



Essays on Hybrid Modeling of Machine Learning Algorithms
and Financial Time Series Models

by

Yi Luo

A thesis submitted to the Economics Department
at Lancaster University for the degree of
Doctor of Philosophy in Economics

June, 2022

When youth departs, may wisdom prove enough.

I dedicate this thesis to my grandfathers Mr. Shijie Luo and Mr. Yongfa Chen who have strong faith in my talent and capabilities. I also dedicate this thesis to those who love me and help me throughout the path.

谨以此文纪念我的祖父罗世杰先生，我的外祖父陈永发先生，感谢你们无条件信任我的天赋与能力。我也将此文献给所有爱我和帮助过我的大家。

Declaration

I declare that this thesis is my own work and has not been submitted for the award of a higher degree elsewhere. This thesis contains no material previously published or written by any other person except where references have been made in the thesis.

Yi Luo

June, 2022

Essays on Hybrid Modeling of Machine Learning Algorithms and Financial Time Series Models

Yi Luo

Economics Department, Lancaster University

A thesis submitted for the degree of *Doctor of Philosophy in Economics*. June, 2022

Abstract

This thesis attempts to model and forecast realized volatility and stock market tail risk using hybrid models integrating Machine Learning algorithms with Financial Time Series models. One of the advantages of Machine Learning approaches is that it can well approximate a wide range class of linear and nonlinear functions, forming the input-output map by learning the data rather than assuming the data generating process. Traditional Time Series models, however, focus on reproducing the stylized facts of target variables through statistical modeling. By hybridizing these two types of models, we find that Machine Learning approaches well complement Financial Time Series models in variable screening, complex relationship detection and nonlinearity modeling. In addition, it is found that instead of using raw data in the Machine Learning algorithms, Financial Time Series models generate more effective features that significantly improves learning ability of those algorithms.

Acknowledgements

If someone told me that I can complete Doctorate Degree several years ago, I would never have imagined that I could actually make it. This achievement would never be possible without warm help from many people. First and for most, I would like to express my genuine gratitude to my supervisors, Professor Marwan Izzeldin and Professor Mike Tsionas, for all inspiring discussions, experienced guidance and continuous caring. In particular, I want to thank Professor Marwan for: 1) providing me with this fantastic opportunity to realize my dreams; 2) strong belief in my potential and capabilities; 3) kind encouragement and support when I face difficulties; 4) enlightening discussions about research, teaching, life and more. You are the best supervisor I have ever met and I would always remember that you said I am your best student!

I am very grateful to our Head of Department, Professor Efthymios Pavlidis, for the initial PhD application talk in 2017. You told me that PhD is a lonely journey and it is a huge decision so that I must think carefully before choosing to apply. Many people may think owing a Doctorate Degree is more or less a great achievement but very few of them actually know how tremendous effort that have to be made to achieve the final goal. I was lucky to have this serious talk at the beginning of my PhD journey. Let me pass special tribute to Gerry Steele who offers professional and extremely helpful editorial suggestions on my papers.

I would like to thank Lancaster University Management School for offering me honorable studentship without which my achievement would never be possible. I am also thankful to academic and administrative staff of the Economics Department at Lancaster University for great discussions and support throughout these years. I wish my PhD colleagues all the best in their life and career and I hope we can keep in touch. I especially thank to Aya Ghalayini for sweet discussions and delicious Lebanese food.

I want to thank myself for never giving up even when things were very difficult. Three years and eight months, I am finally here. I have become stronger, smarter and more confident. I was a Top Accounting student before pursuing PhD and I chose to start from zero in a cutting-edge research direction. Not everyone can accept falling from the top to the bottom and head for the top again. I did it. I want to tell everyone that fate can be changed as long as you are determined enough. Until today, I have

completed three PhD chapters in my interested area, achieved MIT MicroMaster Degree in Statistics and Data Science where I spent thousands of hours, taught diversified students in various modules, presented in several prestigious conferences, polished my coding skills and found a decent job. This is not because I am smart but because I am insistent. There were countless breaking moment in the past few years that I could not even think about it. However, it is those hard times that shapes a unique me and I am truly grateful to everything on the path.

I am grateful for having so many amazing people in my life. My mom is the greatest woman in this world. She is such a strong woman as if nothing can defeat her but I know she also needs to be protected. I want to tell her that I love her, respect her and would protect her. My dad never speaks for himself but I know that he loves me although we haven't spent much time together in our lives. I never told my dad that I also love him more than anything. Let me express special gratitude to my husband for his unconditional love and care. I would say it was magic that links us. After all these years, I can still feel endless love between you and me. You always asking me what am I thinking when I look at you. I am thinking about you and every moment I spent with you. You are so amazing that all beautiful words fit you well.

Last but not least, I want to thank my friends Emily Wang, Jun Dai, Shanshan Zhou, Dr. Xiaohan Xue, Dr. Huan Yang, Dr. Yuting Bai, Dr. Xingzhi Yao for all the sweet times and insightful discussions.

Contents

Introduction	4
1 The Impact of Periodicity on Volatility-Volume Relations: Evidence from Mixture Model	5
1.1 Introduction	5
1.2 Theoretical background	8
1.3 Methodology	11
1.3.1 Periodicity estimation in returns	11
1.3.1.1 Constant periodicity	13
1.3.1.2 Day-by-Day periodicity	13
1.3.2 Periodicity estimation in volume and number of trades	14
1.3.2.1 The inner loop of STL	15
1.3.2.2 The outer loop of STL	16
1.4 Empirical framework	17
1.4.1 Mixture model	18
1.4.2 Linear regression model	19
1.4.3 Random forest	20
1.5 Results	21
1.6 Conclusion	23
Appendix A Chapter 1	25
A.1 Tables and Figures	25
2 Forecasting RV: A Hybrid Model Integrating BiLSTM with HAR-type Models	38

2.1	Introduction	38
2.2	Models	43
2.2.1	HAR model	43
2.2.2	HAR-TCJ model	44
2.2.3	Neural Network models	45
2.2.3.1	Deep Feedforward Neural Network model	46
2.2.3.2	Bidirectional Long Short-term Memory Network model	47
2.2.4	Proposed hybrid models - BiLSTM with HAR-type models	49
2.3	Data and experiment	51
2.4	Results	53
2.5	Conclusion	55
Appendix B Chapter 2		57
B.1	Tables and Figures	57
3	When MIDAS Meets LASSO: Forecasting Tail Risk Using Effective Macroeconomic Variables	68
3.1	Introduction	68
3.2	Models	72
3.2.1	The ES-CAViaR-MIDAS model	72
3.2.2	ES-CAViaR-MIDAS with variable selection	76
3.2.3	Tuning parameter selection	76
3.2.4	Estimation framework	77
3.3	Data and empirical study	78
3.3.1	Macroeconomic variables	79
3.3.2	Financial variables	80
3.3.3	Principal Component Analysis (PCA) for variables	81
3.4	Empirical analysis	82
3.4.1	Dynamic variable selection using Adaptive-Lasso	82
3.4.2	Backtesting approaches	84
3.4.2.1	Unconditional coverage (UC) test	85
3.4.2.2	Dynamic quantile (DQ) regression	85
3.4.2.3	Quantile score function	86

3.4.2.4	Bootstrap test for ES	86
3.4.2.5	Dynamic Expected Shortfall (DES) regression	87
3.4.2.6	FZ loss function for (VaR, ES)	87
3.4.3	Backtesting results	88
3.4.4	Robustness checks	91
3.4.4.1	Restricted Beta Weighting schemes	91
3.5	Conclusion	92
Appendix C Chapter 3		94
C.1	Tables and Figures	94
4	Concluding Remarks and Further Developments	112
	References	116

List of Figures

A.1	Intraday return	27
A.2	Intraday raw trading volume	28
A.3	Intraday raw number of trades	29
A.4	ACF plot for raw and filtered trading volume	30
A.5	ACF plot for raw and filtered number of trades	31
A.6	Mixture model correlation plot for trading volume	32
A.7	Mixture model correlation plot for number of trades	33
A.8	Mixture model posterior probability plot for trading volume	34
A.9	Mixture model posterior probability plot for number of trades	35
A.10	In-sample Adjusted R squared results under linear regression framework	36
A.11	In-sample Adjusted R squared results under Random Forest framework	37
B.1	LSTM memory cell	64
B.2	Structure of Neural Network models	65
B.3	ACF and PACF plot of SPY's RV	66
B.4	Structure of the proposed HAR-BiLSTM hybrid model	67
C.1	Estimation framework	103
C.2	Dynamic variable selection at $\alpha = 1\%$	104
C.3	Dynamic variable selection at $\alpha = 2.5\%$	105
C.4	Dynamic variable selection at $\alpha = 5\%$	106
C.5	Dynamic variable selection at $\alpha = 10\%$	107
C.6	Dynamic variable selection at $\alpha = 1\%$ with fixed w_1	108
C.7	Dynamic variable selection at $\alpha = 2.5\%$ with fixed w_1	109
C.8	Dynamic variable selection at $\alpha = 5\%$ with fixed w_1	110

C.9 Dynamic variable selection at $\alpha = 10\%$ with fixed w_1 111

List of Tables

A.1	Summary statistics	25
A.2	Out-of-sample performance of linear regression framework	26
A.3	Out-of-sample performance of Random Forest framework	26
B.1	Summary statistics	57
B.2	Input variables of different models	58
B.3	Selected tuning parameter values for $h = 1$	59
B.4	Selected tuning parameter values for $h = 5$	60
B.5	Out-of-sample forecast losses: pre-crisis sample	61
B.6	Out-of-sample forecast losses: post-crisis sample	62
B.7	Out-of-sample forecast losses: last sample	63
C.1	Summary statistics	95
C.2	Variable correlation matrix	96
C.3	Principal Component Analysis for 20 macroeconomic and financial variables	97
C.4	Description of competing models	97
C.5	Backtesting results for estimated risk measures (VaR, ES) of the S&P 500 index	98
C.6	Out-of-sample performance rankings for various levels of α	100
C.7	Backtesting results for estimated risk measures (VaR, ES) of the S&P 500 index with fixed w_1	101

Introduction

Machine Learning (ML) and Deep Learning (DL) techniques are becoming more and more popular nowadays both in academia and in financial industry. Numerous studies have been published on ML models with superior performances than classical time series forecasting techniques. Meanwhile, the widespread application of automated electronic trading systems coupled with increasing demand for higher yields keeps forcing researchers and practitioners to continue working on implementing better models. In the last few years, DL has strongly emerged as the best performing predictor class within the ML field in various implementation areas. However, both ML and DL models have been criticized for the "black box" problem which makes the results interpretation very difficult since it does not provide any insights about the resulted input-output map (Lai et al., 2009). Moreover, these approaches suffer from the overfitting problem and the fantastic training sample performance can be hardly generalized to the testing sample.

Financial time series models, on the other hand, focus on modeling the data generating process and stylized facts of target variables. The estimated parameters of Financial Time Series models carry clear economic explanation and are often easy to be interpreted. However, one of the problems is that these models usually have good in-sample fitting but the out-of-sample forecasting performance can be deteriorated when underlying assumptions are violated. In addition, possible nonlinear relationship between Financial Time Series and other variables (i.e., macroeconomic variables) is hard to be modeled through traditional statistical modeling which often requires complicated data processing before estimation.

In view of these, this thesis attempts to find a path which enables two models complementing each other and achieving better forecasting performance. The empirical

application focuses on predicting realized volatility and stock market tail risk. In particular, Chapter 1 uses statistical modeling to extract periodicity, which is a noisy factor, in realized volatility and trading volume. By filtering periodicity from trading volume, the predictability of market activity improves significantly in forecasting future realized volatility under both linear regression and Machine Learning framework. In Chapter 2, to better forecast realized volatility, we propose a nonparametric neural network-based hybrid model which takes components generated from Financial Time Series models as input variables. The proposed hybrid model significantly outperforms all benchmark models across all forecasting horizons. Chapter 3 targets at predicting stock market tail risk using effective low-frequency variables. We propose a new quantile-based MIDAS model to integrate low-frequency information and the most important predictors are selected using Machine Learning approach. Our model achieves the minimum loss in forecasting tail risk. We now provide details of each chapter as follows.

