set five parameter sets, which yield five different results of prediction accuracy.

Based on the previous studies(Patel et al., 2015; Lee et al., 2019), the prediction accuracy of LR for a single stock is the same as the accuracy of LR, whereas those of other four models are defined as the averages of top three results in prediction accuracy among the five parameter sets for measuring the average performance of machine learning models. In contrast, a model-specific prediction accuracy for each sector is defined as the average of the prediction accuracy of stocks associated in each sector. The same approach is used to define the overall performance of five models. We calculate the averages of the prediction accuracy of all stocks. In case



Example of training set (3M lag, 1W prediction period)	
Without ETE network indicator (only plain log-returns)	$[x = (r_{t_0}, r_{t_0+1}, \ldots, r_{t_0+59}), y = \frac{r_{t_{0+64}} - r_{t_{0+60}}}{r_{t_{0+60}}}]$ $\cdots$ $[x = (r_{t_1-63}, r_{t_1-62}, \ldots, r_{t_1-4}), y = \frac{r_{t_1} - r_{t_{1-4}}}{r_{t_1}}]$
With ETE network indicator	$[x = (r_{t_0}, r_{t_0+1}, \ldots, r_{t_0+59}, \boldsymbol{ETE^{in}}_{\boldsymbol{t_0+59}}, \boldsymbol{ETE^{out}}_{\boldsymbol{t_0+59}},), y = \frac{r_{t_{0+64}} - r_{t_{0+60}}}{r_{t_{0+60}}}]$ $\cdots$ $[x = (r_{t_1-63}, r_{t_1-62}, \ldots, r_{t_1-4}, \boldsymbol{ETE^{in}}_{\boldsymbol{t_1-4}}, \boldsymbol{ETE^{out}}_{\boldsymbol{t_1-4}}), y = \frac{r_{t_1} - r_{t_{1-4}}}{r_{t_1}}]$

Figure 4.3: A scenario of the proposed machine learning framework

of the improvement of prediction accuracy incurred by ETE network indicator, the LR is defined as the simple difference of each stock, and the MLP, RF, XGB and LSTM are defined as the average of the top three among the accuracy differences between parameter sets. Furthermore, the same approach is applied to the adjusted accuracy, which can be obtained by (4.1).

In this study, we evaluate the stock price direction prediction performance using adjusted accuracy. Note that the adjusted accuracy focuses on the consistency of the accuracy of the overall stocks and sectors and the improvement focuses on identifying the influence of the ETE network indicator between the same parameter sets of each model. Then, the prediction process is repeated 10 times to check the average and standard deviation of defined adjusted accuracy. Finally, the overall framework can be summarized as a step-by-step procedure described in Figure 4.4.

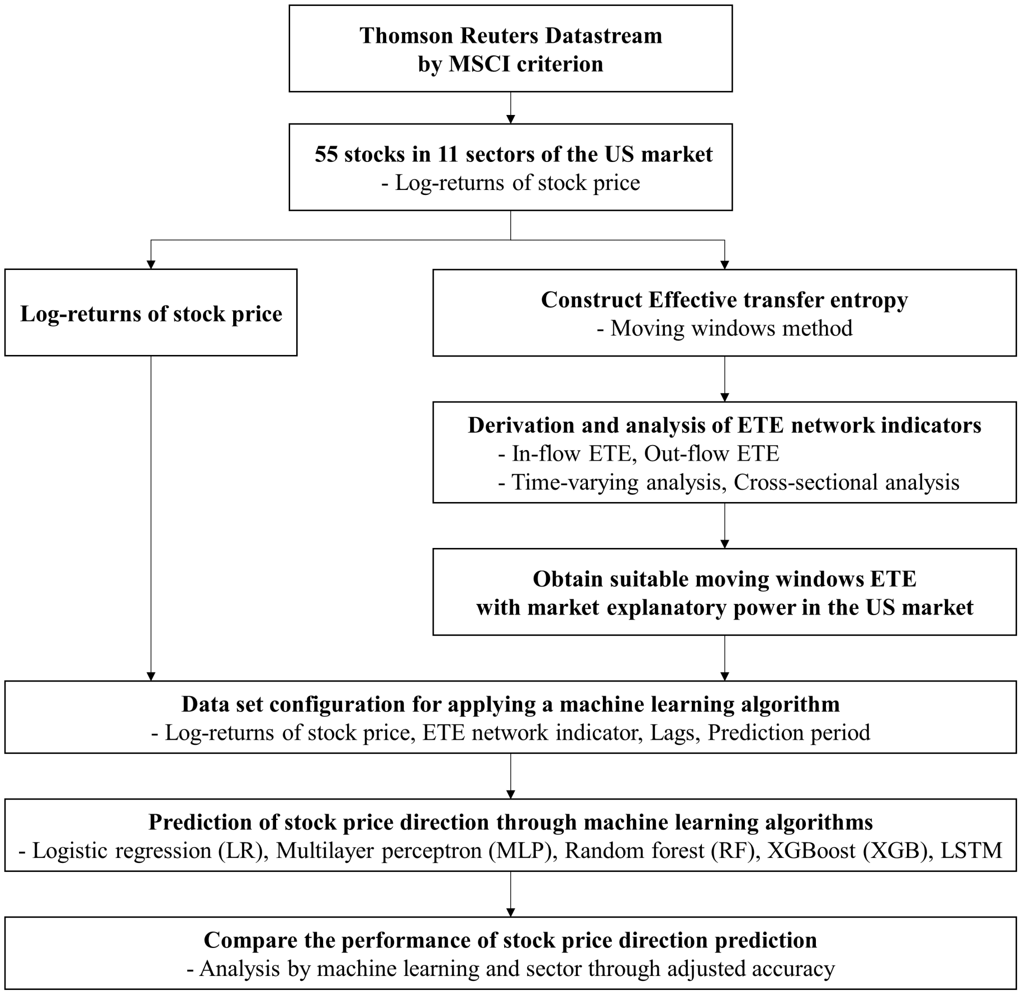


Figure 4.4: Procedure for the stock price prediction using the ETE

4.3 Results

4.3.1 Prediction performance in different models

We intend to utilize the evolution of ETE, which has a market explanatory power, as an input variable in predicting the direction of future stock price based on the LR, MLP, RF, XGB and LSTM. Especially, the prediction performance of the model using the inflow and outflow ETE values, which can be defined as ETE network indicators, as input variables are compared with that of the model without the ETE values. Table 4.2 summarizes the average prediction adjusted accuracy and their standard deviation of all models for different moving windows, lags, and prediction periods.

The results show that there is no significant difference in prediction adjusted accuracy among different moving windows. Also, there is no significant difference according to the lag, whereas the adjusted accuracy for the 6M lag is higher than that of others and the adjusted accuracy tends to be higher in the shorter lags. More importantly, the improvement of prediction performance by utilizing the ETE network indicator is detected in all parameter sets. The improvement tends to decrease as the prediction period increases and increase as the lag increases. Therefore, we suggest that the ETE network indicators can be used to improve the performance of the stock price forecast in the US market.

Table 4.3 to 4.7 for adjusted accuracy summarize prediction performances for different machine learning algorithms, moving windows, lags, prediction periods, and integration of ETE indicator.

From the perspective of tendency of adjusted accuracy, there is little difference

Table 4.2: Overall adjusted accuracies in different dataset structure

MW	Lag		With ETE			Without ETE			Improvement		
			1W	1M	3M	1W	1M	3M	1W	1M	3M
3M	3M	Mean	1.32	1.27	1.13	1.25	1.26	1.06	0.58	0.36	0.32
		Std	0.051	0.032	0.024	0.028	0.023	0.014	0.051	0.045	0.035
	6M	Mean	1.33	1.33	1.19	1.25	1.29	1.12	0.59	0.44	0.41
		Std	0.022	0.036	0.023	0.058	0.036	0.015	0.041	0.047	0.029
	1Y	Mean	1.25	1.27	1.14	1.19	1.26	1.07	0.6	0.48	0.43
		Std	0.045	0.039	0.02	0.032	0.037	0.038	0.048	0.049	0.035
6M	3M	Mean	1.37	1.35	1.16	1.3	1.35	1.13	0.65	0.43	0.25
		Std	0.058	0.03	0.038	0.042	0.031	0.026	0.029	0.021	0.035
	6M	Mean	1.41	1.4	1.17	1.31	1.39	1.17	0.66	0.51	0.3
		Std	0.031	0.035	0.027	0.026	0.037	0.016	0.034	0.029	0.01
	1Y	Mean	1.29	1.3	1.13	1.15	1.24	1.12	0.66	0.53	0.37
		Std	0.06	0.02	0.031	0.026	0.029	0.014	0.049	0.028	0.022

*Note: MW, With ETE, Without ETE, and Improvement are the abbreviations of the moving window, the performance when predicted by the mixture of log-returns and the ETE network indicator, the performance when predicted only by its plain log-returns, improvement of prediction adjusted accuracy incurred by ETE network indicator, the mean and standard deviation of the adjusted accuracy performed 10 times, respectively.

Table 4.3: Adjusted accuracy for different machine learning algorithms (LR)

Model	**MW**	**Lag**		**With ETE**			**Without ETE**			**Improvement**		
				1W	1M	3M	1W	1M	3M	1W	1M	3M
LR	3M	3M	Mean	1.17	1.18	1.12	0.96	1.04	0.95	0.48	0.17	0.08
			Std	0.07	0.041	0.02	0.069	0.043	0.019	0.075	0.05	0.024
		6M	Mean	1.17	1.17	1.1	0.98	1.08	0.97	0.51	0.26	0.21
			Std	0.074	0.036	0.027	0.077	0.05	0.025	0.059	0.063	0.03
		1Y	Mean	1.02	1	0.95	0.9	0.94	0.83	0.55	0.33	0.44
			Std	0.072	0.039	0.022	0.122	0.036	0.023	0.164	0.062	0.041
	6M	3M	Mean	1.18	1.23	1.1	1.04	1.17	1.05	0.58	0.21	-0.01
			Std	0.094	0.046	0.027	0.087	0.045	0.015	0.057	0.03	0.028
		6M	Mean	1.23	1.33	1.06	1.03	1.22	1.03	0.63	0.55	0.12
			Std	0.085	0.032	0.016	0.083	0.041	0.013	0.076	0.057	0.03
		1Y	Mean	1.04	1.02	0.93	0.9	0.99	0.86	0.63	0.48	0.38
			Std	0.111	0.035	0.016	0.106	0.036	0.019	0.093	0.031	0.021

*Note: MW, With ETE, Without ETE, Improvement, Mean, Std are the abbreviations of the moving window, the performance when predicted by the mixture of log-returns and the ETE network indicator, the performance when predicted only by its plain log-returns, the improvement of prediction adjusted accuracy incurred by ETE network indicator, the mean and standard deviation of the adjusted accuracy performed 10 times, respectively.

Table 4.4: Adjusted accuracy for different machine learning algorithms (MLP)

Model	**MW**	**Lag**		**With ETE**			**Without ETE**			**Improvement**		
				1W	1M	3M	1W	1M	3M	1W	1M	3M
MLP	3M	3M	Mean	1.69	1.63	1.37	1.67	1.59	1.22	1.01	0.83	0.72
			Std	0.082	0.088	0.093	0.151	0.123	0.059	0.129	0.147	0.066
		6M	Mean	1.67	1.73	1.37	1.5	1.69	1.31	1.04	0.93	0.83
			Std	0.077	0.119	0.078	0.119	0.13	0.081	0.133	0.077	0.075
		1Y	Mean	1.63	1.83	1.38	1.59	1.73	1.37	1.01	0.96	0.81
			Std	0.139	0.122	0.098	0.097	0.089	0.114	0.144	0.115	0.087
	6M	3M	Mean	1.68	1.66	1.35	1.68	1.72	1.33	1.04	0.84	0.75
			Std	0.141	0.124	0.109	0.166	0.185	0.093	0.165	0.098	0.1
		6M	Mean	1.71	1.72	1.38	1.69	1.71	1.37	1.07	0.92	0.85
			Std	0.154	0.086	0.123	0.057	0.101	0.122	0.08	0.071	0.094
		1Y	Mean	1.73	1.74	1.46	1.58	1.68	1.48	1.14	0.94	0.76
			Std	0.139	0.11	0.101	0.122	0.156	0.08	0.154	0.088	0.068

*Note: MW, With ETE, Without ETE, Improvement, Mean, Std are the abbreviations of the moving window, the performance when predicted by the mixture of log-returns and the ETE network indicator, the performance when predicted only by its plain log-returns, the improvement of prediction adjusted accuracy incurred by ETE network indicator, the mean and standard deviation of the adjusted accuracy performed 10 times, respectively.

Table 4.5: Adjusted accuracy for different machine learning algorithms (RF)

Model	**MW**	**Lag**		**With ETE**			**Without ETE**			**Improvement**		
				1W	1M	3M	1W	1M	3M	1W	1M	3M
RF	3M	3M	Mean	1.19	1.18	1	1.25	1.16	0.91	0.32	0.05	0.07
			Std	0.08	0.03	0.031	0.057	0.029	0.018	0.122	0.021	0.034
		6M	Mean	1.36	1.27	1.07	1.38	1.28	1.01	0.49	0.05	0.13
			Std	0.081	0.039	0.031	0.078	0.047	0.026	0.131	0.054	0.042
		1Y	Mean	1.22	1.21	1.03	1.22	1.19	0.97	0.84	0.18	0.26
			Std	0.061	0.044	0.023	0.059	0.045	0.021	0.213	0.039	0.04
	6M	3M	Mean	1.37	1.21	1.05	1.31	1.24	0.98	0.97	0.04	-0.03
			Std	0.071	0.033	0.028	0.089	0.036	0.022	0.091	0.044	0.02
		6M	Mean	1.43	1.3	1.1	1.37	1.31	1.04	1.06	0.19	-0.01
			Std	0.037	0.035	0.033	0.06	0.041	0.032	0.136	0.055	0.032
		1Y	Mean	1.24	1.18	1.06	1.18	1.16	1	1.16	0.47	0.22
			Std	0.079	0.055	0.036	0.053	0.051	0.029	0.083	0.101	0.031

*Note: MW, With ETE, Without ETE, Improvement, Mean, Std are the abbreviations of the moving window, the performance when predicted by the mixture of log-returns and the ETE network indicator, the performance when predicted only by its plain log-returns, the improvement of prediction adjusted accuracy incurred by ETE network indicator, the mean and standard deviation of the adjusted accuracy performed 10 times, respectively.

Table 4.6: Adjusted accuracy for different machine learning algorithms (XGB)

Model	**MW**	**Lag**		**With ETE**			**Without ETE**			**Improvement**		
				1W	1M	3M	1W	1M	3M	1W	1M	3M
XGB	3M	3M	Mean	1.27	1.26	1.12	1.32	1.39	1.21	0.41	0.04	-0.09
			Std	0.086	0.05	0.056	0.082	0.066	0.051	0.138	0.057	0.058
		6M	Mean	1.28	1.24	1.16	1.38	1.36	1.17	0.4	0.08	-0.03
			Std	0.068	0.069	0.067	0.086	0.053	0.036	0.072	0.068	0.056
		1Y	Mean	1.23	1.2	1.19	1.29	1.39	1.14	0.37	0.18	0.1
			Std	0.081	0.063	0.042	0.104	0.094	0.034	0.105	0.086	0.042
	6M	3M	Mean	1.33	1.32	1.1	1.4	1.4	1.21	0.53	0.08	-0.1
			Std	0.133	0.045	0.046	0.152	0.062	0.058	0.104	0.047	0.046
		6M	Mean	1.37	1.32	1.13	1.4	1.4	1.21	0.54	0.16	-0.07
			Std	0.081	0.102	0.053	0.105	0.05	0.022	0.152	0.057	0.049
		1Y	Mean	1.29	1.31	1.12	1.28	1.31	1.18	0.47	0.26	0
			Std	0.143	0.066	0.031	0.097	0.052	0.024	0.11	0.048	0.04

*Note: MW, With ETE, Without ETE, Improvement, Mean, Std are the abbreviations of the moving window, the performance when predicted by the mixture of log-returns and the ETE network indicator, the performance when predicted only by its plain log-returns, the improvement of prediction adjusted accuracy incurred by ETE network indicator, the mean and standard deviation of the adjusted accuracy performed 10 times, respectively.

Table 4.7: Adjusted accuracy for different machine learning algorithms (LSTM)

Model	**MW**	**Lag**		**With ETE**			**Without ETE**			**Improvement**		
				1W	1M	3M	1W	1M	3M	1W	1M	3M
LSTM	3M	3M	Mean	1.6	1.25	1.13	1.32	1.27	1.04	0.73	0.71	0.83
			Std	0.187	0.075	0.026	0.095	0.094	0.064	0.219	0.158	0.141
		6M	Mean	1.49	1.42	1.24	1.33	1.24	1.13	0.74	0.77	0.89
			Std	0.084	0.169	0.056	0.162	0.125	0.062	0.112	0.216	0.157
		1Y	Mean	1.63	1.45	1.24	1.42	1.45	1.13	0.86	0.69	0.64
			Std	0.13	0.148	0.082	0.104	0.099	0.098	0.18	0.156	0.072
	6M	3M	Mean	1.5	1.4	1.2	1.34	1.35	1.1	0.77	0.79	0.8
			Std	0.124	0.121	0.101	0.144	0.047	0.048	0.116	0.273	0.171
		6M	Mean	1.66	1.42	1.25	1.48	1.41	1.19	0.8	0.79	0.63
			Std	0.137	0.099	0.149	0.099	0.102	0.043	0.106	0.124	0.126
		1Y	Mean	1.62	1.58	1.17	1.29	1.44	1.18	0.85	0.69	0.65
			Std	0.109	0.078	0.061	0.155	0.089	0.091	0.137	0.116	0.139

*Note: MW, With ETE, Without ETE, Improvement, Mean, Std are the abbreviations of the moving window, the performance when predicted by the mixture of log-returns and the ETE network indicator, the performance when predicted only by its plain log-returns, the improvement of prediction adjusted accuracy incurred by ETE network indicator, the mean and standard deviation of the adjusted accuracy performed 10 times, respectively.

in adjusted accuracy between the moving windows as in overall performance. Note that the improvements by integrating ETE network indicators are detected in all models, whereas in LR, RF and XGB, there are some conditions in which the decreases in adjusted accuracy in the 3M prediction period. The improvement of LR and MLP tends to decrease as the prediction period increases, and LSTM shows a certain improvement regardless of the prediction period. The RF shows a good improvement in the 1W prediction period.