In **Chapter 1**, we contribute to the volatility-volume relation literature by examining the volatility-volume relation from a new perspective. First, we propose to use the Seasonal-Trend Decomposition Procedure Based on Locally Estimated Smoothing (STL) to estimate the periodic component in the trading activity. It is fast and easy to apply. Second, by applying the mixture model and posterior probability analysis, we jointly model the periodicity in financial return and trading volume, allowing varying mean and variance of normal distribution. We provide evidence for Mixture of Distribution Hypothesis at the periodicity level. We find that periodicity in trading volume is highly correlated with periodic return in the early trading day while such correlation becomes weaker during the rest of the day. This indicates that the volatility-volume relation may be time-varying and the failure to account this in statistical modeling may explain the contradictory results in the empirical literature. Third, by filtering periodicity from trading volume, the predictability of trading volume on return volatility is significantly improved under both linear regression and Random Forest improves significantly. This suggests that the periodicity in trading volume may be a noisy factor that needs to be carefully handled in investigating the volatility-volume relation.

In **Chapter 2**, we propose a hybrid model integrating Neural Networks (NN) with

Heterogeneous Autoregressive-type (HAR) models to forecast realized volatility. In addition to the well documented long memory in realized volatility, it has found that realized volatility subjects to regime-switching and structural breaks which requires modeling of both self-similarity and nonlinearity. In viewing of this, some authors suggest to use smooth transition or threshold model. However, these methods require huge amount of data to determine the number of regimes before estimating the model and the sample log-likelihood can easily trap in local maxima, leading to poor out-of-sample performance (Pavlidis et al., 2012). With the development of computational resources, NN has become increasingly popular in forecasting financial time series given its superior learning ability. However, it is criticized by lacking explicit economic interpretation. Another fold of literature therefore focuses on building the semiparametric HAR-based hybrid models in which NN-based nonlinear transformation term is added to the original HAR. However, by using the simple additive combination, this approach may underestimate the relationship between the HAR components and nonlinear term (Taskaya-Temizel and Casey, 2005). To overcome the limitations, we propose the nonparametric NN-based hybrid models which takes various input features generated from HAR-type models. The proposed model approximates possible linearities and nonlinearities through effective learning while taking HAR-type components to improve interpretability. The empirical results show that NN-based hybrid models outperform all other benchmarks in the out-of-sample forecasting. This indicates that HAR-type components can be served as effective features in NN structure.

In **Chapter 3**, a new framework for the joint estimation and forecasting of Value at Risk (VaR) and Expected Shortfall (ES) is proposed, which incorporates low-frequency macroeconomic and financial indicators into the quantile-based MIDAS model. Low-frequency macroeconomic and financial information have been found important in predicting stock market volatility while is much less exploited in forecasting tail risk. The reasons might be attributed to the lack of an appropriate framework that deals with frequency disalignment and an effective variable screening device that selects the strongest predictors among large numbers of candidate variables. To address these problems, we propose a new quantile-based MIDAS model in which the low-frequency information is integrated. By using an innovative Machine-Learning approach

that maximizes the penalized Asymmetric Laplace (AL) likelihood function with an Adaptive-Lasso penalty, the most informative variables are selected in a “big data” setting. The dynamic selection process enables the visualizing of the variable-selection evolution. In the empirical analysis, three variables (namely, realized volatility, term spread and housing starts) are consistently selected for most of the rolling windows and serve to the strongest predictors of future tail risk. Moreover, the average number of selected variables increases in predicting more extreme tail risk which indicates that more extreme VaR and ES may rely on additional information. The out-of-sample backtesting results show that our method passes most backtests with relatively higher p -values and achieves the minimum loss in the joint forecasting of VaR and ES.

Chapter 1

The Impact of Periodicity on Volatility-Volume Relations: Evidence from Mixture Model

1.1 Introduction

The key of the modern microstructure theory is that trades possess information and lead to a persistent impact on security price. There are a lot of theoretical and empirical literature investigating the relationship of volatility and trading volume. Researchers find it important as the volatility-volume relation possesses information about how the market participants process the new information and how the market price reacts to the new information.

In early empirical literature, volatility is mainly measured by absolute returns and trading volume is mainly measured by number of traded shares under consistent time intervals. The data used is mainly weekly or monthly returns ([Giot, Laurent, and Petitjean \(2010\)](#)). However, with the development of high-frequency data, realized volatility becomes a common proxy of volatility measurement, while trading activity is classified as trade sizes (i.e., the number of shares per trade) and trade frequency (i.e., number of trades). The theoretical models that explain the volatility-volume relation are classified into three classes: the competitive microstructure model ([Glosten and Milgrom \(1985\)](#)), the strategic microstructure model ([Kyle \(1985\)](#)) and the mixture

type of distribution model ([Clark \(1973\)](#)).

Both the competitive and the strategic microstructure models are the extensions of the Sequential Arrival of Information Hypothesis ([Tauchen and Pitts \(1983\)](#)). Competitive model indicates that the market makers and individual investors experience adverse selection problem when trading with informed investors. Due to the different perception of the importance of the information and the different actions accordingly, investors who possess better information tend to execute larger-sized trades. Therefore, competitive models support a positive relationship between the volatility and trade size. However, the strategic model assumes different behavior of the informed investors. It assumes that the informed investors tend to stealth themselves by breaking up the large trades into many smaller transactions ([Giot, Laurent, and Petitjean \(2010\)](#)). Therefore, the strategic model supports a weaker positive relation between the volatility and trading size as part of its impact may be transferred into the number of trades.

The Mixture of Distribution Hypothesis (MDH) offers an intuitively appealing explanation for the strongly positive correlation between return volatility and trading volume. The MDH is primarily statistical models. Central to the hypothesis is that the price volatility and trading volume are driven by the same information arrival rate or news process. Increase in price would be resulted by the unexpected good news arrival, while the decrease in price is driven by the bad news. As the market adjusts itself to a new equilibrium, an above-average trading activity would appear when those events occur ([Luu and Martens \(2003\)](#)). [Tauchen and Pitts \(1983\)](#) raise the fundamental mixture model where the return and trading volume were jointly normally distributed and subjected to the same information arrival rate. [Andersen \(1996\)](#) extends this research by distinguishing the informed and uninformed trading volume among which only the informed trading volume and volatility are jointly determined by the information arrival rate. Moreover, many recent empirical literature provide strong evidence for the MDH and indicate that it is the number of trades rather than the average trade size reflects the amount of information arrivals. For example, [Jones et al. \(1994\)](#) decompose the realized variance into continuous component and jumps component and find that the price volatility is driven by the number of trades under equal time interval. [K. Chan and W.-M. Fong \(2000\)](#) find that it is the order imbalance that influences the volatility-volume relation. [Giot, Laurent, and Petitjean \(2010\)](#) find

that the positive volatility-volume relation is only applied for the continuous component and both the trade size and order imbalance add no significant explanatory power beyond the number of trades. Furthermore, [R. D. Huang and Masulis \(2003\)](#) take trade size into consideration (i.e., large trades and small trades) and find that trade frequency influences the volatility for large trades while both trade frequency and trade size affect the volatility for small trades.

Despite the intuitive explanation offered by the MDH, it subjects to certain limitations. First, it is well documented in the literature that financial return is fat-tailed rather than normally distributed. Violation of such assumption may explain the low predictability of trading volume on return volatility. Second, in addition to information flow variable, both return and trading volume may be jointly affected by many other factors (i.e., investor expectation and trading preference). In view of this, the conditional normality can be easily violated if those noisy factors are not controlled.

We contribute to the volatility-volume literature by considering the MDH in a more micro way. Instead of jointly modeling the return and trading volume, we model periodicity in both return and trading volume through the mixture model. The empirical research which provides the evidence for the existence of the intraday returns patterns can be traced to [McInish and Wood \(1985\)](#) who discover that the return volatility displays a U-shaped pattern over the trading day. That is, volatility is low in the middle of the day and high at the opening and closing time during the day. [Müller et al. \(1990\)](#) demonstrate similar patterns in the foreign exchange market. However, no rigorous statistical modelling of periodicity factor in financial return process until the work of [Andersen and Bollerslev \(1997\)](#). They propose a statistical model allowing periodicity in the high frequency return calculation process and the realized variance shows clear long memory feature after filtering periodicity. They find that ignoring the periodicity may distort the relationship between return volatility and microstructural variables. In other words, return periodicity can be viewed as a noisy factor that negatively affect volatility estimation and other studies of microstructural relationship. Their study is extended by [Boudt et al. \(2011\)](#) who finds that periodicity in return volatility biases existing jump tests and therefore proposes periodicity robust jump test. Many studies have provided evidence that the volatility changes systematically over the trading day and such pattern is related to the intraday trading volume variation which

indicates that the volatility periodicity should be affected by periodicity in trading volume. However, to the best of our knowledge, the periodicity in trading volume is much less exploited and there is lack of statistical investigation on the relationship between periodicity in financial return and trading volume.

Our contributions are as follows. First, we propose to use the Seasonal-Trend Decomposition Procedure Based on Locally Estimated Smoothing (STL) to estimate the periodic component in the trading activity. It is fast and easy to apply. Second, by applying the mixture model and posterior probability analysis, we jointly model the periodicity in financial return and trading volume, allowing varying mean and variance of normal distribution. We provide evidence for MDH at the periodicity level. We find that periodicity in trading volume is highly correlated with periodic return in the early trading day while such correlation becomes weaker during the rest of the day. This indicates that the volatility-volume relation may be time-varying and the failure to account this in statistical modeling may explain the contradictory results in the empirical literature. Third, by filtering periodicity from trading volume, the predictability of trading volume on return volatility is significantly improved under both linear regression and Random Forest improves significantly. This suggests that the periodicity in trading volume may be a noisy factor that needs to be carefully handled in investigating the volatility-volume relation.

The rest of this chapter is summarized as follows. In section 2, we review the relevant literature. In section 3, we introduce the theoretic models and research methodology. In section 4, we describe the empirical framework. In section 5, we report and analyze the empirical results. We conclude in section 6.

1.2 Theoretical background

The volatility-volume relation have been in the central stage for the empirical research for decades. This is not only because the price changes are clearly caused by the trading activity but also the relation is of great importance for the asset pricing, portfolio allocation and risk management. Also, empirical research is still trying to investigate whether there are other elements that drive the volatility, which is key to the valuation and prediction. [Shahzad et al. \(2014\)](#) suggest that the market would be more efficient

and less volatile if the investors are mainly informed investments.

Empirical research on the volatility-volume relation can be traced to [Osborne \(1959\)](#), who models the price change as diffusion process with variance being dependent on the number of trades. This may indicate a positive relationship between the absolute price change and the trading volume. This work is further developed by [Clark \(1973\)](#) and [L. Harris \(1989\)](#). However, because Osborne assumes the transactions are uniformly distributed in time, the volume-price issue was not directly addressed.

Early empirical research that focuses on volume-price was conducted by [Granger and Morgenstern \(1963\)](#). They analyze weekly data of New York Stock Exchange from 1939 to 1961 but find no relation between the price index and the aggregate level of volume. In 1964, they applied new data which includes daily and transaction data for individual stocks but still found no relation between the absolute price change and the trading volume. Although they do not find robust relation between the price change and trading volume, they do find that there is a positive relationship between the daily volume and the difference of daily high and daily low. However, [T. W. Epps and M. L. Epps \(1976\)](#) indicate that the relationship is caused by the distribution of the price change. The failure of [Granger and Morgenstern \(1963\)](#) motivated [Ying \(1966\)](#) to further investigate the relationship between price change and trading volume. By using Standard and Poor's 500 composite price index and the volume of NYSE traded shares, [Ying \(1966\)](#) find a positive relationship between the trading volume and absolute price change. This finding is later supported by [Crouch \(1970\)](#), [Clark \(1973\)](#), [Morgan \(1976\)](#), [Tauchen and Pitts \(1983\)](#), [Rutledge \(1979\)](#), [L. Harris \(1989\)](#), [Lamoureux and Lastrapes \(1990\)](#), and [Liesenfeld \(2001\)](#).

The above literature all emphasize the role of the trading size (i.e., trading volume), with the positive relationship between volume and price change having been documented for decades. However, many literature start to focus on the explanatory power of number of trades. By applying daily data of 853 stocks from NASDAQ Exchange, [Jones et al. \(1994\)](#) find that it is the number of trades or the occurrence of transactions that drive the volatility rather than the trading size. The trading size almost has no explanatory power. This finding is later supported by [Easley et al. \(1997\)](#) and [Hasbrouck \(1999\)](#).

However, [K. Chan and W.-M. Fong \(2000\)](#) argue that it is too early to conclude

that the trading size has no information content. First, [Barclay and Warner \(1993\)](#) indicate that the informed trades would prefer medium-sized trades if they want to disguise their true intention in the market by breaking up trades, in which case the medium-sized trade class may have the greatest volatility impact and we may not be able to investigate the role of trading size when using the average trading size. Second, [Jones et al. \(1994\)](#) only apply data from NASDAQ. Therefore, whether the conclusion is valid for NYSE is still unknown. It is widely documented that the NYSE has different market microstructure compared with NASDAQ, which could affect the results. By applying data from Trades and Quotes (TAQ) database, [K. Chan and W.-M. Fong \(2000\)](#) classify transactions into three groups, which are small-sized, medium sized and large-sized trades. Using both NYSE and NASDAQ stocks data, they then regress the absolute return and the number of trades in those three groups in order to see if the number of trades has more explanatory power in the specified groups. They confirm a robust and positive relationship for both trading volume and number of trades on the volatility in all three groups and two markets. Still, number of trades has more explanatory power than the trading volume. This finding is further supported by [R. D. Huang and Masulis \(2003\)](#), [Avramov et al. \(2006\)](#), [Izzeldin \(2007\)](#), [Giot, Laurent, and Petitjean \(2010\)](#), and [T. Wang and Z. Huang \(2012\)](#).

Despite investigating the explanatory power of trading volume and number of trades, many literature give attention to other elements that would affect the volatility-volume relation. For example, [Kyle \(1985\)](#) and [Admati and Pfleiderer \(1988\)](#) examine the role of net order flow (the absolute value of the difference between the number of buyer-initiated trades and the seller-initiated trades) or the order imbalance in determining the volatility-volume relation. In such situation, the market makers may not be able to distinguish the informed trader and the uninformed trader due to the information asymmetry. Therefore, they will infer the information from the net flow in order to adjust the asset price. When there are excessive buying orders, they would revise the price upward and vice versa. This was later supported by [Glosten and L. E. Harris \(1988\)](#), [K. Chan and W.-M. Fong \(2000\)](#), and [Shahzad et al. \(2014\)](#).

However, [C. C. Chan and W. M. Fong \(2006\)](#) re-examine the impact of order imbalance and the explanatory power of market activity on the return volatility. Unlike their previous studies, they use the realized variance to measure the volatility rather

than the absolute returns. Again, [C. C. Chan and W. M. Fong \(2006\)](#) confirm that the realized variance is a more accurate estimator of true latent volatility. Moreover, they find that it is the number of trades that plays a dominant role in explaining the volatility. In contrast, neither the trading volume nor the order imbalance adds the explanatory power significantly beyond the number of trades. [Shahzad et al. \(2014\)](#) also confirm that the order imbalance plays little role in explaining the volatility-volume relation.

In conclusion, most of these research belong to the following three categories. First, some authors investigate the information transmission between geographically separated financial markets which are trading sequentially. Second, some literature focus on the lead-lag relation between the financial markets that are trading simultaneously. Third, some authors investigate the role of different microstructure variables in determining the intraday volatility. Our study falls into the third category where we try to model a noisy factor, periodicity, in both realized volatility and trading volume and investigate if it has significant impact on volatility-volume relation.

1.3 Methodology

1.3.1 Periodicity estimation in returns

Let r_i be the consecutive compounded intraday return observations ($i = 1, \dots, DM$), where D is the number of days in our sample period and M is the return interval. After dividing one trading day into unity, $\Delta = 1/M$ is the sampling frequency, meaning the time between two consecutive observations. By assuming r_i is a random variable with normal distribution and zero mean, the high frequency return data generating process is given by,

$$r_i = \sigma_i u_i = p_i f_i u_i \tag{1.1}$$

where σ_i is the standard deviation and can be decomposed into a combination of a deterministic factor p_i and an average volatility factor f_i . Moreover, p_i is the function that contains periodic elements such as the day of the week and news announcements effect. u_i is random innovation and $u_i \stackrel{iid}{\sim} \mathcal{N}(0, 1)$.

Therefore, the impact of deterministic factor p_i is dependent on the choice of time horizon. We obtain the time horizon in $[0, D]$ with length τ and $[\tau/\Delta]$ is the groups of consecutive observations, which is divided by DM observations. According to the standardization condition that the mean of squared periodicity factor equals to one, for all $i = 1, \dots, DM$, we have,

$$\frac{1}{[\tau/\Delta]} \sum_{s \in N_i} p_s^2 = 1 \quad (1.2)$$

where the indices s belongs to the collection N_i and stays in the same time horizon as i .

The returns in Eq.(1.1) is the discrete changes of the consecutive log-price process, which is subjected to Brownian SemiMartingale diffusion. Therefore, the mean return in Eq.(1.1) is zero and the volatility function can be divided into two components, which is slowly time-varying component and a possibly fast periodically varying component. The former component is modeled as f_i while the latter one is the periodicity factor, which is p_i . Following [Boudt et al. \(2011\)](#), when taking jumps into consideration, Eq.(1.1) can be modeled as,

$$r_i = p_i f_i u_i + a_i \quad (1.3)$$

where a_i is a random variable with zero mean if there are no jumps but non-zero mean when jumps occur.

In order to normalize the r_i , we choose to use the normalized version of the Bi-power variation over the time horizon and take the square root, which is,

$$\widehat{f}_i = \sqrt{\frac{\pi}{2} \frac{1}{[\frac{\tau}{\Delta}] - 1} \sum_{l=s+2}^{s+[\tau/\Delta]} |r_l| |r_{l-1}|} \quad (1.4)$$

Therefore, the periodic components of the intraday volatility can be estimated through the standardized high frequency return, which is,

$$\bar{r}_i = \frac{r_i}{\widehat{f}_i} \quad (1.5)$$

1.3.1.1 Constant periodicity

When the standardized returns are subjected to the same periodicity factor, this means $\bar{r}_{1,i}, \dots, \bar{r}_{n_i,i}$ shares the same periodicity factor as \bar{r}_i . We use the scaled version of the standardized return to estimate the periodicity component and the most suitable scale factor recommended by [Boudt et al. \(2011\)](#) is the Shortest Half scale estimator, which can reduce the jump bias as much as possible. Firstly, we order the standardized returns to ensure that $\bar{r}_{1,i} \leq \bar{r}_{2,i} \leq \dots \leq \bar{r}_{n_i,i}$, then the Shortest Half scale is given by,

$$ShortH_i = 0.741 \cdot \min \left\{ \bar{r}_{(h_i),i} - \bar{r}_{(1),i}, \dots, \bar{r}_{(n_i),i} - \bar{r}_{(n_i-h_i+1),i} \right\} \quad (1.6)$$

where $h_i = 1 + \frac{n_i}{2}$.

Thus, we have the scaled ShortH periodicity estimator,

$$\hat{p}_i^{ShortH} = \frac{ShortH_i}{\sqrt{\frac{1}{[\tau/\Delta]} \sum_{s \in N_i} ShortH_S^2}} \quad (1.7)$$

[Rousseeuw and Leroy \(1988\)](#) find that although the ShortH periodicity estimator is robust to jumps, its sensitivity to jumps is still lower than the Weighted Standard Deviation periodicity estimator. In this chapter, we use the Weighted Standard Deviation periodicity estimator to gain more robustness in jumps. It is given by,

$$\hat{p}_i^{WSD} = \frac{WSD_i}{\sqrt{\frac{1}{[\tau/\Delta]} \sum_{s \in N_i} WSD_S^2}} \quad (1.8)$$

where $WSD_S = \sqrt{1.081 \cdot \frac{\sum_{l=1}^{n_S} w_{l,s} \bar{r}_{l,s}^2}{\sum_{l=1}^{n_S} w_{l,s}}}$ and the weight is based on $w \left(\bar{r}_{1,s} / \hat{p}_i^{ShortH} \right)$

1.3.1.2 Day-by-Day periodicity

The periodicity we estimate above is the constant periodicity, which means the periodicity does not vary in the entire time horizon. Given this assumption is somehow unrealistic, we also choose to estimate the day-by-day periodicity, which means the periodicity changes from day to day. This estimation is more realistic because many literature have shown that the volatility is high when the market opens on Monday morning. The format of day-by-day periodicity is quite similar as the constant

periodicity except having subscript that represents the day of the week. Therefore, the day-by-day Shortest Half scale is given by,

$$\text{ShortH}_i^{(d)} = 0.741 \cdot \min \left\{ \bar{r}_{(h_i),i} - \bar{r}_{(1),i}, \dots, \bar{r}_{(n_i),i} - \bar{r}_{(n_i-h_i+1),i} \right\}, d = 1, \dots, 5 \quad (1.9)$$

Then, we have the day-by-day Shortest Half periodicity estimator,

$$\hat{p}_i^{\text{ShortH}^{(d)}} = \frac{\text{ShortH}_i^{(d)}}{\sqrt{\frac{1}{[\tau/\Delta]} \sum_{s \in N_i} \left(\text{ShortH}_S^{(d)} \right)^2}} \quad (1.10)$$

Similarly, we use the day-by-day Weighted Standard Deviation periodicity estimator, which is given by,

$$\hat{p}_i^{\text{WSD}^{(d)}} = \frac{\text{WSD}_i^{(d)}}{\sqrt{\frac{1}{[\tau/\Delta]} \sum_{s \in N_i} \left(\text{WSD}_S^{(d)} \right)^2}} \quad (1.11)$$

where $\text{WSD}_S^{(d)} = \sqrt{1.081 \cdot \frac{\sum_{l=1}^{n_S} w_{l,s} \bar{r}_{l,s}^2}{\sum_{l=1}^{n_S} w_{l,s}}}$ and the weight is based on $w \left(\bar{r}_{l,s} / \hat{p}_i^{\text{ShortH}^{(d)}} \right)$

1.3.2 Periodicity estimation in volume and number of trades

With respect to the intraday volume and number of trades, we also witness strong periodicity in those series. To align with the periodicity frequency in volatility, we use STL method to estimate constant and day-by-day periodicity in volume. STL was proposed by [Cleveland et al. \(1990\)](#) which states that a seasonal time series can be decomposed into three components: trend component, seasonal component and remainder component. The volume and number of trades series can be decomposed as follows,

$$V_t = T_t + S_t + R_t \quad (1.12)$$

$$N_t = T_t + S_t + R_t$$

where V_t, N_t, T_t, S_t, R_t denote the volume series, number of trades series, trend component, seasonal component and remainder component respectively, for $t = 1, \dots, D$.

However, we find that the seasonal component in the summation version above won't affect the final mean estimate of volume and number of trades. Therefore, we take log transformation on those two series and now the decomposition becomes,

$$\ln(V_t) = \ln(T_t) + \ln(S_t) + \ln(R_t) \quad (1.13)$$

$$\ln(N_t) = \ln(T_t) + \ln(S_t) + \ln(R_t)$$

We now give the six important parameters of STL as follows,

$n_{(p)}$ = number of observations in each cycle of seasonal component

$n_{(i)}$ = total number of iterations of the inner loop

$n_{(o)}$ = total number of iterations of the outer loop

$n_{(l)}$ = smoothing parameter for the Low-Pass filtering procedure

$n_{(t)}$ = parameter for trend smoothing procedure

$n_{(s)}$ = parameter for seasonal smoothing procedure

1.3.2.1 The inner loop of STL

The seasonal component and the trend component of volume series are updated during each iteration of the inner loop by seasonal smoothing process and trend smoothing process.

Let $\ln(S_t^{(k)})$ and $\ln(T_t^{(k)})$ for $t = 1, \dots, D$ be the seasonal and trend components after the k th iteration finishes; they are defined at all times for $t = 1, \dots, D$ even if $\ln(V_t)$ and $\ln(N_t)$ is missing. Therefore, the updated seasonal component and trend component, $\ln(S_t^{(k+1)})$ and $\ln(T_t^{(k+1)})$, are computed by the following steps.