From the perspective of comparison between models, the adjusted accuracy of MLP and LSTM show higher than others, it implies that the prediction performance of MLP and LSTM is high and consistent. And the adjusted accuracy of LR shows that the prediction performance of the LR is not stable as those of other models.

In addition to tendency analysis of adjusted accuracy, Table 4.8 summarizes the result of a paired t-test to confirm that the adjusted accuracy improvement by the ETE network indicator has been statistically increased for each model, moving windows of ETE, lag and prediction period.

In MLP and LSTM, the null hypothesis of paired t-test is rejected under all conditions, the adjusted accuracy is statistically improved due to ETE network indicator. In RF and XGB, it can be seen that the adjusted accuracy is statistically improved in the 1W prediction period. In LR, the adjusted accuracy is statistically improved at a relatively low significance level compared to other models.

Overall, we discover that all five machine learning algorithms have improved the adjusted accuracy through the ETE network indicator and suggest that the MLP and LSTM are the most suitable models for predicting future stock price direction predictions when considered the adjusted accuracy and the paired t-test

Table 4.8: Paired t-test for different machine learnings

Model	**Lag**	**MW = 3M**			**MW = 6M**		
		1W	**1M**	**3M**	**1W**	**1M**	**3M**
LR	3M	0.0006	0.0154^{**}	0.0062^{*}	0.0171^{**}	0.3302^{***}	0.4283^{***}
	6M	0.0010	0.1488^{***}	0.0310^{**}	0.0009	0.0717^{***}	0.5304^{***}
	1Y	0.0403^{**}	0.2550^{***}	0.0487^{**}	0.0153^{**}	0.5753^{***}	0.2527^{***}
MLP	3M	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	6M	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1Y	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
RF	3M	0.0001	0.0060^{*}	0.0001	0.0000	0.1005^{***}	0.0004
	6M	0.0000	0.0240^{**}	0.0017^{*}	0.0000	0.0186^{**}	0.0024^{*}
	1Y	0.0000	0.0028^{*}	0.0033^{*}	0.0000	0.0024^{*}	0.0015^{*}
XGB	3M	0.0000	0.0001	0.0015^{*}	0.0000	0.0000	0.0052^{*}
	6M	0.0000	0.0000	0.0002	0.0000	0.0000	0.0111^{**}
	1Y	0.0001	0.0018^{*}	0.0000	0.0000	0.0000	0.0002
LSTM	3M	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	6M	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1Y	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

*Note: MW is the abbreviations of the moving window. Also, the star superscripts *, **, *** refer to 5%, 1%, and 0.1% statistical significances which mean that the paired t-test is **not rejected** at significance level $\alpha = 0.05, 0.01, 0.001$.

simultaneously. Note that the MLP and LSTM have the characteristic of consistent improvement in prediction performance.

4.3.2 Prediction performance in different sectors

Since the adequacy of models established, the detailed analysis of the prediction performances in the different sectors is evaluated based on the adjusted accuracy of the MLP, RF, XGB, and LSTM. Note that we exclude the LR due to its poor performance in the adjusted accuracy. Table 4.9 to 4.12 and Table 4.13 to 4.16 summarize the adjusted accuracy and their improvement by ETE network indicator of each sector, respectively, for different moving windows, lags, prediction periods. Note that the adjusted accuracy indicates the performance when predicted by the mixture of log-returns and the ETE network indicator.

Table 4.9 to 4.12 show that the adjusted accuracy of four algorithms tends to be higher for short-term predictions than long-term. In Table 4.13 to 4.16, the improvement of the adjusted accuracy is observed for all conditions of MLP and LSTM. For the 1W prediction period of RF and XGB, the improvement of the adjusted accuracy is generally observed in all sectors regardless of lag and the decrease in adjusted accuracy is observed in some sectors in 1M or 3M prediction periods where TELE and INFT show the increase of adjusted accuracy in most conditions.

Furthermore, we summarize the adjusted accuracy with their improvement in each sector for different moving windows based on the average of performance for all lags and prediction periods in Table 4.17. Although the performance of machine learning algorithms differs for each sector, we depict sectors with good performance regardless of the algorithm.

Table 4.9: Adjusted accuracy of each sector (MLP)

Sector	MW = 3M									MW = 6M								
	Lag = 3M			Lag = 6M			Lag = 1Y			Lag = 3M			Lag = 6M			Lag = 1Y		
	1W	1M	3M	1W	1M	3M	1W	1M	3M	1W	1M	3M	1W	1M	3M	1W	1M	3M
TELE	1.01	1.86	1.35	1.42	2.26	1.21	0.99	1.57	0.90	1.06	1.52	1.06	1.23	2.01	1.06	1.49	1.44	0.70
	(0.36)	(0.71)	(0.67)	(0.42)	(0.99)	(0.66)	(0.36)	(0.46)	(0.42)	(0.35)	(0.36)	(0.49)	(0.64)	(0.54)	(0.52)	(0.67)	(0.30)	(0.36)
COND	2.37	1.94	2.93	2.36	2.73	3.02	2.38	2.70	2.63	2.22	2.37	2.04	2.18	1.98	2.06	2.05	2.41	1.97
	(1.36)	(1.18)	(1.71)	(1.47)	(1.25)	(1.27)	(1.60)	(1.37)	(2.84)	(3.09)	(2.91)	(1.67)	(0.87)	(0.82)	(1.43)	(0.87)	(1.45)	(0.65)
ENRG	2.11	1.25	0.94	1.80	1.27	0.88	1.89	1.63	1.12	1.92	1.07	0.96	1.89	1.26	1.07	2.12	1.46	1.19
	(0.19)	(0.16)	(0.11)	(0.17)	(0.17)	(0.07)	(0.22)	(0.37)	(0.08)	(0.25)	(0.09)	(0.07)	(0.16)	(0.19)	(0.08)	(0.32)	(0.25)	(0.06)
HLCA	1.30	0.92	0.59	1.20	0.94	0.58	1.12	1.08	0.69	1.45	0.99	0.58	1.27	0.88	0.63	1.42	1.00	0.87
	(0.42)	(0.20)	(0.19)	(0.25)	(0.28)	(0.13)	(0.37)	(0.27)	(0.19)	(0.36)	(0.18)	(0.13)	(0.51)	(0.09)	(0.25)	(0.56)	(0.17)	(0.21)
INDS	6.64	5.26	3.36	4.54	5.13	4.03	3.52	5.28	3.92	6.97	4.49	3.18	3.35	4.85	4.36	5.36	4.80	4.71
	(3.20)	(0.93)	(1.15)	(1.32)	(1.18)	(1.03)	(2.90)	(1.33)	(1.33)	(2.83)	(1.16)	(0.75)	(1.09)	(0.60)	(1.06)	(3.56)	(1.26)	(1.40)
MTRS	1.45	1.42	1.77	1.45	1.72	1.74	1.60	1.73	1.74	1.51	1.83	1.94	1.67	1.69	1.83	1.61	1.73	1.76
	(0.24)	(0.24)	(0.10)	(0.24)	(0.14)	(0.07)	(0.35)	(0.23)	(0.08)	(0.15)	(0.15)	(0.10)	(0.41)	(0.17)	(0.12)	(0.34)	(0.20)	(0.15)
FINC	2.02	2.15	1.36	1.66	1.80	1.64	1.56	2.16	1.34	1.66	1.70	1.29	1.44	1.92	1.67	1.76	1.62	1.47
	(0.64)	(0.49)	(0.29)	(0.43)	(0.44)	(0.48)	(0.35)	(0.60)	(0.33)	(0.44)	(0.30)	(0.38)	(0.24)	(0.39)	(0.48)	(0.42)	(0.25)	(0.38)
CONS	1.69	1.97	2.03	1.68	2.33	1.79	1.61	2.07	2.25	1.66	2.32	1.90	1.98	2.16	1.85	1.64	2.59	2.07
	(0.32)	(0.32)	(0.56)	(0.25)	(0.77)	(0.29)	(0.56)	(0.61)	(0.41)	(0.24)	(0.54)	(0.17)	(0.61)	(0.45)	(0.22)	(0.51)	(0.83)	(0.30)
INFT	1.20	2.05	0.89	1.20	1.78	0.82	1.41	1.90	0.98	1.43	2.12	0.93	1.35	2.24	1.01	1.66	2.22	1.00
	(0.28)	(0.73)	(0.23)	(0.33)	(0.66)	(0.18)	(0.52)	(0.51)	(0.31)	(0.60)	(0.78)	(0.23)	(0.39)	(1.06)	(0.37)	(0.58)	(1.00)	(0.20)
UTIL	3.08	2.20	1.98	2.76	2.18	1.97	2.99	2.21	1.96	2.79	2.21	1.94	2.87	2.22	1.95	2.67	2.23	1.95
	(0.20)	(0.01)	(0.01)	(0.37)	(0.03)	(0.01)	(0.66)	(0.03)	(0.03)	(0.25)	(0.02)	(0.01)	(0.34)	(0.04)	(0.02)	(0.48)	(0.05)	(0.02)
REES	6.31	8.01	3.45	3.81	9.08	4.35	5.35	12.10	6.23	5.34	6.76	4.59	4.53	7.48	4.63	4.90	7.92	6.05
	(1.38)	(1.16)	(0.88)	(1.21)	(1.58)	(0.86)	(1.08)	(5.21)	(1.73)	(0.69)	(0.93)	(0.74)	(0.53)	(2.09)	(0.57)	(1.20)	(1.23)	(1.18)

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the standard deviation of adjusted accuracy performed 10 times.

Table 4.10: Adjusted accuracy of each sector (RF)

Sector	MW = 3M									MW = 6M								
	Lag = 3M			Lag = 6M			Lag = 1Y			Lag = 3M			Lag = 6M			Lag = 1Y		
	1W	1M	3M	1W	1M	3M	1W	1M	3M	1W	1M	3M	1W	1M	3M	1W	1M	3M
TELE	0.38	0.94	0.41	0.74	1.41	0.40	0.42	0.96	0.33	0.46	1.52	0.47	0.82	1.72	0.53	0.37	1.01	0.40
	(0.16)	(0.15)	(0.10)	(0.15)	(0.23)	(0.09)	(0.17)	(0.31)	(0.03)	(0.19)	(0.29)	(0.07)	(0.20)	(0.44)	(0.12)	(0.14)	(0.35)	(0.14)
COND	0.81	0.85	1.42	1.20	0.94	1.58	0.82	0.54	0.83	1.20	1.22	1.04	1.20	1.16	1.00	0.97	0.59	0.60
	(0.23)	(0.15)	(0.32)	(0.31)	(0.18)	(0.37)	(0.19)	(0.14)	(0.14)	(0.32)	(0.33)	(0.15)	(0.22)	(0.23)	(0.14)	(0.32)	(0.11)	(0.11)
ENRG	1.66	1.03	0.58	1.55	0.83	0.65	1.42	0.94	0.83	1.75	0.75	0.46	1.51	0.82	0.51	1.56	0.94	0.67
	(0.17)	(0.20)	(0.04)	(0.21)	(0.12)	(0.04)	(0.08)	(0.10)	(0.05)	(0.14)	(0.08)	(0.02)	(0.10)	(0.13)	(0.05)	(0.13)	(0.09)	(0.03)
HLCA	0.82	0.54	0.44	0.87	0.62	0.49	1.26	0.68	0.55	1.05	0.50	0.55	0.94	0.63	0.57	1.15	0.63	0.61
	(0.37)	(0.05)	(0.03)	(0.31)	(0.05)	(0.02)	(0.51)	(0.05)	(0.03)	(0.26)	(0.04)	(0.02)	(0.25)	(0.03)	(0.01)	(0.33)	(0.03)	(0.02)
INDS	3.10	5.00	1.96	3.37	4.25	2.20	2.98	4.82	1.57	3.85	3.91	1.48	4.03	4.16	1.78	2.81	4.18	1.41
	(1.41)	(0.74)	(0.40)	(1.15)	(0.37)	(0.27)	(0.66)	(0.38)	(0.14)	(2.33)	(0.58)	(0.13)	(1.09)	(0.54)	(0.21)	(0.68)	(0.36)	(0.17)
MTRS	1.28	0.64	1.55	1.42	0.98	1.67	1.42	1.15	1.39	1.35	1.17	1.57	1.75	1.42	1.70	1.54	1.68	1.41
	(0.43)	(0.09)	(0.11)	(0.24)	(0.13)	(0.07)	(0.17)	(0.19)	(0.12)	(0.13)	(0.11)	(0.07)	(0.37)	(0.11)	(0.11)	(0.22)	(0.11)	(0.14)
FINC	1.67	1.64	0.88	1.37	1.49	1.06	1.26	1.83	1.37	1.51	1.56	1.07	1.43	1.32	1.50	1.23	1.24	1.73
	(0.38)	(0.20)	(0.08)	(0.22)	(0.24)	(0.15)	(0.40)	(0.25)	(0.11)	(0.32)	(0.19)	(0.09)	(0.36)	(0.17)	(0.23)	(0.26)	(0.13)	(0.16)
CONS	1.06	1.22	1.34	1.52	1.47	1.47	1.30	1.35	1.76	1.17	1.28	2.28	1.44	1.42	2.47	1.18	1.42	2.43
	(0.15)	(0.07)	(0.10)	(0.38)	(0.21)	(0.10)	(0.24)	(0.15)	(0.23)	(0.17)	(0.15)	(0.16)	(0.15)	(0.16)	(0.20)	(0.20)	(0.12)	(0.20)
INFT	0.90	1.75	0.49	0.85	1.63	0.61	0.71	2.41	0.74	1.15	0.96	0.78	0.85	1.11	0.75	0.68	1.13	0.91
	(0.19)	(0.29)	(0.10)	(0.20)	(0.27)	(0.09)	(0.20)	(0.62)	(0.13)	(0.29)	(0.24)	(0.10)	(0.11)	(0.23)	(0.09)	(0.09)	(0.20)	(0.08)
UTIL	2.81	2.21	1.99	3.40	2.19	1.97	3.24	2.35	2.04	2.49	1.78	1.19	3.08	2.11	1.45	2.92	2.29	2.01
	(0.37)	(0.01)	(0.01)	(0.24)	(0.01)	(0.01)	(0.17)	(0.04)	(0.01)	(0.43)	(0.22)	(0.13)	(0.40)	(0.19)	(0.25)	(0.36)	(0.03)	(0.02)
REES	4.83	4.07	2.01	4.04	3.69	2.15	4.78	3.08	2.15	4.90	3.99	2.42	4.39	4.01	2.43	4.64	3.78	2.96
	(1.06)	(0.36)	(0.05)	(1.03)	(0.56)	(0.05)	(0.94)	(0.18)	(0.08)	(0.84)	(0.43)	(0.36)	(0.81)	(0.55)	(0.31)	(0.98)	(0.28)	(0.15)

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the standard deviation of adjusted accuracy performed 10 times.