Step 1: Computing the detrended series $\ln(V_t) - \ln(T_t^{(k)})$

Step 2: Subseries Smoothing Process. For the subseries of values at each position of the seasonal cycle (it is obviously one day in our case), we use loess regression to smooth the above detrended series. To apply loess, the polynomial degree parameter d and a positive integer q must be chosen. In our case, $q = n_{(s)}$ and $d = 1$. Smoothed values are calculated at all time positions and at the position just before the first-time position of the subseries and just after the last-time position. we can get a temporary seasonal series $\ln(C_t^{(k+1)})$ for t ranging from $-n_{(p)} + 1$ to $N + n_{(p)}$.

Step 3: Low-Pass Filtering of Smoothed Subseries. The whole set of low-pass

filtering consists of four processes. First, a moving average of length $n_{(p)}$ is applied to $\ln(C_t^{(k+1)})$. Then, another moving average of length $n_{(p)}$ is performed, followed by a moving average of length 3. Finally, after a loess smoothing with $d = 1$ and $q = n_{(l)}$, we can get the output of low-pass filtering process $\ln(L_t^{(k+1)})$, which is defined over $t = 1, \dots, D$.

Step 4: Detrending Smoothed Subseries. The $(k+1)$ th seasonal component is given by $\ln(S_t^{(k+1)}) = \ln(C_t^{(k+1)}) - \ln(L_t^{(k+1)})$ for $t = 1, \dots, D$. We subtract $\ln(L_t^{(k+1)})$ so that the low-frequency power is avoided in the seasonal component.

Step 5: Deseasonalizing. The deseasonalized series is computed as $\ln(V_t) - \ln(S_t^{(k+1)})$. Note that the deseasonalized series is missing if $\ln(V_t)$ is missing at a particular time position.

Step 6: Trend Smoothing. Loess smoothing is applied to the deseasonalized series with $d = 1$ and $q = n_{(t)}$. In this step, we can get $\ln(T_t^{(k+1)})$. Smoothed values are calculated at each time position $t = 1, \dots, D$ regardless of the missing values.

1.3.2.2 The outer loop of STL

After we get T_t and S_t from the inner loop, the remainder component is given by,

$$\ln(R_t) = \ln(V_t) - \ln(T_t) - \ln(S_t) \quad (1.14)$$

note that $\ln(R_t)$ is not defined where $\ln(V_t)$ is missing. In the outer loop, we define a weight for each time position where $\ln(V_t)$ is observed so that the outlier will have a small or zero weight if $|\ln(R_t)|$ is very large.

Let $m = 6 \text{ median}(|\ln(R_t)|)$. Then the robustness weight at time position t is given by,

$$w_t = B(|\ln(R_t)|/m) \quad (1.15)$$

where B is the bisquare weight function $B(u) = \begin{cases} (1 - u^2)^2 & \text{for } 0 \leq u < 1 \\ 0 & \text{for } u > 1 \end{cases}$.

Now in the smoothing process of Step 2 and 6, the neighborhood weight in the loess regression is multiplied by w_t and it becomes reliability weights.

1.4 Empirical framework

To empirically investigate how periodicity contributes to the volatility-volume relation, Microsoft (MSFT), Xerox (XRX) and SPDR SP 500 TRUST ETF (SPY) are selected at 5-min sampling frequency. The sampling period is from 03 January, 2000 to 31st December, 2016, covering 4277 trading days in total. Descriptive statistics are reported in Table A.1.

Figure A.1 - A.3 present the intraday return, trading volume and number of trades respectively for the selected stocks. Clear U-shaped pattern in return, trading volume and number of trades is observed. This indicates that return, trading volume and number of trades are periodically high at the beginning and ending of the trading day while are periodically low during the lunch time. Although both intraday return and trading activity display the U-shaped curve, their periodic pattern is different which requires separate estimation. For example, trading activity is higher at the end of trading day compared with the beginning of the trading day. However, such observation is reversed for intraday return.

As is discussed in the previous section, periodicity in return is estimated by the Weighted Standard Deviation estimator while that in trading activity is estimated by the STL method. Figure A.4 - A.5 present the Autocorrelation Function (ACF) plot for the raw and filtered trading volume and number of trades. For all stocks, the autocorrelation for the raw trading activity subjects to periodic changes which suggests strongly deterministic seasonality in the data. After estimating and filtering the periodicity from trading activity, we observe a clear long-memory feature for the filtered trading activity in which its autocorrelation is highly persistent and does not decay to zero even after 10 lags. This indicates that the STL method successfully replicates the periodic trading activity and reproduces long-memory stylized facts of trading activity.

[INSERT FIGURE A.1 - A.3 ABOUT HERE]

[INSERT TABLE A.1 ABOUT HERE]

[INSERT FIGURE A.4 - A.5 ABOUT HERE]

1.4.1 Mixture model

Andersen (1996) extends the MDH by building a joint model where the financial return and informed trading volume are normally distributed conditional on information flow variable. As discussed in the introduction, instead of direct modeling on return and trading volume, we focus on modeling the periodicity in return and trading volume via mixture model.

Mixture model is a type of Machine Learning algorithms and assumes all data points are generated from a mixture of Gaussian distributions with unknown mean and variance (Duda et al. (1973)). It is widely used in Finance literature to detect common factors that affect both return volatility and stock price by clustering and has rich empirical application in asset pricing (see, e.g., Frey and McNeil (2002), Ausin and Galeano (2007), Fotopoulos (2017), Mehlitz and Auer (2021)). Mixture model reports not only the cluster results but also the probability that each data point belongs to that cluster. By applying mixture model on periodicity in return and trading activity, we expect to visualize the underlying dynamics and examine if there are common factors that drive the periodicity both in return and trading activity.

Define finite mixture models with K components of form,

$$h(y | x, \varphi) = \sum_{k=1}^K \pi_k f(y | x, \theta_k) \quad (1.16)$$

$$\pi_k \geq 0, \sum_{k=1}^K \pi_k = 1$$

where y is the periodicity in absolute return with conditional density h , x is the periodicity in trading activity (volume and number of trades in our case), π_k is the prior probability of component k , θ_k is the component specific parameters with f being its density function, and $\varphi = (\pi_1, \dots, \pi_K, \theta'_1, \dots, \theta'_K)'$ is the vector of all parameters. Notice that $K = 2$ in our case and we call them Cluster 1 (C_1) and Cluster 2 (C_2) respectively.

The posterior probability that the observation (x, y) belongs to the cluster k is given by,

$$P(k | y, x, \varphi) = \frac{\pi_k f(y | x, \theta_k)}{\sum_k \pi_k f(y | x, \theta_k)} \quad (1.17)$$

The log-likelihood of the sample of N observations $\{(x_1, y_1), \dots, (x_N, y_N)\}$ is given by,

$$\log L = \sum_{n=1}^N \log h(y_n | x_n, \varphi) = \sum_{n=1}^N \log \left(\sum_{k=1}^K \pi_k f(y_n | x_n, \theta_k) \right) \quad (1.18)$$

Given that the above equation cannot be maximized directly, we use the iterative EM algorithm ([Dempster et al. \(1977\)](#)) to estimate the parameter vector φ via the maximum likelihood estimation:

E-step: estimate the posterior probabilities for each observation,

$$\hat{p}_{nk} = P(k | y_n, x_n, \hat{\varphi}) \quad (1.19)$$

The prior probabilities can then be updated as,

$$\hat{\pi}_k = \frac{1}{N} \sum_{n=1}^N \hat{p}_{nk}$$

M-step: maximize the log-likelihood for each component separately by using the posterior probabilities as weights,

$$\max_{\theta_k} \sum_{n=1}^N \hat{p}_{nk} \log f(y_n | x_n, \theta_k) \quad (1.20)$$

The parameters are updated iteratively via E-step and M-step. By default, mixture model stops to update the parameters when the improvement in log-likelihood is less than 0.1.

1.4.2 Linear regression model

Linear regression model is popular in volatility-volume literature (see, e.g., [K. Chan and W.-M. Fong \(2000\)](#), [Giot, Laurent, and Petitjean \(2010\)](#)). It is easy to apply and its parameters naturally carry explicit economic explanation. To examine the explanatory power and predictability of raw and filtered trading activity on return volatility, linear regression model is conducted and used as benchmark model.

We use the realized volatility (RV) to measure the stock return volatility. The market activity is measured by trading volume and the number of trades. The raw and

filtered RV are defined as,

$$RV_t^{raw} = \sqrt{\sum_{i=1}^D r_{i,t}^2} \quad (1.21)$$

$$RV_t^{filtered} = \sqrt{\sum_{i=1}^D \bar{r}_{i,t}^2}$$

where $r_{i,t}$ is the intraday raw return and $\bar{r}_{i,t}$ is the intraday return without periodicity.

Following [Giot, Laurent, and Petitjean \(2010\)](#), the linear regression equation for trading volume is given by,

$$RV_t^{raw/filtered} = \alpha_0 + \sum_{j=1}^{12} \alpha_j RV_{t-j} + \alpha_{13} M_t + \alpha_{14} V_t^{raw/filtered} + \epsilon_t \quad (1.22)$$

Similarly, the OLS regression equation for number of trades is,

$$RV_t^{raw/filtered} = \beta_0 + \sum_{j=1}^{12} \beta_j RV_{t-j} + \beta_{13} M_t + \beta_{14} N_t^{raw/filtered} + \eta_t \quad (1.23)$$

where RV_{t-j} is the lagged RV included to account for persistence in RV, M_t is the Monday dummy which takes the value one when day t is Monday.

1.4.3 Random forest

Under the linear regression framework, we include lagged RV to account for its high persistence. However, RV is also found to be subjected to regime-switching and structural breaks which requires modeling of nonlinear effects ([Pavlidis et al. \(2012\)](#)). Moreover, the predictability of trading activity on return volatility is often controversial under the linear regression framework after controlling other microstructural noise variables. This has brought attention to researchers who conduct rigorous statistical test and confirm that nonlinearity dominates the volatility-volume relation. The nonlinearity has imposed great challenges in statistical modelling and empirical research in volatility-volume relation is restricted by linear models.

To account for the nonlinearity not only in RV itself but also in between RV and trading activity, we use Random Forest (RF) as the nonlinear modeling framework. RF is an ensemble Machine Learning algorithm that uses multiple decision trees to extract

complex information from data and to obtain the regression results by averaging across all trees (Breiman (2001)). Node, branch and leaf are three essential components of decision tree. Each input variable is treated as a node of the decision tree and branches represent decisions. Leaf is at the end of branch with each leaf storing prediction results of the decision tree. Compared with other tree-based methods, RF has certain advantages. First, RF relies on the bootstrap aggregation (bagging) where each tree randomly selects small portions of training sample and is therefore grown by randomly choosing input subset (Belgiu and Drăguț (2016)). This effectively reduces the correlation between trees in other tree-based algorithms and enhance the model performance by drawing results based on plenty of independent trees. Second, through ensembling, RF is more robust to outliers in the data and produces more accurate results. The regression results are made by averaging across all trees which can effectively reduce the impact of outliers. Moreover, such impact will be further alleviated through bagging where training sample data can be learned multiple times with replacement. RF is still favorable when compared with other Machine Learning algorithms. For example, RF is less likely to be overfitting given its simple structure and less tuning hyperparameters. Empirically, RF has found to perform the best in various tasks among other Machine Learning algorithms (see, e.g., Creamer and Freund (2004), W. Chen et al. (2021), Demirer et al. (2021), Gradojevic et al. (2021), Vrontos et al. (2021))

We apply RF to account possible nonlinearity in volatility-volume relation and to examine the importance of periodicity under the high-dimensional learning framework.

1.5 Results

In this section, we first report the results of mixture model and then the regression results.

Figure A.6 - A.7 present the correlation plots between the periodicity in financial return and in trading activity which are modeled through a mix of two Gaussian distributions. For all stocks, two different correlation patterns are observed with light dots indicating a strong correlation and darker dots suggesting the opposite. We name these two correlation patterns as high-correlation cluster (C_1) and low-correlation

cluster (C_2). All data points are nicely classified into two clusters (C_1 and C_2) with C_1 indicating the possible existence of a common factor that affect the periodic pattern in both return and trading activity. C_2 , however, may represent noise which contributes to weaker correlation.