Table 4.11: Adjusted accuracy of each sector (XGB)

Sector	**MW = 3M**									**MW = 6M**								
	Lag = 3M			**Lag = 6M**			**Lag = 1Y**			**Lag = 3M**			**Lag = 6M**			**Lag = 1Y**		
	1W	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**
TELE	0.55	0.78	0.93	0.85	0.83	0.67	0.87	0.88	1.03	0.64	1.27	0.77	0.89	1.76	0.97	1.13	1.59	1.29
	(0.44)	(0.14)	(0.43)	(0.29)	(0.14)	(0.33)	(0.70)	(0.27)	(0.34)	(0.41)	(0.34)	(0.18)	(0.47)	(0.41)	(0.22)	(0.76)	(0.36)	(0.32)
COND	1.00	1.08	1.02	1.06	0.92	1.18	1.05	0.61	1.22	1.28	1.93	1.25	1.11	1.53	1.11	1.06	1.11	1.00
	(0.35)	(0.28)	(0.18)	(0.29)	(0.23)	(0.30)	(0.44)	(0.28)	(0.28)	(0.63)	(0.60)	(0.25)	(0.27)	(0.20)	(0.24)	(0.33)	(0.26)	(0.19)
ENRG	1.58	1.20	0.71	1.72	1.40	0.68	1.66	1.41	0.82	2.00	1.08	0.58	1.69	1.30	0.65	1.83	1.75	0.62
	(0.65)	(0.39)	(0.06)	(0.47)	(0.40)	(0.08)	(0.54)	(0.39)	(0.16)	(1.74)	(0.23)	(0.11)	(0.70)	(0.30)	(0.07)	(0.96)	(0.74)	(0.09)
HLCA	1.38	0.78	0.69	0.91	0.69	0.75	1.35	0.72	0.74	1.06	0.90	0.57	1.53	0.82	0.60	1.22	0.89	0.67
	(0.60)	(0.14)	(0.07)	(0.39)	(0.11)	(0.07)	(0.83)	(0.10)	(0.13)	(0.45)	(0.13)	(0.11)	(0.73)	(0.11)	(0.10)	(0.66)	(0.12)	(0.12)
INDS	2.71	2.59	2.42	2.54	2.45	2.45	2.55	2.80	2.55	2.63	2.88	1.91	3.32	2.64	1.91	1.89	2.71	1.74
	(0.75)	(0.19)	(0.33)	(0.93)	(0.31)	(0.25)	(0.65)	(0.47)	(0.35)	(0.70)	(0.48)	(0.18)	(2.22)	(0.51)	(0.49)	(0.98)	(0.46)	(0.40)
MTRS	1.16	0.74	1.04	1.22	0.83	1.26	1.49	0.80	1.14	1.38	1.00	1.27	1.52	1.13	1.35	1.59	1.26	1.31
	(0.31)	(0.17)	(0.15)	(0.41)	(0.18)	(0.16)	(0.73)	(0.12)	(0.14)	(0.49)	(0.11)	(0.07)	(0.75)	(0.27)	(0.17)	(0.51)	(0.20)	(0.14)
FINC	1.50	2.38	0.99	1.42	2.28	1.06	1.33	2.17	1.00	1.89	1.86	2.14	1.78	2.10	1.56	1.53	1.86	1.87
	(0.77)	(0.77)	(0.12)	(0.49)	(1.07)	(0.15)	(0.39)	(0.75)	(0.15)	(0.63)	(0.38)	(0.58)	(0.86)	(0.36)	(0.36)	(0.49)	(0.43)	(0.40)
CONS	1.41	1.82	1.89	1.86	1.85	1.86	1.35	1.71	1.84	1.41	1.63	3.83	1.46	1.53	4.33	1.28	1.45	4.19
	(0.52)	(0.25)	(0.29)	(0.99)	(0.43)	(0.21)	(0.29)	(0.39)	(0.16)	(0.26)	(0.21)	(0.62)	(0.24)	(0.25)	(0.57)	(0.21)	(0.16)	(1.12)
INFT	1.11	1.54	0.64	0.93	1.49	0.76	0.80	1.75	0.76	1.10	0.98	0.81	0.91	0.83	1.07	0.91	1.09	0.93
	(0.28)	(0.48)	(0.15)	(0.20)	(0.30)	(0.16)	(0.33)	(0.21)	(0.15)	(0.26)	(0.24)	(0.08)	(0.28)	(0.19)	(0.13)	(0.15)	(0.21)	(0.20)
UTIL	2.80	2.29	2.02	2.37	2.50	2.10	3.09	2.62	2.13	2.45	1.27	0.93	2.94	1.31	0.98	2.17	1.45	1.01
	(0.62)	(0.25)	(0.11)	(0.38)	(0.18)	(0.07)	(1.91)	(0.17)	(0.15)	(0.65)	(0.20)	(0.10)	(0.96)	(0.18)	(0.09)	(0.40)	(0.27)	(0.08)
REES	4.31	3.06	1.54	4.00	3.08	1.64	4.15	4.07	1.41	3.74	3.02	1.57	3.64	2.84	1.29	4.10	2.24	1.20
	(1.40)	(0.84)	(0.14)	(1.16)	(1.11)	(0.10)	(0.88)	(0.88)	(0.19)	(1.32)	(0.65)	(0.24)	(1.44)	(0.80)	(0.14)	(2.03)	(0.71)	(0.12)

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the standard deviation of adjusted accuracy performed 10 times.

Table 4.12: Adjusted accuracy of each sector (LSTM)

Sector	**MW = 3M**									**MW = 6M**								
	Lag = 3M			**Lag = 6M**			**Lag = 1Y**			**Lag = 3M**			**Lag = 6M**			**Lag = 1Y**		
	1W	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**
TELE	3.17	3.22	6.78	1.85	4.54	6.87	2.13	4.68	8.57	1.95	3.60	7.05	2.19	3.99	6.96	2.85	1.98	8.65
	(1.70)	(0.90)	(3.01)	(1.28)	(1.90)	(3.06)	(0.87)	(1.30)	(2.02)	(1.06)	(1.63)	(2.38)	(1.03)	(1.94)	(1.82)	(1.16)	(2.05)	(4.59)
COND	1.76	2.38	7.13	1.72	2.60	6.13	1.52	4.16	4.43	1.26	2.67	3.68	0.97	4.36	2.96	2.06	3.71	2.77
	(2.03)	(2.74)	(1.86)	(0.76)	(2.66)	(1.40)	(1.46)	(3.81)	(3.23)	(1.06)	(2.26)	(2.35)	(1.29)	(1.49)	(4.32)	(1.17)	(2.07)	(1.12)
ENRG	1.13	1.60	1.75	1.74	2.27	2.59	1.24	1.39	0.94	0.99	1.37	2.04	0.96	2.09	2.15	1.30	1.15	1.95
	(0.67)	(0.73)	(0.39)	(0.95)	(0.66)	(0.58)	(0.76)	(0.76)	(1.08)	(0.60)	(0.75)	(0.50)	(0.47)	(0.48)	(0.86)	(0.34)	(0.33)	(1.06)
HLCA	2.13	2.47	6.63	2.34	4.46	5.37	2.89	2.80	8.35	2.31	4.23	5.67	1.45	4.82	9.40	2.49	2.14	6.55
	(1.17)	(1.15)	(2.25)	(0.82)	(1.31)	(1.50)	(1.44)	(2.10)	(1.87)	(1.03)	(2.02)	(1.50)	(0.75)	(2.47)	(1.67)	(1.21)	(1.46)	(1.93)
INDS	2.10	4.87	3.44	1.98	5.22	3.16	2.77	5.23	4.45	2.45	4.68	3.50	1.96	5.32	2.77	2.47	5.85	3.99
	(1.55)	(1.38)	(3.13)	(1.41)	(1.41)	(2.55)	(1.22)	(2.01)	(3.70)	(1.01)	(1.59)	(2.18)	(0.99)	(1.50)	(1.41)	(1.14)	(2.61)	(2.95)
MTRS	1.82	3.73	4.19	0.95	2.23	3.24	1.88	2.73	1.40	1.68	3.16	2.54	1.27	2.60	1.92	1.95	2.34	3.88
	(0.80)	(1.36)	(1.56)	(0.93)	(0.38)	(2.38)	(0.38)	(1.38)	(1.79)	(0.40)	(1.47)	(1.13)	(0.57)	(1.26)	(0.67)	(0.67)	(1.04)	(1.56)
FINC	1.37	2.54	4.59	1.52	2.38	4.38	1.96	2.31	3.14	1.82	2.11	4.21	1.53	3.04	4.16	2.01	3.05	4.42
	(0.81)	(0.63)	(1.34)	(0.39)	(0.99)	(1.94)	(0.41)	(1.08)	(2.35)	(0.35)	(0.88)	(1.32)	(0.35)	(0.78)	(1.58)	(0.55)	(1.24)	(1.23)
CONS	1.78	2.18	3.42	1.51	2.06	3.23	2.66	1.57	2.83	1.47	1.10	2.02	2.14	1.49	3.50	2.04	2.03	1.82
	(1.16)	(1.11)	(1.76)	(0.65)	(3.03)	(1.36)	(1.61)	(1.71)	(1.77)	(0.58)	(0.99)	(0.69)	(0.83)	(0.54)	(1.37)	(1.12)	(0.91)	(0.66)
INFT	3.46	2.01	2.67	2.56	2.70	4.31	2.44	2.07	3.69	2.55	4.11	2.64	2.54	3.41	3.01	2.34	3.28	2.40
	(0.98)	(0.99)	(1.19)	(1.07)	(1.09)	(1.75)	(0.85)	(2.79)	(2.31)	(0.58)	(1.84)	(1.40)	(0.98)	(2.05)	(1.23)	(1.27)	(1.05)	(1.70)
UTIL	0.47	1.15	1.92	1.14	2.50	2.11	1.16	2.36	2.11	0.86	0.83	2.14	0.74	1.56	2.20	1.07	1.66	1.46
	(0.47)	(0.67)	(3.09)	(0.99)	(1.22)	(1.37)	(0.92)	(0.74)	(1.12)	(0.56)	(0.29)	(1.03)	(0.46)	(1.11)	(1.35)	(0.32)	(0.74)	(1.08)
REES	0.97	0.76	0.53	0.74	2.12	3.04	0.86	1.38	5.09	0.51	0.70	2.04	1.22	2.40	2.91	0.64	1.79	5.61
	(0.75)	(1.34)	(0.90)	(0.35)	(1.41)	(0.97)	(0.47)	(2.05)	(5.08)	(0.67)	(0.43)	(1.50)	(0.54)	(1.41)	(2.45)	(0.53)	(0.93)	(3.23)

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the standard deviation of adjusted accuracy performed 10 times.

Table 4.13: Adjusted accuracy improvement by ETE network indicators of each sector (MLP)

Sector	**MW = 3M**									**MW = 6M**								
	Lag = 3M			**Lag = 6M**			**Lag = 1Y**			**Lag = 3M**			**Lag = 6M**			**Lag = 1Y**		
	1W	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**
TELE	1.47	1.91	1.30	1.88	1.40	1.67	1.26	1.22	1.32	1.57	1.27	1.62	1.70	1.36	1.57	1.91	1.10	0.95
	(0.58)	(0.74)	(0.31)	(1.22)	(0.59)	(1.07)	(0.60)	(0.50)	(0.26)	(0.78)	(0.41)	(0.67)	(0.79)	(0.66)	(0.50)	(1.15)	(0.28)	(0.57)
COND	1.09	0.84	1.04	1.97	1.22	1.16	1.18	1.28	0.96	0.87	0.92	1.08	1.28	1.08	1.25	1.65	1.16	0.89
	(0.42)	(0.45)	(0.66)	(1.31)	(0.31)	(0.47)	(0.57)	(0.40)	(0.39)	(0.56)	(0.47)	(1.08)	(0.53)	(0.30)	(1.05)	(2.20)	(0.47)	(0.48)
ENRG	1.22	0.45	0.18	0.94	0.66	0.12	1.35	0.91	0.45	1.74	0.73	0.48	1.45	0.90	1.05	1.74	0.84	0.53
	(0.66)	(0.32)	(0.33)	(0.37)	(0.43)	(0.40)	(0.43)	(0.38)	(0.36)	(1.18)	(0.45)	(0.54)	(0.36)	(0.72)	(0.63)	(0.89)	(0.45)	(0.27)
HLCA	1.15	1.22	1.15	1.22	1.17	1.18	1.37	1.16	1.11	1.28	1.15	0.93	1.43	1.02	1.19	1.58	1.26	1.00
	(0.71)	(0.45)	(0.27)	(0.41)	(0.46)	(0.36)	(0.74)	(0.49)	(0.43)	(0.42)	(0.34)	(0.48)	(0.54)	(0.46)	(0.32)	(1.23)	(0.50)	(0.32)
INDS	1.96	1.13	0.41	1.69	1.24	0.95	1.30	1.16	0.74	1.61	0.86	0.74	1.19	1.07	1.16	1.38	1.17	0.86
	(1.38)	(0.28)	(0.39)	(0.93)	(0.52)	(0.55)	(0.87)	(0.54)	(0.48)	(0.98)	(0.51)	(0.25)	(0.42)	(0.28)	(0.41)	(0.53)	(0.44)	(0.40)
MTRS	1.08	0.61	0.62	0.78	1.13	0.65	1.01	1.28	1.06	0.70	1.07	0.92	0.96	1.07	1.00	1.10	1.11	0.88
	(0.32)	(0.21)	(0.31)	(0.43)	(0.44)	(0.10)	(0.26)	(0.41)	(0.36)	(0.52)	(0.45)	(0.33)	(0.35)	(0.28)	(0.38)	(0.50)	(0.37)	(0.35)
FINC	0.99	1.20	0.94	1.87	1.19	0.92	1.22	1.21	0.83	0.91	1.12	0.85	1.28	1.30	0.89	1.26	1.07	0.89
	(0.35)	(0.42)	(0.19)	(2.39)	(0.31)	(0.38)	(0.38)	(0.26)	(0.33)	(0.35)	(0.30)	(0.25)	(0.55)	(0.50)	(0.20)	(0.79)	(0.44)	(0.41)
CONS	1.17	0.91	0.71	1.14	0.97	0.82	0.78	1.39	0.86	1.44	0.99	0.81	1.39	1.37	1.12	1.35	1.06	1.12
	(0.33)	(0.32)	(0.17)	(0.45)	(0.33)	(0.28)	(0.23)	(0.81)	(0.29)	(0.79)	(0.27)	(0.38)	(0.89)	(0.41)	(0.55)	(1.12)	(0.33)	(0.51)
INFT	0.99	1.08	1.81	1.33	1.34	1.41	1.27	1.26	1.05	1.55	1.29	1.15	1.33	1.50	1.06	1.33	1.45	1.51
	(0.41)	(0.60)	(1.35)	(0.46)	(0.47)	(0.49)	(0.67)	(0.65)	(0.43)	(0.93)	(0.74)	(0.33)	(0.63)	(1.02)	(0.25)	(0.31)	(0.70)	(1.16)
UTIL	1.33	0.83	0.53	1.06	0.80	0.82	1.37	0.95	0.77	1.43	0.84	0.77	1.12	0.98	0.88	1.15	0.92	0.84
	(0.54)	(0.25)	(0.31)	(0.35)	(0.16)	(0.32)	(0.48)	(0.27)	(0.22)	(0.69)	(0.23)	(0.32)	(0.37)	(0.45)	(0.23)	(0.46)	(0.28)	(0.35)
REES	1.14	0.69	0.36	0.98	0.84	0.78	1.28	1.46	0.70	0.94	0.43	0.34	1.28	0.81	0.55	1.55	0.90	0.62
	(0.77)	(0.27)	(0.55)	(0.43)	(0.40)	(0.36)	(0.52)	(1.86)	(0.43)	(0.39)	(0.51)	(0.38)	(0.56)	(0.37)	(0.48)	(0.42)	(0.25)	(0.34)

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the standard deviation of adjusted accuracy improvement performed 10 times.

Table 4.14: Adjusted accuracy improvement by ETE network indicators of each sector (RF)

Sector	**MW = 3M**									**MW = 6M**								
	Lag = 3M			**Lag = 6M**			**Lag = 1Y**			**Lag = 3M**			**Lag = 6M**			**Lag = 1Y**		
	1W	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**
TELE	0.61	0.52	0.93	1.08	0.58	1.08	1.25	0.84	1.11	1.13	0.77	0.93	1.61	0.78	0.99	1.50	1.24	0.76
	(0.27)	(0.15)	(0.15)	(0.42)	(0.13)	(0.08)	(0.74)	(0.26)	(0.19)	(0.39)	(0.07)	(0.14)	(0.47)	(0.13)	(0.07)	(0.81)	(0.39)	(0.15)
COND	-0.01	-0.38	-0.05	0.65	-0.39	0.07	0.91	-0.22	0.49	1.27	0.08	-1.20	1.67	0.58	-0.23	1.81	0.81	0.79
	(0.44)	(0.04)	(0.17)	(0.78)	(0.11)	(0.14)	(0.87)	(0.13)	(0.17)	(0.42)	(0.50)	(0.43)	(1.12)	(0.32)	(0.38)	(0.78)	(0.13)	(0.21)
ENRG	1.61	0.67	0.60	1.44	0.54	0.06	1.66	0.77	-0.23	1.13	0.08	0.13	1.38	0.42	-1.25	1.47	0.35	-0.30
	(0.59)	(0.15)	(0.22)	(0.76)	(0.24)	(0.16)	(0.44)	(0.38)	(0.07)	(0.28)	(0.38)	(0.08)	(0.94)	(0.26)	(0.38)	(0.56)	(0.24)	(0.08)
HLCA	-0.13	0.15	0.18	-0.05	-0.07	0.22	0.56	0.45	0.36	1.17	-0.08	0.27	1.31	0.49	0.33	1.87	0.22	0.53
	(0.16)	(0.33)	(0.10)	(0.15)	(0.25)	(0.07)	(1.06)	(0.50)	(0.30)	(0.69)	(0.42)	(0.04)	(0.88)	(0.64)	(0.08)	(0.72)	(0.23)	(0.07)
INDS	0.91	0.69	0.08	1.33	0.75	0.10	1.32	0.58	0.14	0.95	-0.40	-0.68	1.20	-0.04	-0.49	1.83	0.35	-0.21
	(0.50)	(0.34)	(0.31)	(0.58)	(0.45)	(0.15)	(0.78)	(0.17)	(0.10)	(0.48)	(0.30)	(0.12)	(0.76)	(0.21)	(0.12)	(2.01)	(0.29)	(0.13)
MTRS	-0.04	-0.46	-0.38	-0.01	-0.38	-0.11	0.58	-0.31	0.20	0.52	-0.37	-0.71	1.62	-0.04	0.22	1.49	0.24	0.87
	(0.20)	(0.03)	(0.07)	(0.36)	(0.05)	(0.18)	(0.61)	(0.07)	(0.11)	(0.33)	(0.15)	(0.12)	(0.98)	(0.31)	(0.19)	(1.09)	(0.30)	(0.42)
FINC	1.56	0.71	0.52	2.44	1.56	0.14	1.87	1.46	0.43	1.51	1.36	0.67	2.01	1.02	0.83	1.69	1.50	1.68
	(1.25)	(0.34)	(0.14)	(2.73)	(0.78)	(0.11)	(0.97)	(0.57)	(0.20)	(0.58)	(0.44)	(0.10)	(1.29)	(0.33)	(0.13)	(1.07)	(0.42)	(0.28)
CONS	0.63	0.14	-0.28	0.59	0.45	-0.19	1.20	0.56	-0.42	1.28	0.50	0.05	1.42	0.83	-0.02	1.35	0.51	-0.31
	(0.52)	(0.51)	(0.20)	(0.47)	(0.63)	(0.09)	(0.54)	(0.34)	(0.11)	(0.49)	(0.55)	(0.05)	(0.62)	(0.79)	(0.09)	(0.46)	(0.38)	(0.10)
INFT	0.68	0.71	0.21	1.01	0.68	0.35	0.98	0.74	0.94	1.06	0.15	0.40	1.07	0.07	0.21	0.96	0.43	0.69
	(0.23)	(0.06)	(0.13)	(0.47)	(0.06)	(0.19)	(0.45)	(0.13)	(0.24)	(0.37)	(0.08)	(0.03)	(0.42)	(0.11)	(0.10)	(0.32)	(0.14)	(0.08)
UTIL	0.89	0.20	-0.23	1.66	0.70	-0.12	1.33	1.40	0.56	0.95	-0.40	-0.53	0.95	-0.28	-0.52	1.03	0.69	-0.38
	(0.48)	(0.35)	(0.16)	(1.25)	(0.53)	(0.09)	(0.55)	(0.87)	(0.62)	(0.54)	(0.06)	(0.02)	(0.99)	(0.13)	(0.06)	(0.47)	(0.62)	(0.06)
REES	1.11	0.08	-0.54	1.13	0.25	-0.64	0.89	0.83	-0.20	1.75	-0.65	-0.44	1.22	-0.35	-0.50	1.30	-0.17	0.20
	(0.58)	(0.27)	(0.07)	(0.55)	(0.46)	(0.11)	(0.38)	(0.45)	(0.12)	(0.99)	(0.27)	(0.07)	(0.55)	(0.20)	(0.08)	(0.46)	(0.38)	(0.26)

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the standard deviation of adjusted accuracy improvement performed 10 times.

Table 4.15: Adjusted accuracy improvement by ETE network indicators of each sector (XGB)

Sector	MW = 3M									MW = 6M								
	Lag = 3M			Lag = 6M			Lag = 1Y			Lag = 3M			Lag = 6M			Lag = 1Y		
	1W	1M	3M	1W	1M	3M	1W	1M	3M	1W	1M	3M	1W	1M	3M	1W	1M	3M
TELE	0.40	0.36	0.97	0.79	0.35	1.24	0.34	0.47	1.46	0.57	0.42	0.78	0.74	0.52	1.33	0.58	1.06	1.20
	(0.40)	(0.29)	(0.36)	(0.42)	(0.22)	(0.55)	(0.50)	(0.19)	(0.82)	(0.54)	(0.29)	(0.22)	(0.38)	(0.20)	(0.36)	(0.49)	(0.33)	(0.35)
COND	0.70	-0.15	-0.33	0.56	-0.22	-0.22	0.37	-0.27	0.32	0.67	0.17	-0.25	0.74	0.50	-0.25	0.65	0.61	0.25
	(0.73)	(0.30)	(0.24)	(0.54)	(0.28)	(0.14)	(0.26)	(0.31)	(0.17)	(0.40)	(0.37)	(0.26)	(0.61)	(0.35)	(0.42)	(0.35)	(0.38)	(0.16)
ENRG	0.60	0.39	-0.01	0.64	0.48	0.03	0.72	0.18	0.37	0.41	0.16	0.08	1.08	0.19	-0.03	0.67	0.23	-0.15
	(0.44)	(0.27)	(0.22)	(0.56)	(0.30)	(0.32)	(0.31)	(0.33)	(0.32)	(0.53)	(0.18)	(0.40)	(1.31)	(0.15)	(0.30)	(0.96)	(0.16)	(0.16)
HLCA	0.01	0.07	-0.01	-0.29	-0.02	-0.02	0.01	0.20	0.06	0.72	0.72	0.03	0.80	0.55	-0.05	0.50	0.80	0.32
	(0.21)	(0.19)	(0.22)	(0.14)	(0.24)	(0.12)	(0.44)	(0.38)	(0.20)	(0.87)	(0.32)	(0.18)	(0.62)	(0.32)	(0.13)	(0.30)	(0.27)	(0.28)
INDS	1.11	0.15	0.04	0.91	0.37	0.10	0.87	0.68	0.25	0.60	-0.04	0.04	0.67	0.31	0.04	0.57	0.83	0.13
	(0.30)	(0.24)	(0.17)	(0.30)	(0.28)	(0.08)	(0.20)	(0.28)	(0.09)	(0.28)	(0.36)	(0.09)	(0.69)	(0.38)	(0.28)	(0.48)	(0.65)	(0.22)
MTRS	0.33	-0.38	-0.57	0.34	-0.16	-0.24	0.38	-0.16	-0.20	0.68	-0.24	-0.31	0.86	-0.02	-0.13	0.46	0.02	-0.05
	(0.47)	(0.21)	(0.12)	(0.37)	(0.20)	(0.13)	(0.44)	(0.20)	(0.11)	(0.42)	(0.23)	(0.11)	(0.39)	(0.38)	(0.10)	(0.33)	(0.26)	(0.09)
FINC	0.85	-0.25	-0.04	0.63	-0.16	0.03	0.78	0.18	-0.06	0.51	0.62	0.77	0.47	1.82	0.79	0.61	0.92	1.01
	(0.56)	(0.20)	(0.25)	(0.39)	(0.25)	(0.15)	(0.23)	(0.56)	(0.20)	(0.41)	(0.39)	(0.27)	(0.44)	(0.93)	(0.27)	(0.41)	(0.48)	(0.35)
CONS	0.22	-0.18	-0.21	0.56	-0.09	-0.04	0.52	-0.02	-0.13	0.93	-0.04	-0.41	0.86	0.69	-0.34	0.70	0.26	-0.57
	(0.41)	(0.27)	(0.13)	(0.36)	(0.46)	(0.17)	(0.52)	(0.35)	(0.15)	(0.76)	(0.28)	(0.09)	(0.22)	(0.99)	(0.11)	(0.23)	(0.30)	(0.10)
INFT	0.92	0.41	0.13	0.91	0.53	0.21	0.56	0.87	0.33	0.94	0.25	0.48	0.80	0.19	0.70	0.55	0.41	0.72
	(0.34)	(0.23)	(0.19)	(0.23)	(0.14)	(0.25)	(0.45)	(0.33)	(0.22)	(0.19)	(0.25)	(0.14)	(0.26)	(0.35)	(0.27)	(0.53)	(0.23)	(0.32)
UTIL	0.60	0.49	-1.07	0.76	0.63	-0.56	0.51	0.35	-0.34	0.48	-0.47	-0.63	0.16	-0.52	-0.64	0.39	-0.60	-0.65
	(0.41)	(0.42)	(0.80)	(0.66)	(0.39)	(0.34)	(0.52)	(0.18)	(0.38)	(0.55)	(0.11)	(0.04)	(0.58)	(0.13)	(0.05)	(0.43)	(0.28)	(0.03)
REES	0.96	-0.39	-0.61	0.93	-0.52	-0.71	0.80	-0.56	-0.61	0.59	-0.57	-0.62	0.64	-0.71	-0.59	0.63	-0.48	-0.52
	(0.52)	(0.49)	(0.06)	(0.83)	(0.35)	(0.08)	(0.43)	(0.57)	(0.08)	(0.48)	(0.23)	(0.08)	(0.65)	(0.20)	(0.05)	(0.35)	(0.27)	(0.09)

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the standard deviation of adjusted accuracy improvement performed 10 times.

Table 4.16: Adjusted accuracy improvement by ETE network indicators of each sector (LSTM)

Sector	**MW = 3M**									**MW = 6M**								
	Lag = 3M			**Lag = 6M**			**Lag = 1Y**			**Lag = 3M**			**Lag = 6M**			**Lag = 1Y**		
	1W	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**
TELE	1.01	1.52	1.18	1.20	1.04	1.07	1.30	1.34	1.25	0.88	1.20	1.49	1.29	1.17	0.69	1.00	0.47	1.09
	(0.62)	(0.74)	(0.29)	(1.08)	(0.13)	(0.36)	(0.62)	(0.49)	(0.36)	(0.68)	(0.96)	(0.43)	(0.71)	(0.65)	(0.23)	(0.44)	(0.47)	(0.64)
COND	0.44	0.34	1.44	0.66	0.32	1.23	0.45	0.71	0.53	0.55	0.70	0.88	0.49	0.83	0.44	0.70	1.09	0.54
	(0.43)	(0.45)	(0.89)	(0.47)	(0.28)	(0.62)	(0.46)	(0.63)	(0.22)	(0.41)	(0.67)	(0.60)	(0.55)	(0.48)	(0.58)	(0.38)	(0.80)	(0.37)
ENRG	1.31	0.92	1.32	1.43	1.71	1.40	0.94	0.89	0.40	1.06	1.35	1.06	1.06	1.34	1.54	1.06	0.91	0.84
	(0.90)	(0.45)	(0.49)	(0.67)	(1.20)	(0.61)	(0.47)	(0.59)	(0.39)	(0.66)	(1.43)	(0.58)	(0.72)	(0.65)	(0.83)	(0.39)	(0.17)	(0.14)
HLCA	0.93	0.63	1.18	1.07	1.15	1.33	1.30	1.27	1.02	1.14	0.98	0.99	0.87	1.16	1.03	0.95	0.70	0.96
	(0.60)	(0.28)	(0.29)	(0.57)	(0.43)	(0.84)	(0.69)	(1.49)	(0.14)	(0.42)	(0.44)	(0.32)	(0.60)	(0.72)	(0.34)	(0.36)	(0.53)	(0.31)
INDS	0.94	1.45	0.97	1.18	1.68	0.98	1.24	1.24	1.12	0.83	1.64	0.90	0.97	0.99	0.48	1.50	1.00	0.76
	(0.39)	(0.43)	(0.69)	(1.11)	(1.25)	(0.77)	(0.73)	(0.52)	(1.21)	(0.36)	(1.07)	(0.77)	(0.48)	(0.18)	(0.23)	(0.70)	(0.35)	(0.78)
MTRS	0.82	1.37	1.19	0.71	1.02	0.75	0.87	0.83	0.46	0.67	0.95	0.97	0.80	1.04	1.03	0.92	0.82	1.04
	(0.28)	(0.41)	(0.27)	(0.71)	(0.65)	(0.75)	(0.23)	(0.31)	(0.55)	(0.15)	(0.36)	(0.47)	(0.27)	(0.42)	(0.27)	(0.24)	(0.36)	(0.27)
FINC	1.13	1.52	1.61	1.38	1.08	1.43	1.48	0.99	1.15	2.24	1.05	1.26	1.18	1.32	2.12	1.27	1.02	1.77
	(0.84)	(0.48)	(0.73)	(0.72)	(0.45)	(0.53)	(0.56)	(0.47)	(0.82)	(1.21)	(0.36)	(0.34)	(0.46)	(0.32)	(2.22)	(0.26)	(0.32)	(1.36)
CONS	0.98	1.11	1.28	0.86	0.69	1.05	1.22	0.65	0.95	0.89	0.70	0.97	1.33	0.78	1.60	1.19	1.17	0.76
	(0.49)	(0.49)	(0.44)	(0.31)	(0.82)	(0.29)	(0.54)	(0.65)	(0.51)	(0.28)	(0.64)	(0.45)	(0.57)	(0.34)	(0.60)	(0.70)	(0.35)	(0.36)
INFT	1.16	0.53	0.78	0.93	0.78	0.97	1.11	0.32	0.79	1.08	1.75	0.61	1.13	0.87	0.78	0.96	0.72	0.52
	(0.27)	(0.25)	(0.24)	(0.36)	(0.60)	(0.30)	(0.26)	(0.54)	(0.20)	(0.23)	(1.48)	(0.22)	(0.37)	(0.21)	(0.48)	(0.49)	(0.29)	(0.37)
UTIL	0.77	0.81	0.57	0.82	1.08	1.17	1.19	0.85	0.70	0.73	0.79	1.12	1.08	0.85	0.93	1.02	0.88	0.54
	(0.53)	(0.47)	(0.69)	(0.43)	(0.50)	(0.65)	(0.78)	(0.26)	(0.14)	(0.39)	(0.40)	(0.75)	(0.87)	(0.29)	(0.49)	(0.34)	(0.28)	(0.47)
REES	0.85	0.39	0.09	0.68	0.73	1.16	0.68	0.49	0.48	0.58	0.40	0.49	1.00	0.78	0.60	0.42	0.66	0.84
	(0.42)	(0.68)	(0.26)	(0.32)	(0.55)	(0.57)	(0.20)	(0.82)	(0.49)	(0.73)	(0.25)	(0.37)	(0.39)	(0.23)	(0.26)	(0.19)	(0.23)	(0.36)

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the standard deviation of adjusted accuracy improvement performed 10 times.

Table 4.17: Adjusted accuracy and their improvements in different sectors

	MLP		**RF**		**XGB**		**LSTM**	
MW	**3M**	**6M**	**3M**	**6M**	**3M**	**6M**	**3M**	**6M**
TELE	1.40	1.29	0.67	0.81	0.82	1.15	1.32	1.19
	(1.49)	(1.45)	(0.89)	(1.08)	(0.71)	(0.80)	(1.21)	(1.03)
COND	2.56	2.14	1.00	1.00	1.02	1.26	1.53	1.39
	(1.19)	(1.13)	(0.12)	(0.62)	(0.08)	(0.34)	(0.68)	(0.69)
ENRG	1.43	1.44	1.05	1.00	1.24	1.28	1.32	1.35
	(0.70)	(1.05)	(0.79)	(0.38)	(0.38)	(0.29)	(1.15)	(1.14)
HLCA	0.94	1.01	0.70	0.74	0.89	0.92	0.86	0.95
	(1.19)	(1.20)	(0.19)	(0.68)	(0.00)	(0.49)	(1.10)	(0.98)
INDS	4.63	4.67	3.25	3.07	2.56	2.40	4.32	3.89
	(1.18)	(1.12)	(0.66)	(0.28)	(0.50)	(0.35)	(1.20)	(1.01)
MTRS	1.62	1.73	1.28	1.51	1.08	1.31	1.77	1.79
	(0.91)	(0.98)	(-0.10)	(0.43)	(-0.07)	(0.14)	(0.89)	(0.92)
FINC	1.74	1.61	1.40	1.40	1.57	1.84	1.71	1.66
	(1.15)	(1.06)	(1.19)	(1.36)	(0.22)	(0.84)	(1.31)	(1.47)
CONS	1.94	2.02	1.39	1.68	1.73	2.35	1.43	1.40
	(0.97)	(1.18)	(0.30)	(0.62)	(0.07)	(0.23)	(0.98)	(1.04)
INFT	1.36	1.55	1.12	0.92	1.09	0.96	1.02	1.29
	(1.28)	(1.35)	(0.70)	(0.56)	(0.54)	(0.56)	(0.82)	(0.94)
UTIL	2.37	2.31	2.47	2.15	2.44	1.61	2.37	2.33
	(0.94)	(0.99)	(0.71)	(0.17)	(0.15)	(-0.28)	(0.88)	(0.88)
REES	6.52	5.80	3.42	3.72	3.03	2.63	5.15	5.56
	(0.91)	(0.82)	(0.32)	(0.26)	(-0.08)	(-0.18)	(0.62)	(0.64)

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the improvement.

In order to further analyze the prediction performance on each sector, we draw the scatter plots for adjusted accuracy in Figure 4.5. Note that the sector-by-sector mean of the prediction performance is on the X-axis, whereas the mean of improvement achieved by integrating ETE network indicator is on the Y-axis. We set the means of $X-$ and $Y-$ axes as red dashed lines and the $\pm 1\sigma$, an indifferent region, as a gray box. Specifically, the first quadrant of the scatter plot is the best-case scenario for the proposed prediction framework containing the sectors with high prediction adjusted accuracy and improvement. The second (fourth) quadrant encompasses sectors that have poor (decent) prediction performance but have decent (poor) performance improvement from ETE indicator. The third quadrant is the worst-case scenario containing sectors with poor performance in both prediction performance and performance improvement from ETE indicator.

In Figure 4.5, INDS, UTIL, and REES are all located in the first and fourth quadrant except for UTIL of MLP in 3M, showing consistent high accuracy. In particular, INDS and REES have the high adjusted accuracy in all models. HCLA except for RF, XGB in 3M and INFT except for LSTM and RF in 6M are located in the second quadrant, showing the high improvement in the consistent prediction performance. Also, the high improvement in adjusted accuracy is observed in TELE for MLP and XGB, FINC for RF, XGB in 6M, and LSTM. Sectors belonging to the third quadrant of the worst-case scenario are COND and MTRS. Especially, MTRS is located in the third quadrant in all conditions, whereas COND is located in the third quadrant of RF in 3M, XGB in 3M, and LSTM. Note that CONS is mostly located in the indifferent region except for XGB in 6M.