Figure A.8 - A.9 present the posterior probability results generated by mixture model. As is discussed in the previous section, mixture model also reports the probability that each data point belongs to C_1 and C_2 . The Z scores of average statistics shows that the trading activity displays a distinct U curve compared with the return which indicates that trading activity has a different periodicity factor that needs to be estimated accordingly. Moreover, for MSFT and XRX, the relationship between the periodicity in trading activity and in return is more likely to fall into C_1 at the beginning of the trading day while is more prone to C_2 in the rest of the day. This means that periodicity in trading activity has a stronger correlation with that in return at early trading day compared with the rest of the day for MSFT and XRX. This finding may suggest that trading activity is more informative in predicting volatility at early times of the day while its power decays as time elapsing. In addition, two correlation patterns may indicate that the volatility-volume relation may be time-varying during the trading day and the failure to account this may explain the contradictory results in the empirical literature. For SPY, however, the periodicity in trading activity is highly correlated with that in return shown by the relatively higher posterior probability of C_1 throughout the day. This provides strong evidence that a common factor may drive the periodic behavior in both trading activity and return. Our findings offer empirical support of the MDH in a more micro way. The discovered evidence for common factor at the periodicity level reveals the volatility-volume dynamics at micro level given the periodicity is an important internal factor for both return and trading activity,

The Adjusted R Squared results are presented in Figure A.10 - A.11 for the linear regression framework and Random Forest (RF) respectively. To see the impact of periodicity in trading activity and the explanatory power of trading activity on volatility, we report the change in Adjusted R Squared when raw and filtered trading activity are added. Under the linear regression framework, adding the filtered trading activity contributes the most in increasing the Adjusted R Squared for all stocks. This indicates that the explanatory power of trading activity is improved by taking out the

periodicity. Moreover, trading volume contributes more to volatility compared with number of trades given that the increase in Adjusted R Squared is more when both raw and filtered trading volume is integrated into the linear regression model. However, the impact of filtering periodicity from trading activity on Adjusted R Squared under RF framework is not obvious. The Adjusted R Squared, as a goodness-of-fit measure for linear models, might not be suitable to explain the fitting of nonlinear models.

Table A.2 - A.3 report the average out-of-sample performance of predicting 1-day ahead RV under linear regression and RF framework. We leave 20% of data as the out-of-sample dataset. We find that using filtered trading activity along with raw volatility consistently achieves the minimum loss in forecasting RV which indicates that filtering periodicity from trading activity significantly improves the forecasting performance under both linear regression and RF framework. Surprisingly, under both frameworks, we find that filtering periodicity from volatility increases the prediction loss if the periodicity in trading activity has been taken out. This may indicate that periodicity in trading activity has greater impact on volatility-volume relation compared with volatility periodicity which requires researchers attention. Moreover, RF achieves smaller prediction loss for all input attributes compared with linear regression model which shows its superior predictive ability. This reconfirms that the relationship between trading activity and volatility may be highly nonlinear and requires high-dimensional methods in Machine Learning algorithm to extract.

[INSERT FIGURE A.6 - A.7 ABOUT HERE]

[INSERT FIGURE A.8 - A.9 ABOUT HERE]

[INSERT FIGURE A.10 - A.11 ABOUT HERE]

[INSERT TABLE A.2 - A.3 ABOUT HERE]

1.6 Conclusion

Volatility-volume relation have been in the central stage for the empirical research for decades. This is not only because the price changes are associated with the trading activity but also the relation is of great importance for the asset pricing, portfolio allocation and risk management. [Andersen and Bollerslev \(1997\)](#) document the importance of accounting for periodicity in both volatility estimation and forecasting

but fail to consider the prominent periodicity in trading activity. To examine the impact of periodicity in market activity, we propose to use the Seasonal-Trend Decomposition Procedure Based on Locally Estimated Smoothing to estimate the periodic component in the trading activity. The STL-filtered trading activity displays long-memory property which reflects the stylized facts of trading volume and number of trades. Moreover, we find that filtering periodicity improves the explanatory power of both volume and number of trades on realized volatility under the linear regression framework. In addition, the predictive performance of both linear regression and Random Forest is enhanced by filtering periodicity from trading activity.

Besides, the relationship between the average absolute return and trading activity can be better manifested using the mixture model. With more analysis on the posterior probabilities of the mixing components, intraday volume and number of trades tend to display a stronger effect on the absolute return in early trading day relative to the rest of the day. This indicates the intraday volatility-volume relation may be time-varying. Furthermore, the relationship between trading activity and volatility may be highly nonlinear suggested by the superior predictive performance of Random Forest.

Appendix A

Chapter 1

A.1 Tables and Figures

Table A.1: Summary statistics

Stock	Raw Realized Volatility				Raw Number of Trades				Raw Trading Volume			
	Mean	Std.	Skew.	Kurt.	Mean	Std.	Skew.	Kurt.	Mean	Std.	Skew.	Kurt.
MSFT	2.68	3.85	5.37	49.52	1458.35	1340.63	4.33	52.58	702109.90	725832.60	9.05	243.54
XRX	6.55	13.73	8.57	112.06	231.45	324.96	4.50	50.51	92497.23	156058.80	19.20	1860.70
SPY	1.04	2.26	10.32	172.66	3184.59	4199.73	3.36	24.50	1268867.00	1605608.00	3.63	27.49

Note: This table reports descriptive statistics of MSFT, XRX and SPY for raw realized volatility, raw number of trades and raw trading volume. Descriptive statistics includes mean (Mean.), standard deviation (Std.), Skewness (Skew.) and Kurtosis (Kurt.).

Table A.2: Out-of-sample performance of linear regression framework

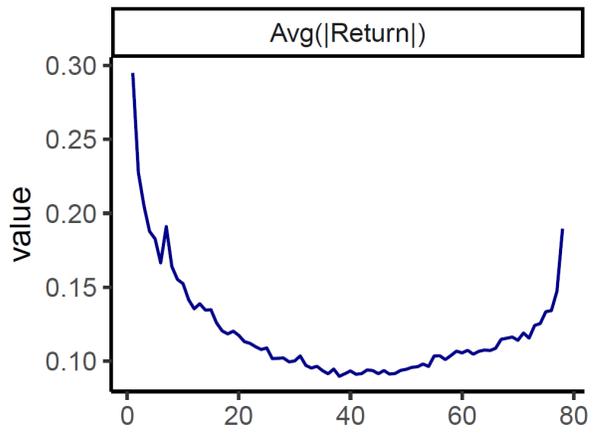
Class	Model	COR	MAE	RMSE	RAE	RRSE
Raw RV	Y = RV-raw; X = Volume-raw	0.7177	0.4851	1.6737	0.4373	0.4919
	Y = RV-raw; X = Volume-filtered-day	0.7319	0.4778	1.6344	0.4305	0.4680**
	Y = RV-raw; X = Volume-filtered-week	0.7241	0.4761	1.6610	0.4319	0.4826
Constant Periodicity Filtered RV	Y = RV-filtered-C; X = Volume-raw	0.7463	0.6321***	1.9057***	0.4444	0.4463***
	Y = RV-filtered-C; X = Volume-filtered-day	0.7332	0.6409***	1.9545***	0.4497	0.4677**
	Y = RV-filtered-C; X = Volume-filtered-week	0.7430	0.6380***	1.9283***	0.4492	0.4529**
Day-by-Day Periodicity Filtered RV	Y = RV-filtered-D; X = Volume-raw	0.7275	0.6506***	1.9899***	0.4554*	0.4758
	Y = RV-filtered-D; X = Volume-filtered-day	0.7222	0.6569***	2.0199***	0.4590*	0.4885
	Y = RV-filtered-D; X = Volume-filtered-week	0.7167	0.6460***	2.0346***	0.4530*	0.4931

Note: This table reports the out-of-sample performance of forecasting 1-day-ahead RV under linear regression Framework. The performance matrices are pearson correlation (COR), mean absolute error (MAE), square root mean squared error (RMSE), relative mean absolute error (RAE), relative square root mean squared error (RRSE). The highest three correlation values are highlighted bold while the lowest three errors in other error matrices are highlighted bold. *, ** and *** indicate that the relative forecasting differences between the corresponding model and the linear regression model with raw volatility and raw trading volume are significant at 10%, 5% and 1% level using the Diebold–Mariano test (Newey–West heteroscedasticity consistent covariance matrix estimator).

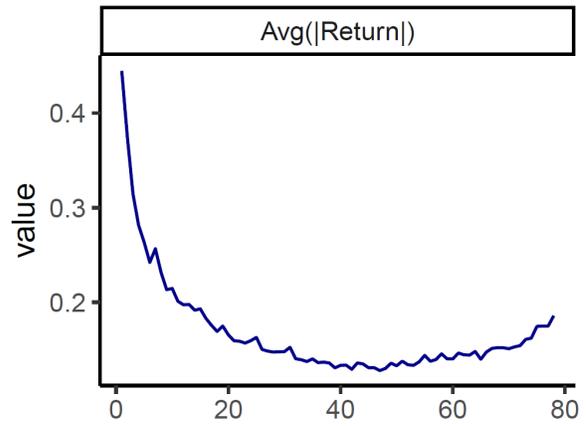
Table A.3: Out-of-sample performance of Random Forest framework

Class	Model	COR	MAE	RMSE	RAE	RRSE
Raw RV	Y = RV-raw; X = Volume-raw	0.8256	0.3450***	1.3511***	0.3110***	0.3205***
	Y = RV-raw; X = Volume-filtered-day	0.8209	0.3461***	1.3694***	0.3119***	0.3286***
	Y = RV-raw; X = Volume-filtered-week	0.8218	0.3485***	1.3673***	0.3161***	0.3270***
Constant Periodicity Filtered RV	Y = RV-filtered-C; X = Volume-raw	0.8428	0.4667	1.5472**	0.3282***	0.2941***
	Y = RV-filtered-C; X = Volume-filtered-day	0.8345	0.4749	1.5856**	0.3332***	0.3078***
	Y = RV-filtered-C; X = Volume-filtered-week	0.8405	0.4708	1.5677**	0.3315***	0.2993***
Day-by-Day Periodicity Filtered RV	Y = RV-filtered-D; X = Volume-raw	0.8365	0.4731	1.5883**	0.3312***	0.3032***
	Y = RV-filtered-D; X = Volume-filtered-day	0.8420	0.4687	1.5687**	0.3275***	0.2946***
	Y = RV-filtered-D; X = Volume-filtered-week	0.8331	0.4734	1.6115	0.3319***	0.3093***

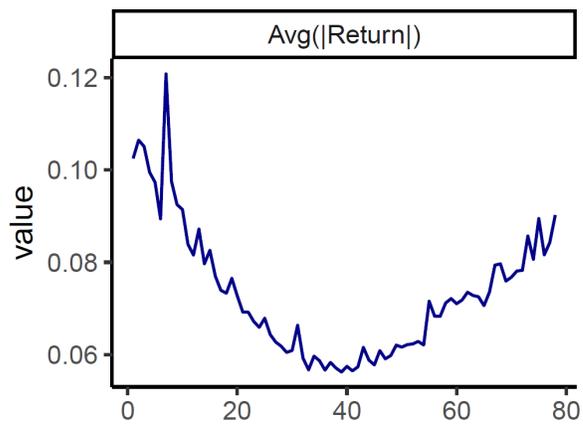
Note: This table reports the out-of-sample performance of forecasting 1-day-ahead RV under Random Forest Framework. The performance matrices are pearson correlation (COR), mean absolute error (MAE), square root mean squared error (RMSE), relative mean absolute error (RAE), relative square root mean squared error (RRSE). The highest three correlation values are highlighted bold while the lowest three errors in other error matrices are highlighted bold. *, ** and *** indicate that the relative forecasting differences between the corresponding model and the linear regression model with raw volatility and raw trading volume are significant at 10%, 5% and 1% level using the Diebold–Mariano test (Newey–West heteroscedasticity consistent covariance matrix estimator).