In conclusion, the first quadrant sector consistently outperforms the other sec-

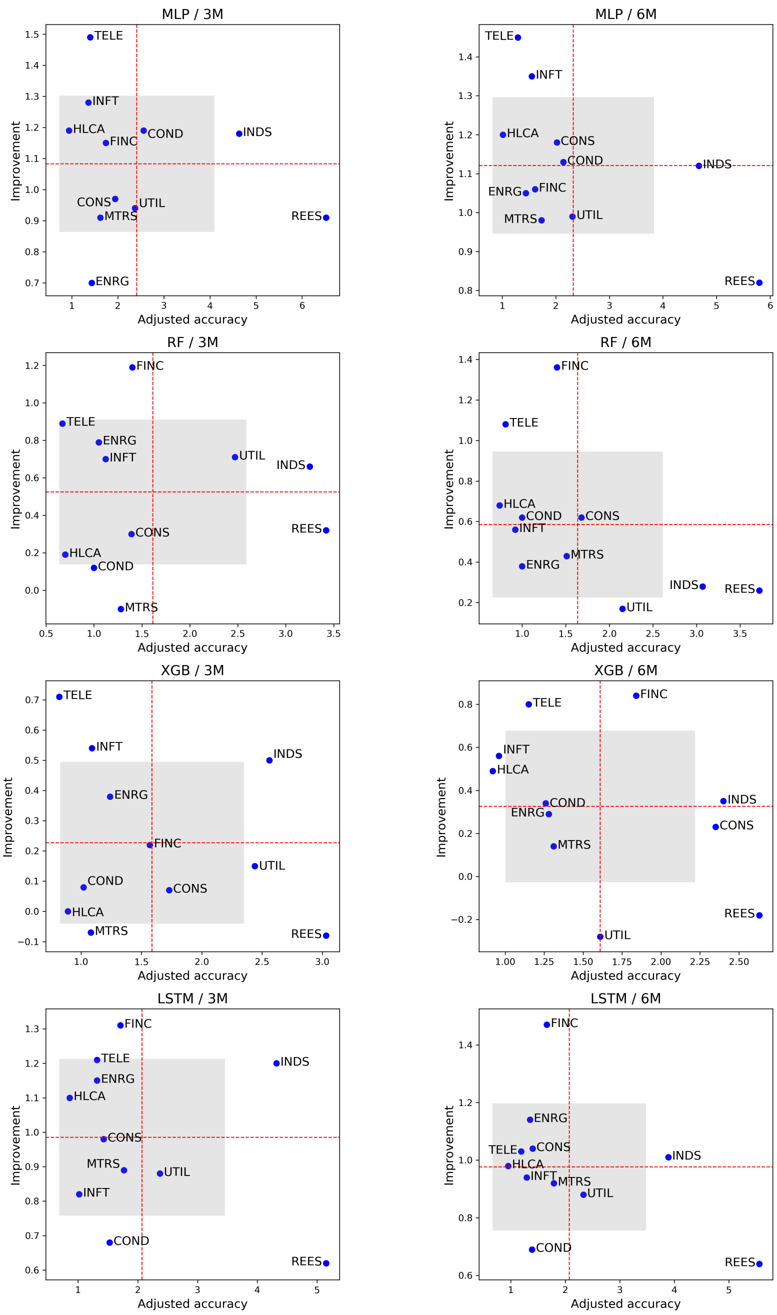


Figure 4.5: Prediction performance vs. prediction improvement

tors in all cases, while the third quadrant sector consistently performs poorly. Thus, we claim that there is a more suitable sector for the application of ETE network indicators in predicting the direction of the stock.

4.4 Summary and Discussion

In this chapter, the evolution of ETE network indicators is utilized as an input variable in predicting the direction of future stock price based on the LR, MLP, RF, XGB, and LSTM. For five machine learning algorithms, we set up a parameter set of 3M and 6M moving windows ETE, 3M, 6M, and 1Y of lag and 1W, 1M, and 3M of the prediction period. For each dataset, we compare the prediction performances of the mixture of log-returns and ETE network indicator against that of plain log-returns in terms of models and sectors through adjusted accuracy.

At first, the average prediction adjusted accuracy of all models for different moving windows, lags, and prediction periods are analyzed. The improvement of prediction performance by utilizing the ETE network indicator is detected in all parameter sets. Notably, as the lag increases, the prediction adjusted accuracy, in general, tends to decrease slightly and the prediction adjusted accuracy tends to decrease in the long term rather than the short term.

Secondly, the prediction performances for different machine learning algorithms are analyzed. From the perspective of adjusted accuracy, we verify that all five machine learning algorithms have improved the prediction performance through the ETE network indicator and suggest that the MLP and LSTM are the most suitable models for predicting future stock price direction predictions considering the adjusted accuracy and paired t-test.

Lastly, the prediction performances for each sector are analyzed. Through the numerical and scatter plot analysis about adjusted accuracy and its improvement by ETE network indicators for MLP, RF, XGB, LSTM of each sector, we claim that there is a more suitable sector for the application of ETE network indicators in predicting the direction of the stock.

In conclusion, the ETE network indicators can be used to improve the performance of the stock price direction for all cases of the LR, MLP, RF, XGB, and LSTM. Notably, the MLP and LSTM as more suitable machine learning algorithms in predicting the direction of stock price based on the adjusted accuracy. After all, the application of Black-Litterman model based on the characteristics of the ETE network indicator and the prediction of stock price direction using machine learning algorithms integrating the ETE network indicator will be discussed in Chapter 5.

Chapter 5

The Black-Litterman model for ETE and machine learning

5.1 The Black-Litterman model

The Black-Litterman model(Black and Litterman, 1990) is proposed to overcome the problems that investors face when they actually apply modern portfolio theory to establish an investment strategy. The Markowitz's mean-variance portfolio is determined solely by expected returns and asset covariance matrices, resulting in significant changes in asset weight despite small changes in these two parameters. The Black-Litterman model introduces various parameters to alleviate this problem and reduce the sensitivity to parameters through the reflection of the investor's view.

In this dissertation, the Black-Litterman model with a bayesian approach is used(Pyo and Lee, 2018). The Black-Litterman model starts with implied excess equilibrium returns which are the expected return based on the information when the investor has prior information about the optimal portfolio obtained from the following quadratic utility maximization problem: $max_w \ \mathbf{w}^T\boldsymbol{\pi} - \frac{1}{2}\lambda\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$.

$$\boldsymbol{\pi} = \lambda\boldsymbol{\Sigma}\mathbf{w}_{mkt} \tag{5.1}$$

where $\boldsymbol{\pi} \in \mathbb{R}^{n\times 1}$, $\boldsymbol{\Sigma} = E[(\mathbf{r}-\boldsymbol{\mu})(\mathbf{r}-\boldsymbol{\mu})^T]$ where mean of excess returns $\boldsymbol{\mu} =$

$E[\mathbf{r}]$, $\mathbf{w}_{mkt} \in \mathbb{R}^{n\times 1}$ are the implied excess equilibrium return vector $\mathbf{r}$, the $n \times n$ covariance matrix of excess returns, and the market capitalization weights of the assets, respectively.

The risk aversion coefficient λ represents the risk premium rate which is an appropriate excess return per unit of risk considered by the investors. The greater the risk aversion coefficient, the greater the excess return required per unit of risk. λ can be obtained by multiplying both sides of (5.1) by $\mathbf{w}_{mkt}^T$ as

$$\lambda = (\bar{r}_{mkt} - r_f)/\sigma^2 \tag{5.2}$$

where $\bar{r}_{mkt} = \mathbf{w}_{mkt}^T \boldsymbol{\pi} + r_f$, r_f, $\sigma^2 = \mathbf{w}_{mkt}^T \boldsymbol{\Sigma} \mathbf{w}_{mkt}$ are the total expected return on the market portfolio, the risk free rate, and the variance of the market portfolio, respectively.

When the prior expected excess returns $\boldsymbol{\mu}$ are as follows,

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\pi}, \tau\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp[-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\pi})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\pi})] \tag{5.3}$$

where τ is a scalar, the risk-adjustment constant. The view vector of returns $\mathbf{q} \in \mathbb{R}^k$ assumed in the Black-Litterman model is a random vector having the following conditional distribution.

$$\mathbf{q}|\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{P}\boldsymbol{\mu}, \boldsymbol{\Omega}) \tag{5.4}$$

where $\mathbf{P}$, $\boldsymbol{\Omega}$ are the $k \times n$ matrix of the asset weights within each view and a diagonal $k \times k$ matrix of the covariance of the views representing uncertainty of the view.

There are two ways to apply view portfolio weights: (1) absolute view that the sum of the weight is 0 and (2) relative view that the sum of the weight is 1.

The marginal distribution for $\mathbf{q}$ is derived from annotation [1].

$$\mathbf{q} \sim \mathcal{N}(\mathbf{P}\boldsymbol{\pi}, \boldsymbol{\Omega} + \mathbf{P}(\tau\boldsymbol{\Sigma})\mathbf{P}^T) \tag{5.7}$$

Then, the new posterior combined expected excess returns are as follows.

$$\boldsymbol{\mu}_{posterior} \sim \mathcal{N}(\boldsymbol{\mu}_{BL}, \boldsymbol{\Sigma}_{BL}) \tag{5.8}$$

where

$$\boldsymbol{\Sigma}_{BL} = [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1} \tag{5.9}$$

$$\boldsymbol{\mu}_{BL} = \boldsymbol{\Sigma}_{BL}[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{q}] \tag{5.10}$$

$$= \boldsymbol{\pi} + \tau\boldsymbol{\Sigma}\mathbf{P}^T[\mathbf{P}(\tau\boldsymbol{\Sigma})\mathbf{P}^T + \boldsymbol{\Omega}]^{-1}(\mathbf{q} - \mathbf{P}\boldsymbol{\pi}) \tag{5.11}$$

The equation (5.9) and (5.10) can be derived from (5.6). The equation (5.11)

[1] When a marginal Gaussian distribution for $\mathbf{x}$ and a conditional Gaussian distribution $\mathbf{y}$ given $\mathbf{x}$ is

$$p(\mathbf{x}) \sim \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}_x)$$
$$p(\mathbf{y}|\mathbf{x}) \sim \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \boldsymbol{\Sigma}_{y|x})$$

The marginal Gaussian distribution for $\mathbf{y}$ and a conditional Gaussian distribution $\mathbf{x}$ given $\mathbf{y}$ are

$$p(\mathbf{y}) \sim \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \boldsymbol{\Sigma}_{y|x} + \mathbf{A}\boldsymbol{\Sigma}_x\mathbf{A}^T) \tag{5.5}$$
$$p(\mathbf{x}|\mathbf{y}) \sim \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}_{x|y}(\mathbf{A}^T\boldsymbol{\Sigma}_{y|x}^{-1}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Sigma}_x^{-1}\boldsymbol{\mu}), \boldsymbol{\Sigma}_{x|y}) \tag{5.6}$$

where $\boldsymbol{\Sigma}_{x|y}^{-1} = (\boldsymbol{\Sigma}_x^{-1} + \mathbf{A}^T\boldsymbol{\Sigma}_{y|x}^{-1}\mathbf{A})$

can be derived by applying the following the Sherman-Woodbury-Morrison formula.

$$(\mathbf{A}+\mathbf{BD}^{-1}\mathbf{C})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}+\mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} \tag{5.12}$$

5.2 Data and experiment set-ups

5.2.1 Data

The classification of industry and its constituent stocks are based on *The MSCI(Morgan Stanley Capital International) USA IMI(Investable Market Index) Sector Indexes* as of September 30, 2019. In this chapter, total return index (RI), which is used instead of the stock price and market value (MV) are used. RI is an index that not only tracks stock price changes but also measures the performance of a group of stocks, assuming that all cash distributions have been reinvested. The price index only considers the price fluctuations such as capital gains or losses of the stocks that make up the index, while the RI includes dividends, interests, rights offerings, and other distributions realized during a given period of time. RI is generally considered a more accurate measure of performance.

According to the data reference date, the existing sector name of 'Telecommunication Services' has been changed to the 'Communication Services', but in this dissertation, the 'Communication services' is referred to as 'Telecommunication Services' for consistency. In addition, there are changes in two stocks among 55 stocks used in Chatper 3 and 4 depending on the data reference date. 'Activision Blizzard, Inc.' (ATVI) instead of FOXA in TELE, and 'Linde plc' (LIN) instead of DWDP in MTRS are reflected. The data period is the same as that used in Chapter 3 and 4.

5.2.2 Experiment set-ups

As mentioned in Chapter 3, the two properties of ETE are identified: (1) ETE tends to rise in the financial crisis periods and (2) In the financial crisis periods, the difference of outflow and inflow ETE of the sector related to the financial crisis shows the most positive value. In Chapter 4, the improvement of prediction adjusted accuracy by integrating the ETE network indicators are detected in all cases of the LR, MLP, RF, XGB, and LSTM, and the MLP and LSTM are suitable models for predicting the stock price direction using ETE network indicators. These results are applied to the investor's view of the Black-Litterman model.

At first, from the perspective of prediction of the stock price direction using machine learning models, the adjusted accuracy and its improvement of the stock price direction prediction by the ETE network indicator are verified against the new data sets of MLP and LSTM. As in the previous chapter, the prediction periods are set for 1 week (1W, 5 days), 1 month (1M, 20 days), and 3 months (3M, 60 days) and applies equally to the rebalancing period (RP). The application of ETE moving windows is only for 60 days moving windows ETE due to the slight difference in performance between 60 and 120 days. The lag is applied as 3M lags considering the tendency for accuracy to improve as the number of lags decreases. The rest of the experimental framework such as training methods and define prediction performance is the same as in Chapter 4. The parameter pairs used in machine learnings are summarized in Table 5.1.

Table 5.1: Parameter set-ups for machine learning algorithms

Model	**Parameters**	**Levels**
MLP	Neuron of first layer	256, 128, 64, 32,16
LSTM	Hidden unit	2, 4, 8, 16, 32

When applying the investor's view of the Black-Litterman model, there are two investor's views as follows.

- **View 1: Absolute View**

  In the prediction of the stock price direction through a machine learning algorithms, for each prediction period, the stocks predicted by positive (negative) cumulative returns will rise (fall) by **q**%.

- **View 2: Relative View**

  Stocks with a negative (positive) difference of outflow and inflow ETE are expected to outperform stocks with a positive (negative) by **q**%.

The methodology of the 'View 1' applies the results predicted only by its plain log-returns and the mixture of log-returns and the ETE network indicator using machine learning models to absolute view of Black-Litterman model. In the stock price direction prediction result through machine learning models, both MLP and LSTM generate up/down results of five parameter sets, and only stocks in which at least three out of five predictions predict the same direction is reflected in the view vector in order to increase the reliability of the investor's view vector.

$$P_{abs} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{5.13}$$

where P_{abs} is an example of an absolute investor's view. When a column is a stock, the first, third, and fourth stocks are applied to investor's view.

The methodology of the 'View 2' considering the problem that the relative view effect cannot be properly reflected due to a large number of absolute views starts from the optimal weight obtained from results predicted only by its plain log-returns to avoid overlapping ETE effects and applies to relative view according to the difference of outflow and inflow ETE. The market capitalization weighting scheme is used for the method of applying the relative view according to Idzorek (2007). For stocks with positive (negative) difference of outflow and inflow ETE, the relative view of each positive (negative) stock is applied as follows; the market cap of each positive (negative) stock divided by the sum of market cap of a positive (negative) stocks. In order to select stocks that are influential in the information flow, only the top 70% of the number of stocks with a positive (negative) difference of outflow and inflow ETE are applied to the relative view.

$$P_{rel} = \begin{pmatrix} 0.3 & 0.5 & 0 & -0.2 & -0.8 & 0 \end{pmatrix} \tag{5.14}$$

where P_{rel} is an example of an relative investor's view. When a column is a stock, the first and second stocks are positive difference of outflow and inflow ETE and the fourth and fifth stocks are negative.

The rest of the detailed parameters except the investor's view of the black-litre model are as follows. The normal return is used as unit of yield. The past period, which is the reflection period of the past to derive the risk aversion coefficient, is set to 1 year (1Y, 240 days). The scalar τ is set to 0.025 by referring to previous studies of He and Litterman (2002). The expected rate of return $\mathbf{q}$ is set as the average rate of return of stocks in the past 240 days at the time point plus its standard deviation.The objective function is taken as the Sharpe ratio. The uncertainty of view $\boldsymbol{\Omega}$ is as follows.

$$\Omega = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_k \end{pmatrix} = \begin{pmatrix} (p_1 \Sigma p_1')\tau & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & (p_k \Sigma p_k')\tau \end{pmatrix} \tag{5.15}$$

For convenience of description, we define an abbreviation for the portfolio of the Black-Litterman model having a certain investor's view as follows. The Black-Litterman portfolio with the investor's view of applying the prediction performance by its plain log-returns is abbreviated to 'woETE' and the mixture of log-returns and the ETE network indicator to 'wETE'. And the Black-Litterman portfolio for the investor's view of applying the return of the stocks with a negative (positive) difference of outflow and inflow ETE is expected to outperform that of the stocks with a positive (negative) are abbreviated to 'dETE(N)>dETE(P)' and 'dETE(P)>dETE(N)', respectively.

The performances of the Black-Litterman model with all these four investor's views are compared with the performance of the market portfolio that weights according to the market value and market index ‘S&P 500’. The performance of the portfolios is evaluated through the cumulative returns for the rebalancing period from the first day of the test set, that is, the cumulative return of portfolios.

The performance metric of the portfolio is defined as follows. From the perspective of the cumulative return of portfolio, cumulative return, Sharpe ratio, and Sortino ratio are used. The process of deriving performance measures through portfolio establishment from stock price direction prediction using MLP and LSTM is repeated 10 times to verify the average and standard deviation for adjusted accuracy and portfolio performance. Finally, the overall framework can be summarized as a step-by-step procedure described in Figure 5.1.

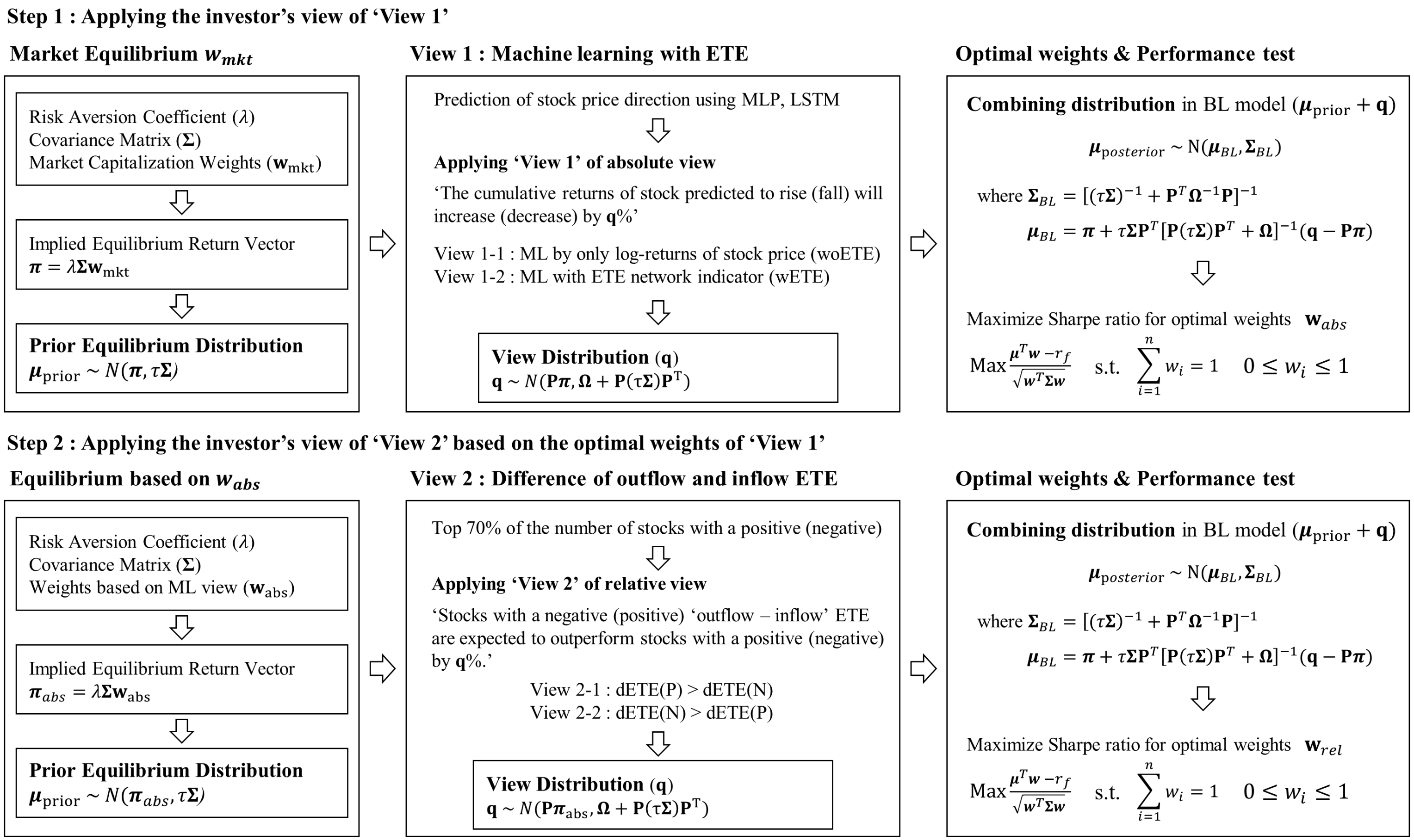


Figure 5.1: Procedure for the Black-Litterman model with ETE and machine learning

5.3 Results

5.3.1 Prediction performance in different models and sectors

The values of outflow ETE minus inflow ETE is shown in Figure 5.2 and Table 5.2 summarizes adjusted accuracy for MLP and LSTM, prediction periods, and integration of ETE network indicator.

Note that Figure 5.2 for the evolution of outflow minus inflow ETE and Figure 3.8 in Chapter 3 shows a similar trend across all sectors. In Table 5.2, although RI is used instead of the stock price, it can be seen that the tendency of adjusted accuracy is not significantly different from the results of the stock prices. Also, in Table 5.3, for all prediction periods and models, the null hypothesis of the paired t-test that the adjusted accuracy of two cases are equal is rejected, confirming that the adjusted accuracy is statistically improved through ETE network indicator.

Table 5.4 summarizes the adjusted accuracy and its improvement of each sector for MLP and LSTM, prediction periods, and integration of ETE network indicator. The adjusted accuracy is improved through the ETE network indicator in all prediction periods except ENRG in the MLP's 3M prediction period. There are sectors with similar adjusted accuracy as the prediction period increases whereas others have decreased adjusted accuracy as the prediction period increases.

5.3.2 Portfolio performances for cumulative return

In this section, the performances of an actual investment strategy based on the proposed portfolios are evaluated. Table 5.5 summarizes the cumulative return, Sharpe ratio, and Sortino ratio for each rebalancing period, model and portfolios.

At first, there is no significant difference in cumulative return according to

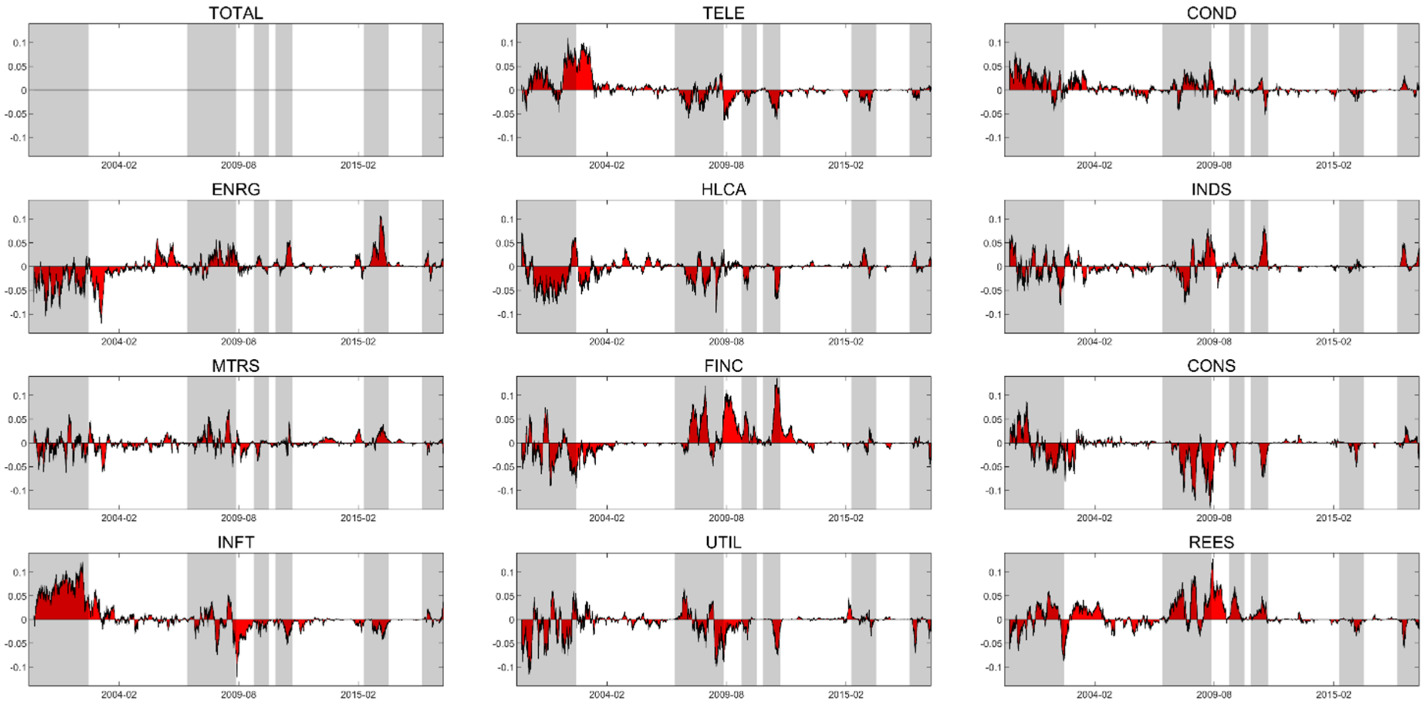


Figure 5.2: Evolution of ‘outflow - inflow’ ETE in different sectors

Table 5.2: Adjusted accuracy for machine learning models

Model		**With ETE**			**Without ETE**			**Improvement**		
		1W	1M	3M	1W	1M	3M	1W	1M	3M
MLP	Mean	1.87	1.76	1.38	1.79	1.76	1.30	1.11	0.80	0.71
	Std	0.097	0.093	0.108	0.115	0.141	0.102	0.143	0.118	0.117
LSTM	Mean	1.95	1.55	1.32	1.47	1.41	1.14	0.65	0.63	0.68
	Std	0.250	0.116	0.083	0.148	0.058	0.068	0.078	0.122	0.053

*Note: With ETE, Without ETE, Improvement, Mean, Std are the abbreviations of the prediction performance when predicted by the mixture of log-return of RI and the ETE network indicator, the prediction performance when predicted only by its plain log-return of RI, the improvement of prediction adjusted accuracy incurred by ETE network indicator, the mean and standard deviation of the adjusted accuracy performed 10 times, respectively.

Table 5.3: Paired t-test for machine learnings

Model	**p-value**		
	1W	1M	3M
MLP	0.000	0.000	0.000
LSTM	0.000	0.000	0.000

rebalancing period or model. The woETE and wETE does not show any tendency according to the rebalancing period, and the portfolios of dETE tend to increase the cumulative return as the rebalancing period increases.

From the perspective of the investor's view of machine learning with ETE, The wETE shows higher performance in cumulative return, Sharpe ratio, and Sortino ratio than the woETE including benchmarks. This indicates that the improvement in the stock price direction prediction adjusted accuracy due to the ETE network indicator also results in an improvement in the yield of the actual portfolio through the Black-Litterman model. Notably, the wETE performances of LSTM are better than those of MLP in all rebalancing period except 1W rebalancing period.

From the perspective of the investor's view of the ETE network indicator, dETE(N)>dETE(P) shows the best cumulative return in all rebalancing period. Meanwhile, dETE(P)>dETE(N) shows the worst performance for all rebalancing period. This supports the results of chapter 3, in which the difference of outflow and inflow ETE of stocks associated with the crisis in the crisis period is positive.

The cumulative returns for each rebalancing period of the portfolios are described in Figure 5.3, 5.4, and 5.5. Note that the cumulative return is defined that the average cumulative return of portfolios for 10 trials. For all rebalancing periods, it can be seen that the cumulative return of the wETE is larger than the woETE in

Table 5.4: Adjusted accuracy and its improvement in different sectors

Model		**MLP**						**LSTM**					
Sector		**Acc (adj)**			**Acc (adj) Imp**			**Acc (adj)**			**Acc (adj) Imp**		
		1W	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**	**1W**	**1M**	**3M**
TELE	Mean	1.35	1.88	1.55	1.67	1.25	1.22	1.56	2.23	1.17	1.22	1.15	0.69
	Std	0.499	0.444	0.652	0.567	0.532	0.487	0.655	0.238	0.627	0.28	0.374	0.165
COND	Mean	2.34	3.40	1.97	1.38	1.01	1.23	3.00	1.23	1.85	0.57	0.19	0.96
	Std	0.904	2.314	0.549	0.834	0.687	0.165	1.551	1.176	0.676	0.335	0.346	0.317
ENRG	Mean	2.02	1.21	0.79	1.23	0.73	-0.04	2.25	1.16	0.87	0.87	1.03	1.47
	Std	0.342	0.152	0.068	0.42	0.468	0.635	0.175	0.203	0.084	0.354	0.242	0.514
HLCA	Mean	1.62	0.90	0.56	1.98	0.77	0.87	1.77	0.86	0.52	0.93	0.95	0.92
	Std	0.488	0.259	0.123	0.993	0.256	0.151	1.022	0.174	0.049	0.272	0.313	0.151
INDS	Mean	4.60	3.91	3.37	1.25	0.99	0.70	8.90	4.71	2.44	0.73	0.74	0.89
	Std	1.754	1.299	1.428	0.669	0.266	0.321	2.762	1.09	1.155	0.396	0.335	0.441
MTRS	Mean	1.79	1.75	1.80	1.28	1.07	0.50	1.56	1.82	2.16	0.57	1.22	0.97
	Std	0.167	0.077	0.065	0.369	0.521	0.515	0.067	0.184	0.304	0.145	0.464	0.564
FINC	Mean	2.02	2.44	1.40	1.20	1.22	1.15	2.08	2.34	1.57	0.89	0.97	1.46
	Std	0.34	0.445	0.112	0.413	0.507	0.319	0.335	0.532	0.259	0.221	0.496	0.576
CONS	Mean	1.81	1.84	2.05	1.65	0.93	0.74	2.18	1.38	1.40	0.70	1.08	0.99
	Std	0.421	0.476	0.298	0.723	0.19	0.227	0.771	0.132	0.297	0.341	0.242	0.259
INFT	Mean	1.48	2.43	0.94	1.34	1.21	1.26	1.53	2.06	1.19	1.07	0.66	0.86
	Std	0.781	0.568	0.246	0.713	0.194	0.561	0.718	1.042	0.289	0.233	0.236	0.278
UTIL	Mean	2.99	2.16	1.98	1.19	1.14	0.70	3.05	2.16	1.98	0.73	1.16	0.76
	Std	0.148	0.006	0.004	0.43	0.509	0.252	0.056	0.012	0.023	0.175	1.19	0.569
REES	Mean	5.04	7.57	3.18	1.12	0.52	0.25	4.99	6.84	3.12	0.70	0.43	0.25
	Std	1.491	0.932	0.416	0.51	0.534	0.364	0.94	1.367	0.86	0.19	0.438	0.225

*Note: Acc (adj), Acc (adj) Imp, Mean, Std are the abbreviations of the adjusted accuracy, improvement of prediction adjusted accuracy incurred by ETE network indicator, the mean and standard deviation of the adjusted accuracy performed 10 times, respectively.

Table 5.5: Performances of the Black-Litterman portfolios

Model	Portfolio		Cum return			Sharpe ratio			Sortino ratio		
			1W	1M	3M	1W	1M	3M	1W	1M	3M
MLP	woETE	Mean	3.769	3.654	3.576	1.244	1.210	1.196	1.641	1.585	1.573
		Std	0.060	0.075	0.065	0.015	0.017	0.018	0.024	0.023	0.025
	wETE	Mean	3.862	3.630	3.652	1.266	1.210	1.213	1.670	1.585	1.596
		Std	0.141	0.090	0.060	0.034	0.023	0.014	0.046	0.031	0.019
	dETE(P)>dETE(N)	Mean	1.527	1.842	2.702	0.314	0.414	0.613	0.433	0.560	0.835
		Std	0.010	0.011	0.014	0.003	0.003	0.003	0.004	0.004	0.004
	dETE(N)>dETE(P)	Mean	4.214	4.172	4.810	0.874	0.889	0.994	1.169	1.195	1.367
		Std	0.273	0.033	0.022	0.036	0.005	0.003	0.050	0.007	0.004
LSTM	woETE	Mean	3.671	3.625	3.659	1.218	1.210	1.218	1.601	1.584	1.599
		Std	0.098	0.067	0.044	0.024	0.015	0.014	0.035	0.019	0.021
	wETE	Mean	3.738	3.688	3.686	1.226	1.223	1.228	1.624	1.606	1.617
		Std	0.142	0.077	0.057	0.033	0.020	0.016	0.043	0.033	0.023
	dETE(P)>dETE(N)	Mean	1.522	1.850	2.697	0.313	0.416	0.612	0.431	0.563	0.834
		Std	0.007	0.010	0.014	0.002	0.003	0.003	0.003	0.004	0.003
	dETE(N)>dETE(P)	Mean	4.119	4.179	4.892	0.859	0.890	1.005	1.150	1.196	1.380
		Std	0.211	0.021	0.048	0.027	0.003	0.006	0.038	0.004	0.007
Benchmark	Market port.	Mean	3.415	3.437	3.411	1.002	1.005	0.998	1.284	1.289	1.280
		Std	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	S&P 500	Mean	3.111	3.111	3.111	0.882	0.882	0.882	1.116	1.116	1.116
		Std	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

*Note: Cum return is the abbreviations of the cumulative return of the portfolio.

most time periods. In addition, the dETE(N)>dETE(P) shows the best cumulative return with high volatility whereas the dETE(P)>dETE(N) shows the worst cumulative return.

5.4 Summary and Discussion

In this chapter, the result of machine learning with ETE and the feature of ETE network indicator are conducted to the investor's view of the Black-Litterman model. The prediction results of the stock price direction of MLP and LSTM that have been verified to be suitable models for stock price prediction in the previous chapter and the sign of difference of outflow and inflow ETE are used to the investor's view. Accordingly, the Black-Litterman portfolios with four investor's views are compared with market portfolio and market index. The RI is used for analysis data instead of stock price to confirm more accurate performance.

At first, the analysis for the evolution of ETE and the prediction performance of the stock price direction using machine learning are conducted. Although RI is used instead of the stock price, it can be seen that the tendency of the adjusted accuracy is not significantly different from the those of stock prices.