(a) MSFT



(b) XRX



(c) SPY

Figure A.1: Intraday return

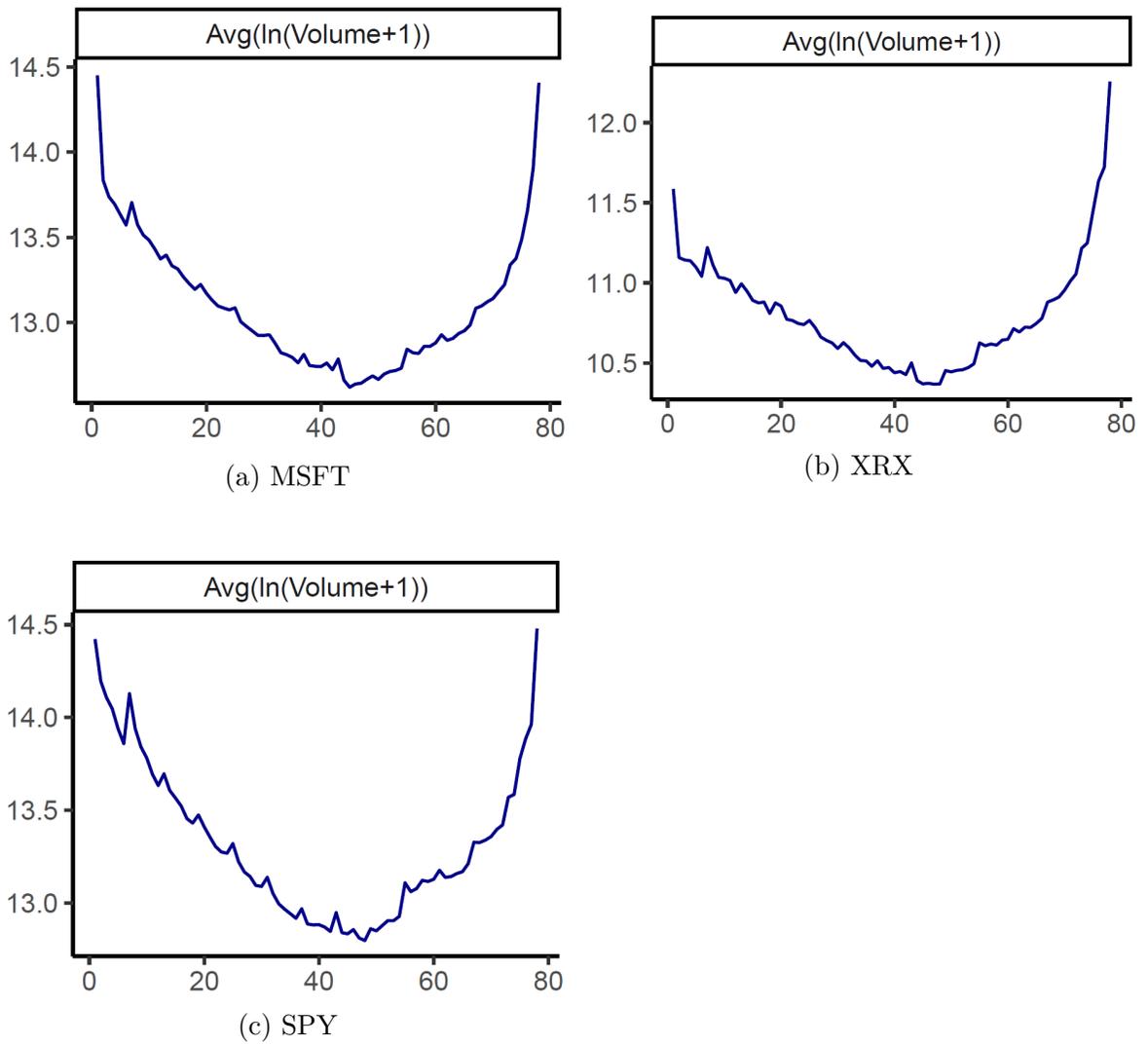
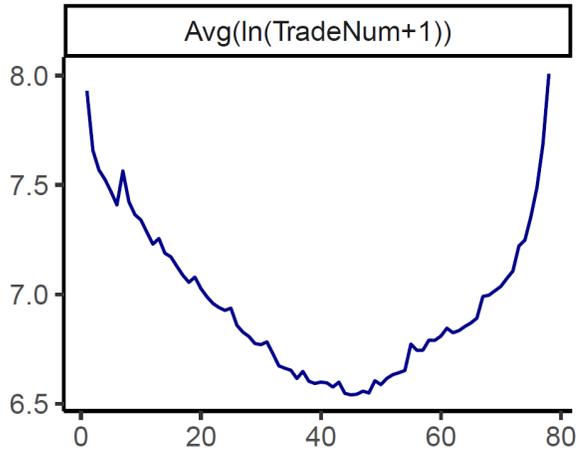
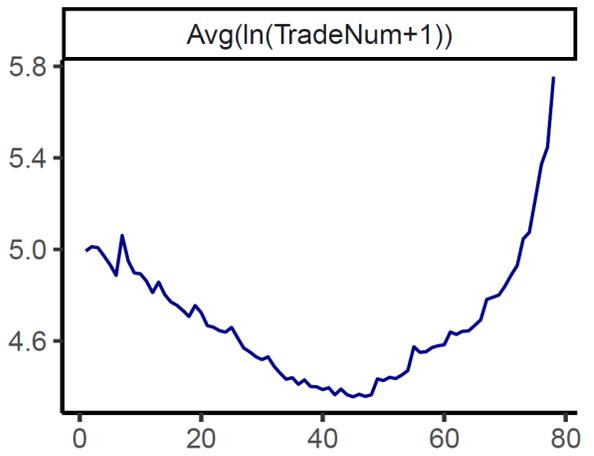


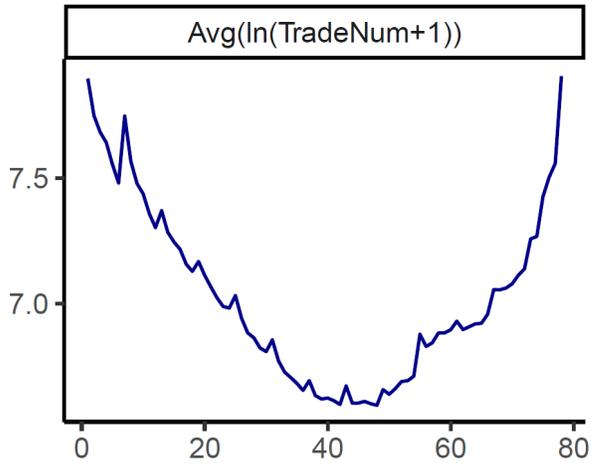
Figure A.2: Intraday raw trading volume



(a) MSFT

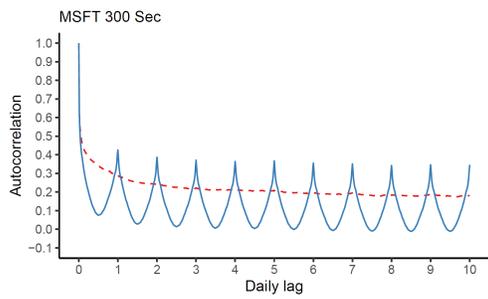


(b) XRX

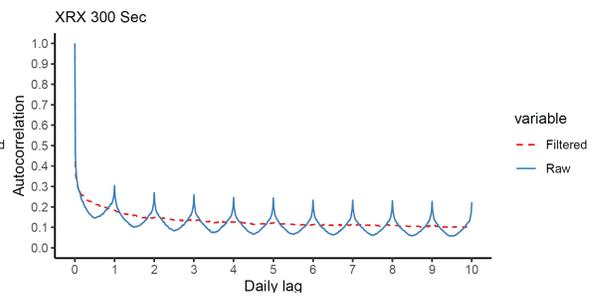


(c) SPY

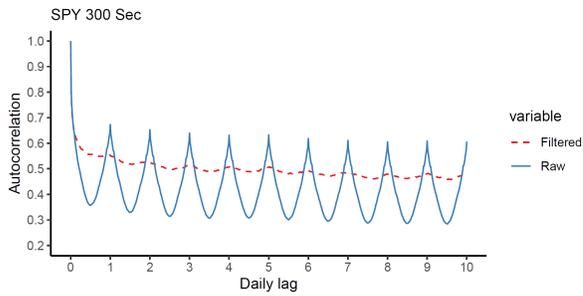
Figure A.3: Intraday raw number of trades



(a) MSFT

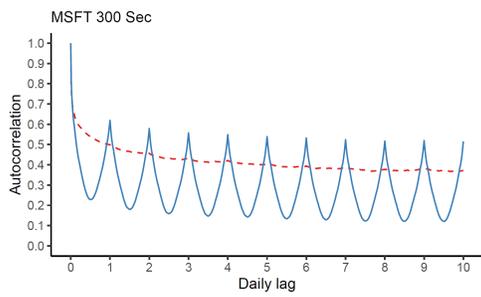


(b) XRX

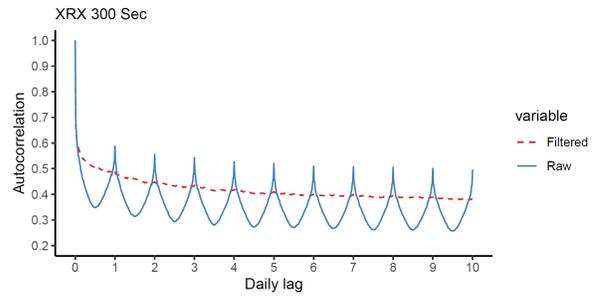


(c) SPY

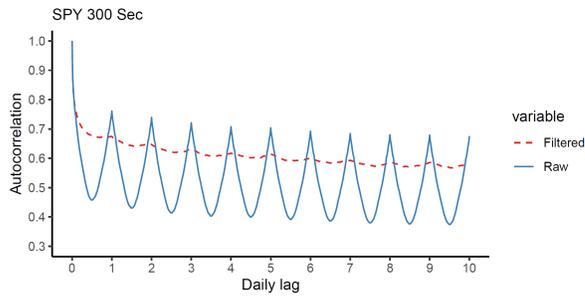
Figure A.4: ACF plot for raw and filtered trading volume