Next, the evaluation of the performance of the proposed portfolios is analyzed. The cumulative return, Sharpe ratio, and Sortino ratio are used as evaluation metrics. Note that the wETE is larger than woETE including benchmarks. In terms of the investor's view of ETE, the dETE(N)>dETE(P) shows lower in the Sharpe ratio compared to the wETE and woETE including some cases of benchmarks, but it shows the best cumulative return. The dETE(P)>dETE(N) shows the worst per-

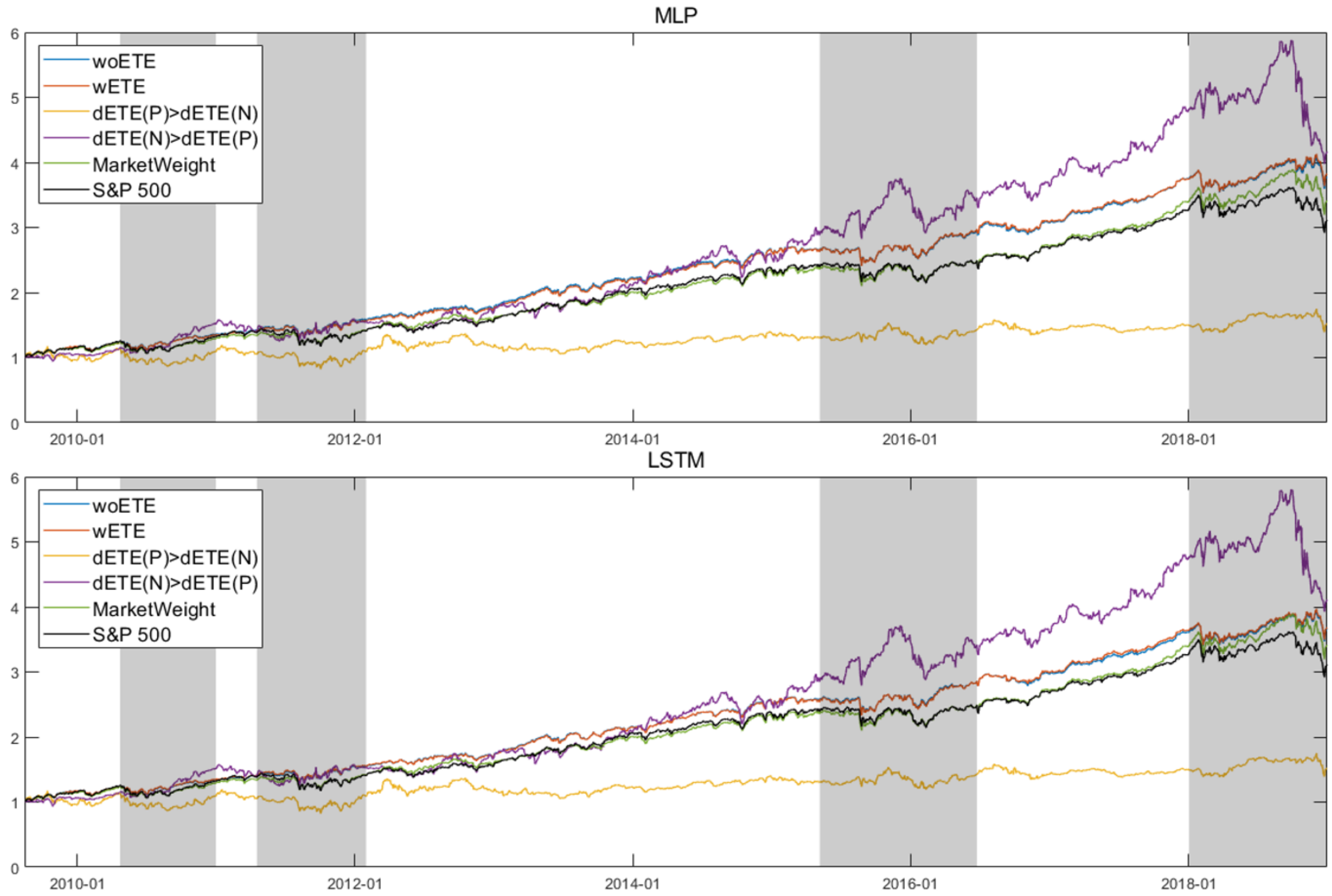


Figure 5.3: Portfolio of cumulative return for rebalancing period (1W)

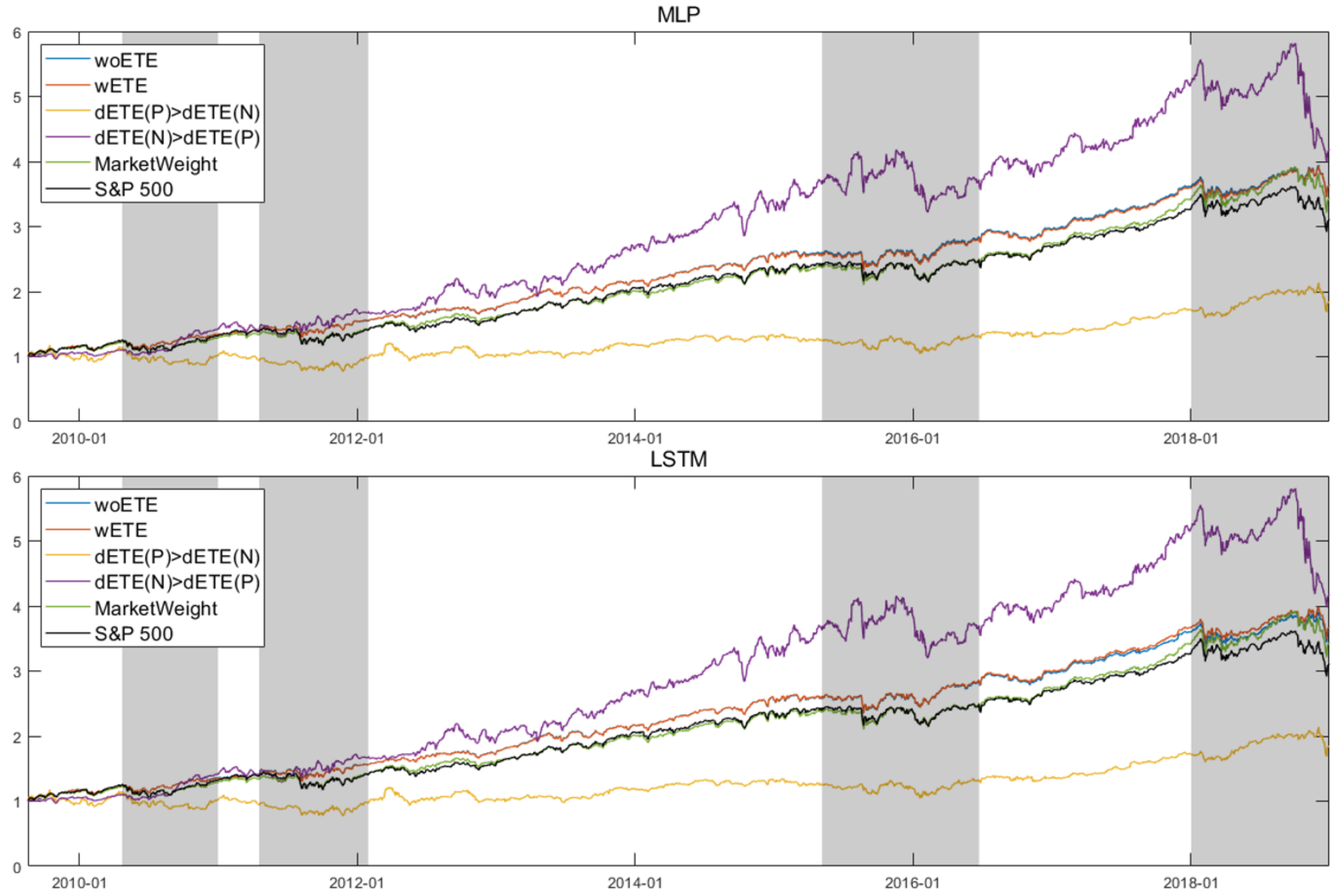


Figure 5.4: Portfolio of cumulative return for rebalancing period (1M)

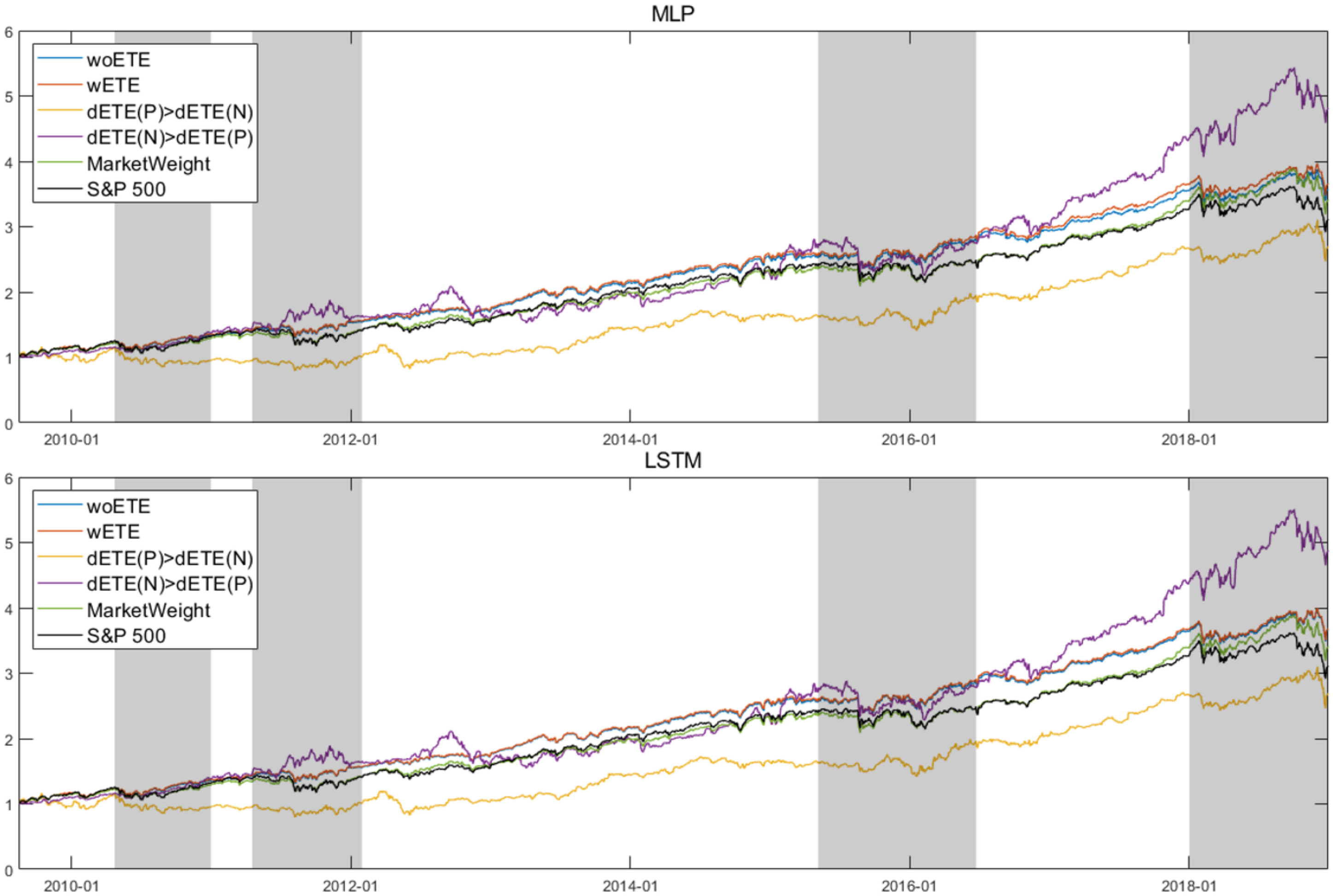


Figure 5.5: Portfolio of cumulative return for rebalancing period (3M)

formance for all parameter sets.

In conclusion, through the Black-Litterman model with the investor's view of machine learning, the use of the prediction results of the stock price direction leads to an improvement in returns, and the possibility of applying the ETE network indicator to investor's view in the Black-Litterman model is confirmed.

Chapter 6

Conclusion

6.1 Contributions and Limitations

Throughout the dissertation, we analyze the statistical explanatory power among major US stocks according to financial events based on the time-varying ETE using the moving window method. Then, we utilize and examine the ETE network indicator as a feature to improve the prediction performance of the direction of the stock price through machine learning algorithms and optimal asset allocation using the Black-Litterman model.

Many previous studies are describing statistical explanatory power between elements in the financial system using TE, which can identify asymmetry information flow between components. In this context, the ETE used in this dissertation is an advanced method for controlling noise in measuring the information flow, which is a disadvantage of TE. Furthermore, the previous research on TE has focused on analyzing market phenomena in connection with information flows, whereas this dissertation focuses on a more practical question such as the utilization of ETE in the financial market. Thus, the novelty of this dissertation lies in the fact that, in our best knowledge, this is the first attempt to integrate the ETE of an individual US stock to analyze the US market and to predict the direction of the stock price.

It is also the first attempt to apply these results to the investor's view of the Black-Litterman model.

The findings of this dissertation can be summarized as follows. At first, we discover that the time-varying ETE based on the 3M and 6M moving windows have market explanatory power using 55 stocks from 11 sectors and six cases of financial crises in the US financial market. Secondly, we detect the increases in the statistical explanatory power of sectors related to the financial crisis and the absolute size of information flow in the market. Thirdly, the utilization of ETE network indicators as new features improves the prediction performance on the stock price direction for all cases of the LR, MLP, RF, XGB and LSTM. Fourthly, we identify the MLP and LSTM as more suitable machine learning algorithms in predicting the direction of stock price based on the adjusted accuracy, newly introduced performance measure based on the concept of risk-adjusted return. Notably, we reveal the suitable sectors for the utilization of ETE network indicators. Lastly, the portfolios of the Black-Litterman model applying the results of stock price direction prediction using machine learnings with the ETE network indicators outperform the market portfolio and market index in terms of the return on risk. From the perspective of practical application of this study, it is meaningful that the prediction of stock price direction using machine learning with the ETE network indicators actually led to an increase in the profit on investment through the Black-Litterman model.

The limitation of this research that should be addressed in future studies is the computation time to obtain the time-varying ETE and multiple results of machine learnings. Especially, it is necessary to optimize the ETE computation algorithm to apply the mechanism to all stocks in S&P 500 broadly.

6.2 Future Work

It is required to increase the number of market constituents by expanding the computational power to derive and analyze ETE. In addition, advanced tuning of machine learning parameters is necessary for good performance in stock price prediction. As confirming the practical usage of ETE in the market through prediction of stock price direction and asset allocation, developing an alarm index to detect market crisis using ETE would be a future work.

Bibliography

Arulraj, M., Pvs, M., and Karthika, R. (2012). Global portfolio optimization for bse sensex using the enhanced black-litterman model. *Procedia engineering*, 38:2987–2997.

Awajan, A. M., Ismail, M. T., and Wadi, S. A. (2018). Improving forecasting accuracy for stock market data using emd-hw bagging. *PloS one*, 13(7):e0199582.

Ballings, M., Van den Poel, D., Hespeels, N., and Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 42(20):7046–7056.

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

Beach, S. L. and Orlov, A. G. (2007). An application of the black–litterman model with egarch-m-derived views for international portfolio management. *Financial Markets and Portfolio Management*, 21(2):147–166.

Bisoi, R., Dash, P., and Parida, A. (2019). Hybrid variational mode decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis. *Applied Soft Computing*, 74:652–678.

Black, F. and Litterman, R. (1990). Asset allocation: combining investor views with market equilibrium. *Goldman Sachs Fixed Income Research*, 115.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial analysts journal*, 48(5):28–43.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Bonanno, G., Caldarelli, G., Lillo, F., and Mantegna, R. N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4):046130.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Brown, I. and Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3):3446–3453.

Caraiani, P. (2017). The predictive power of local properties of financial networks. *Physica A: Statistical Mechanics and its Applications*, 466:79–90.

Chan, S. W. and Franklin, J. (2011). A text-based decision support system for financial sequence prediction. *Decision Support Systems*, 52(1):189–198.

Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., and Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Systems with Applications*, 112:353–371.

Chen, K., Zhou, Y., and Dai, F. (2015). A lstm-based method for stock returns prediction: A case study of china stock market. In *Big Data (Big Data), 2015 IEEE International Conference on*, pages 2823–2824. IEEE.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Chen, Y. and Hao, Y. (2018). Integrating principle component analysis and weighted support vector machine for stock trading signals prediction. *Neurocomputing*, 321:381–402.

Cheng, C.-H. and Yang, J.-H. (2018). Fuzzy time-series model based on rough set rule induction for forecasting stock price. *Neurocomputing*, 302:33–45.

Chi, K. T., Liu, J., and Lau, F. C. (2010). A network perspective of the stock market. *Journal of Empirical Finance*, 17(4):659–667.

Chong, E., Han, C., and Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, 83:187–205.

Chunxia, Y., Xueshuai, Z., Luoluo, J., Sen, H., and He, L. (2016). Study on the contagion among american industries. *Physica A: Statistical Mechanics and its Applications*, 444:601–612.

Clements, M. P., Franses, P. H., and Swanson, N. R. (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, 20(2):169–183.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–232.

Creamer, G. G. (2015). Can a corporate network and news sentiment improve portfolio optimization using the black–litterman model? *Quantitative Finance*, 15(8):1405–1416.

Darbellay, G. A. and Wuertz, D. (2000). The entropy as a tool for analysing statistical dependences in financial time series. *Physica A: Statistical Mechanics and its Applications*, 287(3-4):429–439.

Davis, M. and Lleo, S. (2020). Debiased expert forecasts in continuous-time asset allocation. *Journal of Banking & Finance*, 113:105759.

Dewandaru, G., Masih, R., Bacha, O. I., and Masih, A. M. M. (2015). Combining momentum, value, and quality for the islamic equity portfolio: Multi-style rotation strategies using augmented black litterman factor model. *Pacific-Basin Finance Journal*, 34:205–232.

Dimpfl, T. and Peter, F. J. (2013). Using transfer entropy to measure information flows between financial markets. *Studies in Nonlinear Dynamics and Econometrics*, 17(1):85–102.

Faes, L., Nollo, G., and Porta, A. (2013). Compensated transfer entropy as a tool for reliably estimating information transfer in physiological time series. *Entropy*, 15(1):198–219.

Fang, Y., Bo, L., Zhao, D., and Wang, S. (2018). Fuzzy views on black-litterman portfolio selection model. *Journal of Systems Science and Complexity*, 31(4):975–987.

Fernandes, B., Street, A., Fernandes, C., and Valladão, D. (2018). On an adaptive black–litterman investment strategy using conditional fundamentalist information: A brazilian case study. *Finance Research Letters*, 27:201–207.

Feuerriegel, S. and Gordon, J. (2018). Long-term stock index forecasting based on text mining of regulatory disclosures. *Decision Support Systems*, 112:88–97.

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

Freedman, D. A. (2009). *Statistical models: theory and practice*. cambridge university press.

García-Medina, A., Junior, L. S., Bañuelos, E. U., and Martínez-Argüello, A. (2018). Correlations and flow of information between the new york times and stock markets. *Physica A: Statistical Mechanics and its Applications*, 502:403–415.

Geva, T. and Zahavi, J. (2014). Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news. *Decision support systems*, 57:212–223.

Ghysels, E. and Osborn, D. R. (2001). *The econometric analysis of seasonal time series*. Cambridge University Press.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 37(3):424–438.

Guo, Z., Wang, H., Yang, J., and Miller, D. J. (2015). A stock market forecasting model combining two-directional two-dimensional principal component analysis and radial basis function neural network. *PloS one*, 10(4):e0122385.