(a) MSFT



(b) XRX



(c) SPY

Figure A.5: ACF plot for raw and filtered number of trades

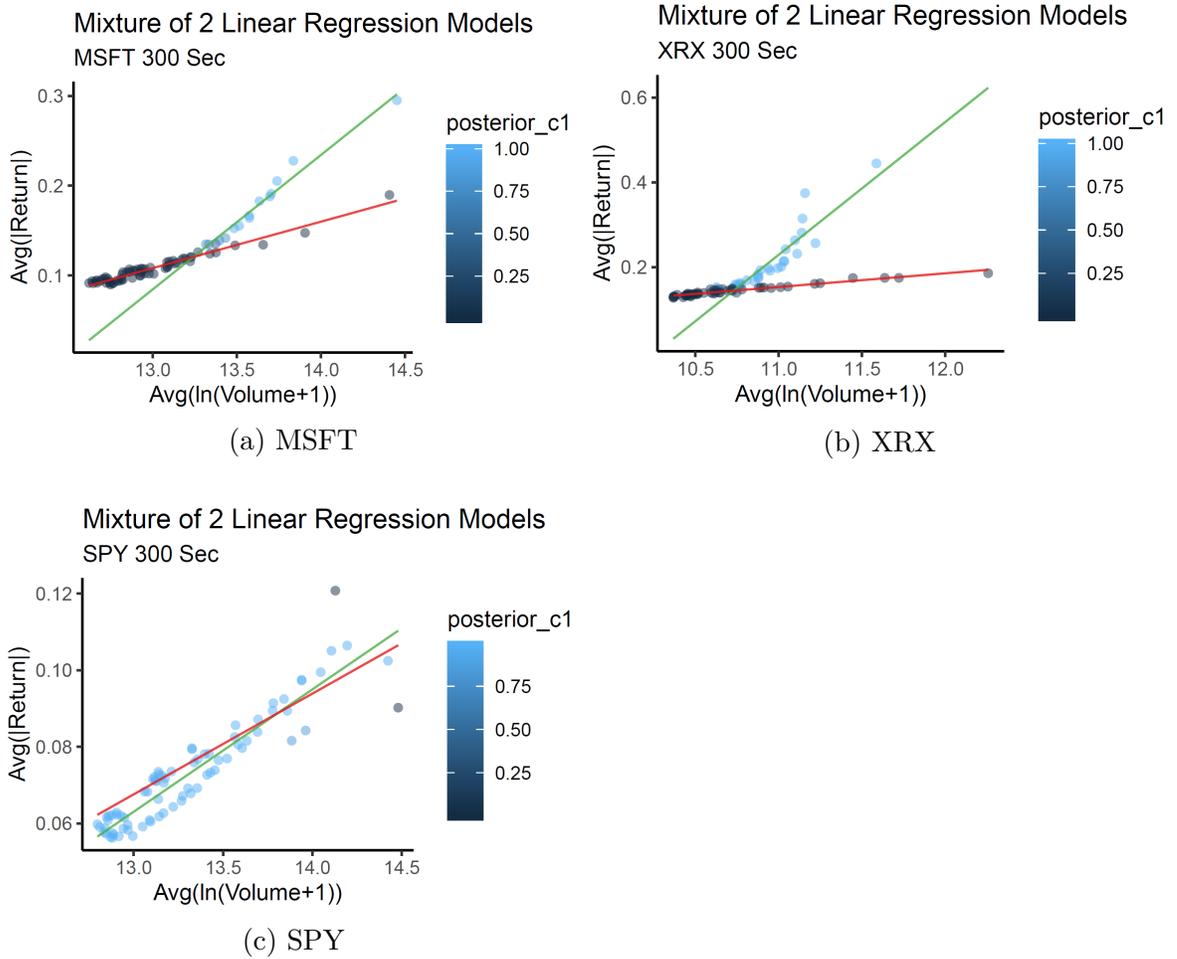


Figure A.6: Mixture model correlation plot for trading volume

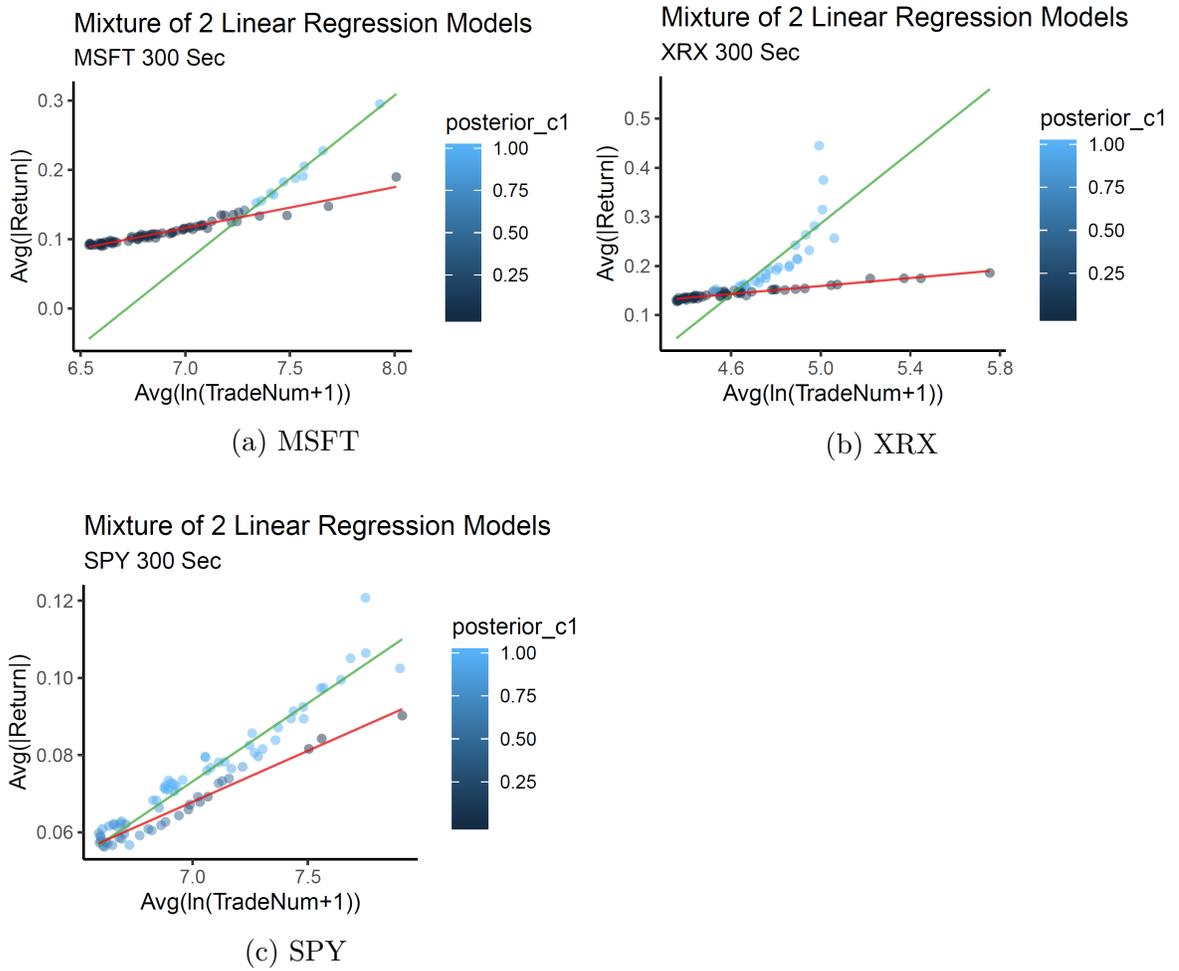


Figure A.7: Mixture model correlation plot for number of trades

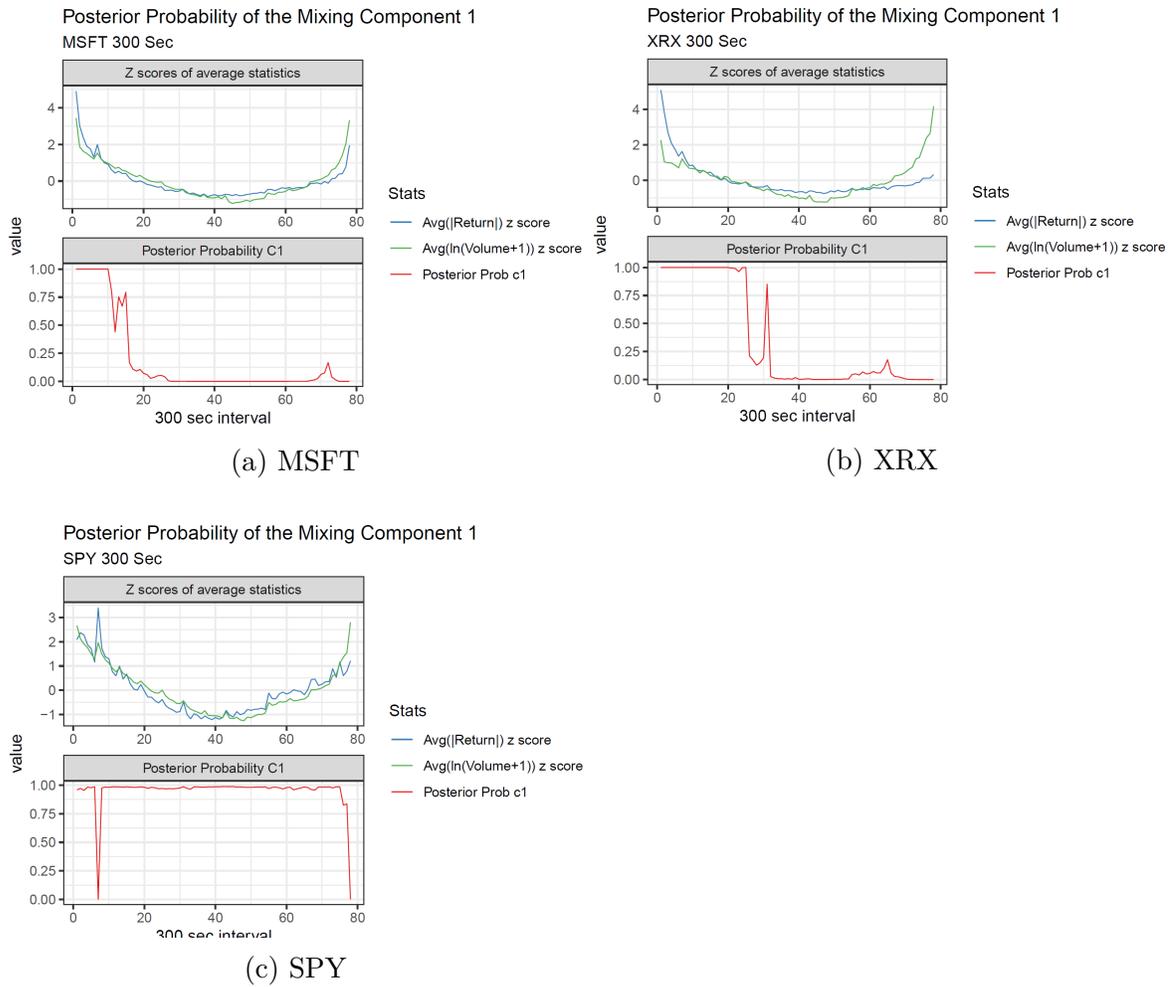


Figure A.8: Mixture model posterior probability plot for trading volume

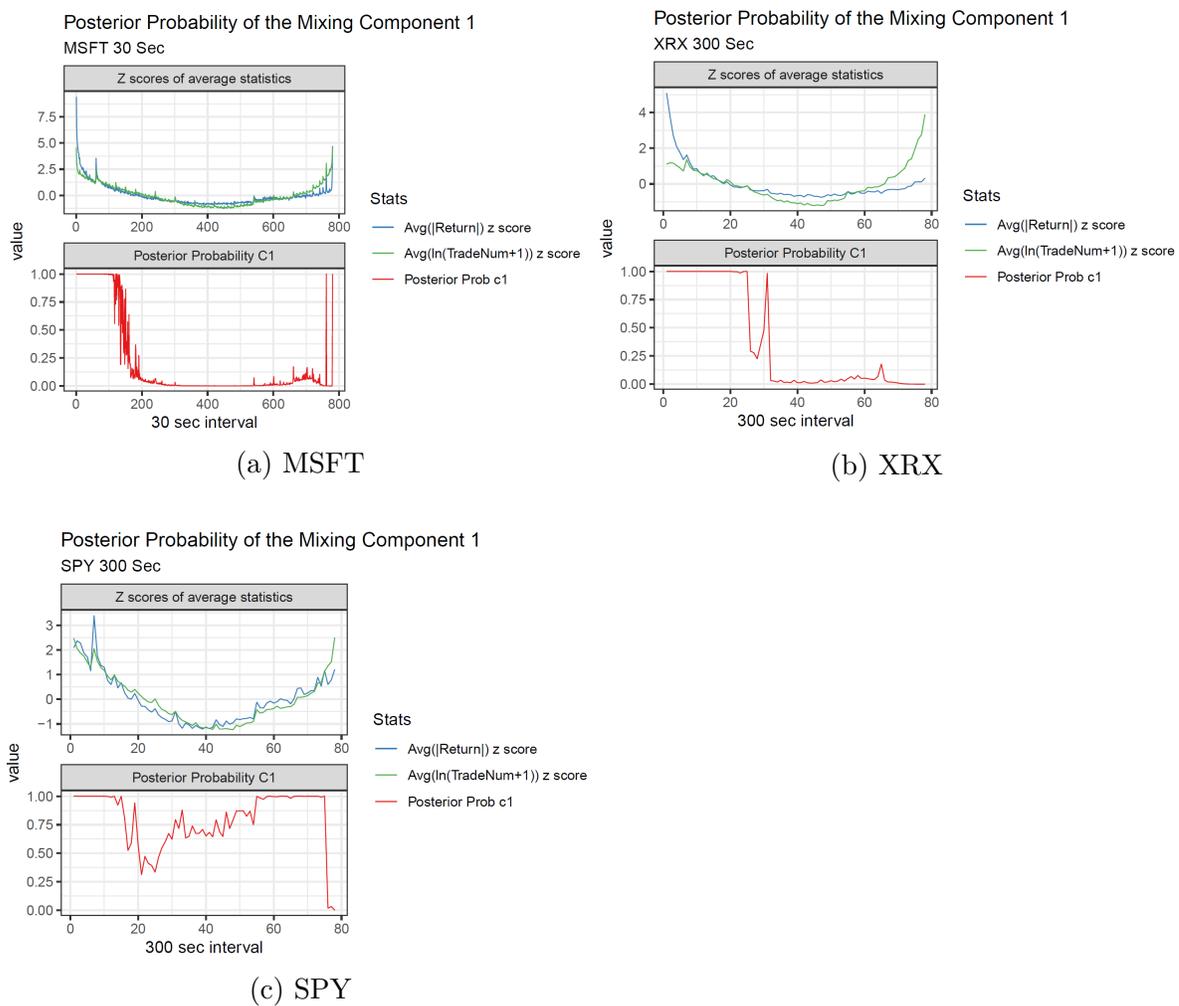


Figure A.9: Mixture model posterior probability plot for number of trades

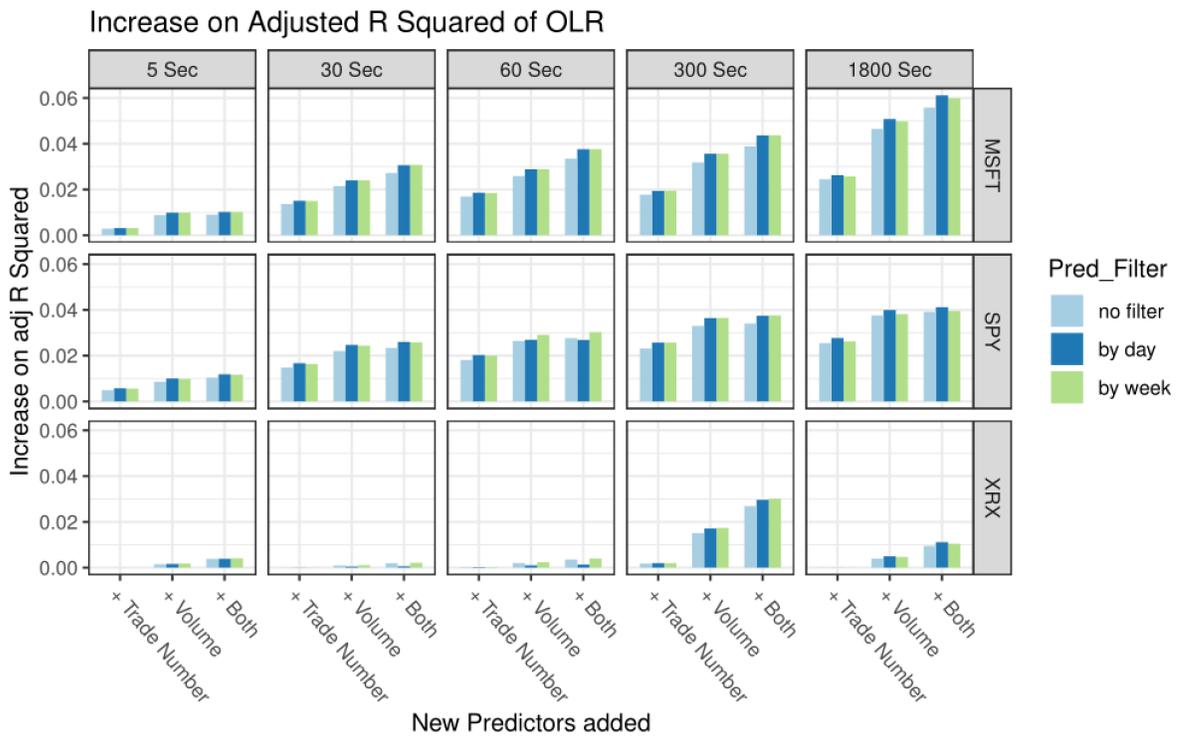


Figure A.10: In-sample Adjusted R squared results under linear regression framework

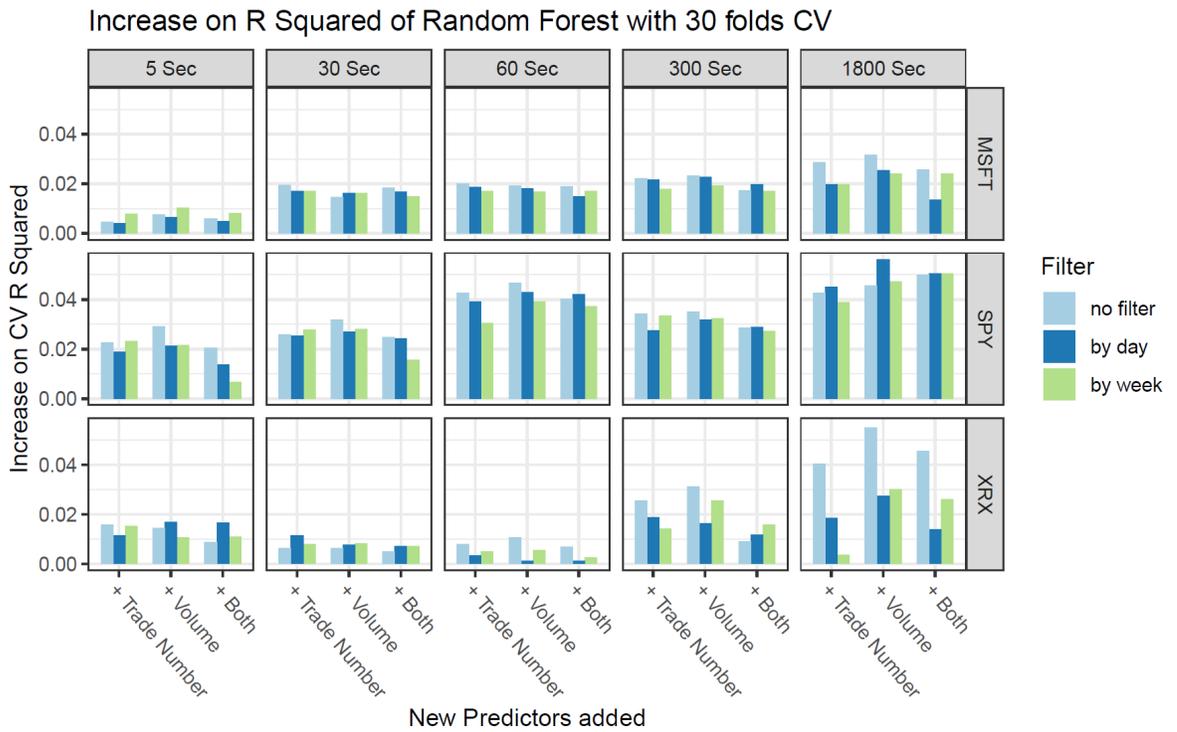


Figure A.11: In-sample Adjusted R squared results under Random Forest framework

Chapter 2

Forecasting RV: A Hybrid Model Integrating BiLSTM with HAR-type Models

2.1 Introduction

Accurate volatility forecasting is critical for various financial applications such as asset pricing, asset allocation and risk management. The key task of financial volatility modeling is to model the stylized facts of financial returns and volatility. Since the early work of [Engle \(1982\)](#) and [Bollerslev \(1986\)](#), Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) family models have been extended to capture many important stylized facts such as leverage effects, long memory, leptokurtosis and volatility clustering. However, for GARCH-type models, problems arise in distributional assumption, estimation complexity and economic interpretation of estimated parameters. In addition, the failure to account nonlinearity distorts the definition of data generating process and imposes great challenges on model estimation. With the substantial development of computational resources, the Artificial Neural Network models (ANNs) have been extensively used to predict stock return volatility, showing superior forecasting performance against Financial Time Series models (see, e.g., [Tino et al. \(2001\)](#), [Hamid and Iqbal \(2004\)](#), [Khan \(2011\)](#), [Nikolaev et al. \(2013\)](#), [Xiong et al. \(2015\)](#), [Doering et al. \(2017\)](#), [Zhou et al. \(2019\)](#), [Bucci \(2020\)](#), [Horvath](#)

et al. (2021)). One of the advantages of this approach is that it can well approximate a wide class of linear and nonlinear functions, forming the input-output map by learning the data rather than assuming the data generating process which complements the limitations of traditional Financial Time Series models.

However, ANNs have been criticized for the "black box" problem which makes the results interpretation very difficult since it does not provide any insights about the resulted input-output map (Lai et al. (2009)). To improve the interpretability of Neural Network models, many studies suggest hybrid modeling, integrating Financial Time Series models with ANNs. Donaldson and Kamstra (1997) develop seminonparametric GARCH-type models which uses ANN transformation to capture nonlinear relationship between past return innovations and future volatility. The logistic function, the most popular nonlinear transformation function used in the ANN literature, is applied to approximate the nonlinearities. The empirical results show that the GARCH-type models using logistic transformation generally outperform other alternatives in forecasting volatility of four stock market index. Another type of hybrid modeling is purely nonparametric approach through the usage of more informative input features (see, e.g., Sjöberg et al. (1995), Faraway and Chatfield (1998), Nanopoulos et al. (2001), Dash et al. (2007), Zheng et al. (2016), J.