Guresen, E., Kayakutlu, G., and Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 38(8):10389–10397.

Hagenau, M., Liebmann, M., and Neumann, D. (2013). Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems*, 55(3):685–697.

Harris, R. D., Stoja, E., and Tan, L. (2017). The dynamic black–litterman approach to asset allocation. *European Journal of Operational Research*, 259(3):1085–1096.

He, G. and Litterman, R. (2002). The intuition behind black-litterman model portfolios. *Available at SSRN 334304*.

Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M., and Bhattacharya, J. (2007). Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 441(1):1–46.

Ho, T. K. (1995). Random decision forests. In *Document analysis and recognition, 1995., proceedings of the third international conference on*, volume 1, pages 278–282. IEEE.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hu, H., Tang, L., Zhang, S., and Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with google trends. *Neurocomputing*, 285:188–195.

Idzorek, T. (2007). A step-by-step guide to the black-litterman model: Incorporating user-specified confidence levels. In *Forecasting expected returns in the financial markets*, pages 17–38. Elsevier.

Jiang, J. and Gu, R. (2016). Using rényi parameter to improve the predictive power of singular value decomposition entropy on stock market. *Physica A: Statistical Mechanics and its Applications*, 448:254–264.

Jiang, M., Liu, J., Zhang, L., and Liu, C. (2020). An improved stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. *Physica A: Statistical Mechanics and its Applications*, 541:122272.

Jizba, P., Kleinert, H., and Shefaat, M. (2012). Rényi's information transfer between financial time series. *Physica A: Statistical Mechanics and its Applications*, 391(10):2971–2989.

Kao, L.-J., Chiu, C.-C., Lu, C.-J., and Chang, C.-H. (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.

Kara, M., Ulucan, A., and Atici, K. B. (2019). A hybrid approach for generating investor views in black–litterman model. *Expert Systems with Applications*, 128:256–270.

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

Kenett, D. Y., Raddant, M., Lux, T., and Ben-Jacob, E. (2012). Evolvement of uniformity and volatility in the stressed global financial village. *PloS one*, 7(2):e31144.

Kenett, D. Y., Shapira, Y., Madi, A., Bransburg-Zabary, S., Gur-Gershgoren, G., and Ben-Jacob, E. (2011). Index cohesive force analysis reveals that the us market became prone to systemic collapses since 2002. *PLoS one*, 6(4):e19378.

Khansa, L. and Liginlal, D. (2011). Predicting stock market returns from malicious attacks: A comparative analysis of vector autoregression and time-delayed neural networks. *Decision Support Systems*, 51(4):745–759.

Kim, H. Y. and Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models. *Expert Systems with Applications*, 103:25–37.

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

Kim, M., Newth, D., and Christen, P. (2016). Macro-level information transfer in social media: Reflections of crowd phenomena. *Neurocomputing*, 172:84–99.

Kim, M. J., Kwak, Y. B., and Kim, S. Y. (2011). Dependence structure of the korean stock market in high frequency data. *Physica A: Statistical Mechanics and its Applications*, 390(5):891–901.

Kocuk, B. and Cornuéjols, G. (2020). Incorporating black-litterman views in portfolio construction when stock returns are a mixture of normals. *Omega*, 91:102008.

Kolm, P. and Ritter, G. (2017). On the bayesian interpretation of black–litterman. *European Journal of Operational Research*, 258(2):564–572.

Kumar, S. and Deo, N. (2012). Correlation and network analysis of global financial indices. *Physical Review E*, 86(2):026101.

Kwon, O. and Oh, G. (2012). Asymmetric information flow between market index and individual stocks in several stock markets. *EPL (Europhysics Letters)*, 97(2):28007.

Kwon, O. and Yang, J.-S. (2008a). Information flow between composite stock index and individual stocks. *Physica A: Statistical Mechanics and its Applications*, 387(12):2851–2856.

Kwon, O. and Yang, J.-S. (2008b). Information flow between stock indices. *EPL (Europhysics Letters)*, 82(6):68003.

Lee, T. K., Cho, J. H., Kwon, D. S., and Sohn, S. Y. (2019). Global stock market investment strategies based on financial network indicators using machine learning techniques. *Expert Systems with Applications*, 117:228–242.

Lei, L. (2018). Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. *Applied Soft Computing*, 62:923–932.

Lim, K., Kim, S., and Kim, S. Y. (2017). Information transfer across intra/inter-structure of cds and stock markets. *Physica A: Statistical Mechanics and its Applications*, 486:118–126.

Lux, T. and Marchesi, M. (1999). Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498.

Maasoumi, E. and Racine, J. (2002). Entropy and predictability of stock market returns. *Journal of Econometrics*, 107(1-2):291–312.

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

Mantegna, R. N. and Stanley, H. E. (1999). *Introduction to econophysics: correlations and complexity in finance.* Cambridge university press.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

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

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

Molchan, G. (1990). Strategies in strong earthquake prediction. *Physics of the Earth and Planetary Interiors*, 61(1-2):84–98.

Molgedey, L. and Ebeling, W. (2000). Local order, entropy and predictability of financial time series. *The European Physical Journal B-Condensed Matter and Complex Systems*, 15(4):733–737.

Moradi, M., Jabbari Nooghabi, M., and Rounaghi, M. M. (2019). Investigation of fractal market hypothesis and forecasting time series stock returns for tehran stock exchange and london stock exchange. *International Journal of Finance & Economics*, pages 1–17.

M'ng, J. C. P. and Mehralizadeh, M. (2016). Forecasting east asian indices futures via a novel hybrid of wavelet-pca denoising and artificial neural network models. *PloS one*, 11(6):e0156338.

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

Nelson, D. M., Pereira, A. C., and de Oliveira, R. A. (2017). Stock market's price movement prediction with lstm neural networks. In *Neural Networks (IJCNN), 2017 International Joint Conference on*, pages 1419–1426. IEEE.

Nobre, J. and Neves, R. F. (2019). Combining principal component analysis, discrete wavelet transform and xgboost to trade in the financial markets. *Expert Systems with Applications*, 125:181–194.

Noh, J. D. (2000). Model for correlations in stock markets. *Physical Review E*, 61(5):5981.

Ntakaris, A., Mirone, G., Kanniainen, J., Gabbouj, M., and Iosifidis, A. (2019). Feature engineering for mid-price prediction with deep learning. *Ieee Access*, 7:82390–82412.

Ola, M. R., Nooghabi, M. J., and Rounaghi, M. M. (2014). Chaos process testing (using local polynomial approximation model) in predicting stock returns in tehran stock exchange. *Asian Journal of Research in Banking and Finance*, 4(11):100–109.

Oliveira, N., Cortez, P., and Areal, N. (2017). The impact of microblogging data for stock market prediction: Using twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications*, 73:125–144.

Ormos, M. and Zibriczky, D. (2014). Entropy-based financial asset pricing. *PloS one*, 9(12):e115742.

Palczewski, A. and Palczewski, J. (2019). Black–litterman model for continuous distributions. *European Journal of Operational Research*, 273(2):708–720.

Palomba, G. et al. (2008). Multivariate garch models and the black-litterman approach for tracking error constrained portfolios: an empirical analysis. *Global Business and Economics Review*, 10(4):379.

Pang, T. and Karan, C. (2018). A closed-form solution of the black–litterman model with conditional value at risk. *Operations Research Letters*, 46(1):103–108.

Patel, J., Shah, S., Thakkar, P., and Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1):259–268.

Peron, T. D. and Rodrigues, F. A. (2011). Collective behavior in financial markets. *EPL (Europhysics Letters)*, 96(4):48004.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., and Stanley, H. E. (2002). Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):066126.

Pozzi, F., Di Matteo, T., and Aste, T. (2013). Spread of risk across financial markets: better to invest in the peripheries. *Scientific reports*, 3:1665.

Pradeepkumar, D. and Ravi, V. (2017). Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. *Applied Soft Computing*, 58:35–52.

Pyo, S. and Lee, J. (2018). Exploiting the low-risk anomaly using machine learning to enhance the black–litterman framework: Evidence from south korea. *Pacific-Basin Finance Journal*, 51:1–12.

Pyo, S., Lee, J., Cha, M., and Jang, H. (2017). Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the korean stock markets. *PloS one*, 12(11):e0188107.

Qiu, M. and Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PloS one*, 11(5):e0155133.

Rather, A. M., Agarwal, A., and Sastry, V. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications*, 42(6):3234–3241.

Razak, F. A. and Jensen, H. J. (2014). Quantifying 'causality'in complex systems: understanding transfer entropy. *PLoS One*, 9(6):e99462.

Reid, D., Hussain, A. J., and Tawfik, H. (2014). Financial time series prediction using spiking neural networks. *PloS one*, 9(8):e103656.

Rosenblatt, F. (1961). Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Technical report, Cornell Aeronautical Lab Inc Buffalo NY.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533.

Sabatini, A. (2000). Analysis of postural sway using entropy measures of signal complexity. *Medical and Biological Engineering and Computing*, 38(6):617–624.

Sandoval, L. (2014). Structure of a global network of financial companies based on transfer entropy. *Entropy*, 16(8):4443–4482.

Sandoval Jr, L. (2014). To lag or not to lag? how to compare indices of stock markets that operate on different times. *Physica A: Statistical Mechanics and its Applications*, 403:227–243.

Satchell, S. and Scowcroft, A. (2000). A demystification of the black–litterman model: Managing quantitative and traditional portfolio construction. *Journal of Asset Management*, 1(2):138–150.

Schiavo, S., Reyes, J., and Fagiolo, G. (2010). International trade and financial integration: a weighted network analysis. *Quantitative Finance*, 10(4):389–399.

Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, 85(2):461.

Sensoy, A., Sobaci, C., Sensoy, S., and Alali, F. (2014). Effective transfer entropy approach to information flow between exchange rates and stock markets. *Chaos, Solitons & Fractals*, 68:180–185.

Sezer, O. B. and Ozbayoglu, A. M. (2018). Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Applied Soft Computing*.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55.

Sharpe, W. F. (1994). The sharpe ratio. *Journal of portfolio management*, 21(1):49–58.

Shin, H., Hou, T., Park, K., Park, C.-K., and Choi, S. (2013). Prediction of movement direction in crude oil prices based on semi-supervised learning. *Decision Support Systems*, 55(1):348–358.

Shynkevich, Y., McGinnity, T. M., Coleman, S. A., and Belatreche, A. (2016). Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning. *Decision Support Systems*, 85:74–83.

Silva, T., Pinheiro, P. R., and Poggi, M. (2017). A more human-like portfolio optimization approach. *European Journal of Operational Research*, 256(1):252–260.

Song, D.-M., Tumminello, M., Zhou, W.-X., and Mantegna, R. N. (2011). Evolution of worldwide stock markets, correlation structure, and correlation-based graphs. *Physical Review E*, 84(2):026108.

Soytas, U. and Sari, R. (2003). Energy consumption and gdp: causality relationship in g-7 countries and emerging markets. *Energy economics*, 25(1):33–37.

Swanson, N. R., Ghysels, E., and Callan, M. (1999). A multivariate time series analysis of the data revision process for industrial production and the composite leading indicator. *Cointegration, Causality and Forecasting: Festschrift in Honor of Clive WJ Granger*.

Thakur, M. and Kumar, D. (2018). A hybrid financial trading support system using multi-category classifiers and random forest. *Applied Soft Computing*, 67:337–349.

Tsai, C.-F. and Hsiao, Y.-C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems*, 50(1):258–269.

Tsaih, R., Hsu, Y., and Lai, C. C. (1998). Forecasting s&p 500 stock index futures with a hybrid ai system. *Decision Support Systems*, 23(2):161–174.

Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience*, 30(1):45–67.

Vỳrost, T., Lyócsa, Š., and Baumöhl, E. (2015). Granger causality stock market networks: Temporal proximity and preferential attachment. *Physica A: Statistical Mechanics and its Applications*, 427:262–276.

Wang, Q., Xu, W., and Zheng, H. (2018). Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles. *Neurocomputing*, 299:51–61.

Xiao, Y. and Valdez, E. A. (2015). A black–litterman asset allocation model under elliptical distributions. *Quantitative Finance*, 15(3):509–519.

Yook, S.-H., Chae, H., Kim, J., and Kim, Y. (2016). Finding modules and hierarchy in weighted financial network using transfer entropy. *Physica A: Statistical Mechanics and its Applications*, 447:493–501.

Yue, P., Cai, Q., Yan, W., and Zhou, W.-X. (2020). Information flow networks of chinese stock market sectors. *IEEE Access*, 8:13066–13077.

Zahedi, J. and Rounaghi, M. M. (2015). Application of artificial neural network models and principal component analysis method in predicting stock prices on tehran stock exchange. *Physica A: Statistical Mechanics and its Applications*, 438:178–187.

Zhang, J., Cui, S., Xu, Y., Li, Q., and Li, T. (2018). A novel data-driven stock price trend prediction system. *Expert Systems with Applications*, 97:60–69.

Zunino, L., Tabak, B. M., Figliola, A., Pérez, D., Garavaglia, M., and Rosso, O. (2008). A multifractal approach for stock market inefficiency. *Physica A: Statistical Mechanics and its Applications*, 387(26):6558–6566.

국문초록

주식 시장은 경제 분야의 중요한 부분으로 광범위하게 연구되고 있다. 특히, 주식 시장의 구성 요소들인 주식 가격과 그 수익률의 관계를 예측하고 분석하는 연구는 투자자들이 최적 투자 전략을 세우기 위해 중요한 과업 중 하나이다. 이러한 맥락에서, 어떠한 시스템의 구성 요소들 간의 관계를 분석하는 데 있어 전이 엔트로피(Transfer entropy)는 비모수 지표로써 상관 관계나 그레인저-인과관계에 비해 요소 간 통계적 설명력을 확인하기에 용이하다. 주식 가격의 예측과 이를 통한 최적 자산 배분 전략에 대한 연구 또한 전통적인 선형 모델부터 최신의 머신 러닝 모델의 적용까지 다양하게 연구되고 있다.

본 학위논문의 목적은 경제물리학과 정보이론 분야에서 사용되는 효율적 전이 엔트로피(Effective transfer entropy, ETE)를 이용하여 미국 주식 시장에서 시장 구성 요소 간 발생하는 정보 흐름의 특징을 파악하여 시장의 특성을 나타낼 수 있는 시장 설명력 있는 ETE 기반의 네트워크 지표를 도출하고, 이 네트워크 지표의 사용이 다양한 머신 러닝 알고리즘을 통한 주가 방향 예측에서 성능 향상을 가져다 주는 지에 대해 연구한다. 나아가, 시장 설명력 있는 ETE 네트워크 지표의 구조적 특징과 머신 러닝 알고리즘을 통한 주가 방향 예측 결과를 투자자 관점을 고려한 최적 포트폴리오 구성 전략인 블랙-리터만 모형(Black-Litterman model)에 적용하여 결과적으로 정보 이론과 머신 러닝 기법을 이용한 실제 투자 전략 활용성에 대해 연구한다.

먼저, 미국 주식 시장의 주요 금융 위기들과 주식들 간의 통계적 설명력을 ETE를 통해 분석함으로써 3개월과 6개월 이동창을 기반으로 하는 ETE가 미국 주식 시장에 대해 설명력 있는 지표임을 확인했다. 해당 지표가 주요 금융 위기에서 그 값이 커지고, ETE 네트워크 지표의 시계열 분석을 통해 각 금융위기에서 해당 금융 위기와 관련된

섹터들이 다른 섹터들에 통계적 설명력이 있는 것을 확인했다.

다음으로, 로지스틱 회귀(Logistic regression, LR), 다층 퍼셉트론(Multilayer perceptron, MLP), 랜덤 포레스트(Random forest, RF), XGBoost(XGB) 및 Long short-term memory network(LSTM)의 5개 머신 러닝 알고리즘에 대해 ETE 네트워크 지표가 새로운 변수로 추가되었을 때 주가 방향 예측에 대한 예측 성능이 향상되는 것을 확인했다. 한편, 예측 모델의 예측 성능 평가에 대한 지표로 금융 분야에서 쓰이는 위험 조정 수익률로부터 도출한 수정 정확도 활용을 제안했고, 이 평가 지표를 이용한 분석을 통해 해당 5개 모델 중 MLP와 LSTM이 미국 주식 시장에 대한 주가 방향 예측에서 더 적합한 모델임을 확인했다.

마지막으로, 시장 설명력 있는 유입 및 유출 ETE 네트워크 지표의 특징과 머신 러닝 알고리즘을 이용한 주가 방향 예측 결과를 블랙-리터만 모형의 투자자 관점에 적용하여, 머신 러닝 알고리즘을 이용한 주가 방향 예측 결과를 투자자 관점에 적용한 블랙-리터만 포트폴리오는 시장 포트폴리오와 시장 인덱스보다 나은 위험 대비 수익률을 보이고, ETE 네트워크 지표를 적용한 블랙-리터만 포트폴리오는 가장 높은 수익률을 보임을 확인했다. ETE와 주가 방향 예측의 사용이 투자 수익률 향상으로 이어지고, 예측 성능을 향상시키면 투자 수익률도 함께 증가하는 결과를 활용하여 투자자들이 ETE와 머신 러닝을 활용한 블랙-리터만 모형을 통해 수익을 극대화 할 수 있는 투자 전략을 수립할 수 있는 가능성에 대해 확인했다.

본 학위논문은 정보 이론의 ETE를 금융 투자 분야에 적용할 수 있도록, 머신 러닝 알고리즘을 이용한 주가 방향 예측과 블랙-리터만 모형을 통한 최적 포트폴리오 구성 전략에 대한 첫 번째 연구이다.

주요어: 정보 이론, 경제물리학, 전이 엔트로피, 머신 러닝, 특성 추출, 예측 알고리즘, 주식 시장, 시계열 분석, 블랙-리터만 모형, 최적 자산 배분

학번: 2016-30252