-F. Chen et al. (2016), Petneházi (2019), Hewamalage et al. (2021)). Features generated by Financial Time Series models have become increasingly popular given their explicit economic interpretation and successful reproduction of important stylized facts of financial volatility. For example, Roh (2007) propose hybrid models integrating Deep Feedforward Neural Network (DFN) with GARCH, EWMA and EGARCH model respectively. Rather than using only past volatility as input for DFN, features generated from those GARCH-type models provide valuable information on the characteristics of stock market volatility such as leverage effects, volatility clustering and excessive kurtosis. The empirical results show that hybrid models significantly outperform both single GARCH-type models and single DFN in forecasting volatility of KOSPI 200 index. This indicates that the predictability and interpretability of DFN can be enhanced by integrating features generated from GARCH-type models. Roh's study has been extended by integrating other GARCH-type models with various Neural Network models in forecasting stock index volatility (see, e.g., Tseng et al. (2008), Hajizadeh et al. (2012), Kristjanpoller et al. (2014),

Maciel et al. (2016), Kim and Won (2018)).

Although GARCH-type models successfully reflect some empirical characteristics of in-sample volatility, such characteristics are hard to be reproduced in the out-of-sample volatility forecasting (Andersen, Bollerslev, Diebold, and Labys (2003)). Moreover, inappropriate distributional assumptions on financial returns and failure of reproducing other important stylized facts (i.e., long memory) may further deteriorate the quality of features generated by GARCH-type models and integrated into the ANNs. With the development of high-frequency data, Andersen and Bollerslev (1998) derive realized volatility (RV), which is a new nonparametric measure of return volatility and is defined as the sum of intraday squared returns. Barndorff-Nielsen and Shephard (2002) show that RV is model-free and unbiased estimate of ex post return variation. RV displays long-memory feature which has been previously modeled by AutoRegressive Fractionally Integrated Moving Average (ARFIMA) process (see, e.g., Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2003)). However, ARFIMA lacks explicit economic interpretation and is nontrivial to be estimated. To improve this, Corsi (2009) proposes Heterogeneous AutoRegressive (HAR) model which is an additive cascade model of three volatility components defined over different time horizons, i.e., daily, weekly and monthly. HAR successfully reproduces the main features of financial returns and volatility (i.e., fat tails, self-similarity and long memory). In addition, compared with GARCH-type models, HAR achieves superior performance in forecasting RV across various financial assets with less estimated parameters. It is also easy to implement and interpret. Despite RV's prominent long-memory feature, it is also found to be subject to regime-switching and structural breaks which requires modeling of nonlinear effects. The nonlinearity has imposed great challenges in modeling and forecasting RV. McAleer and Medeiros (2008) extend the standard HAR into the Heterogeneous Autoregression with Multiple-Regime Smooth Transition (HARST) model, merging long memory and nonlinearities in RV. However, such method requires huge amount of data to determine the number of regimes before estimating the model and the sample log-likelihood can easily trap in local maxima, leading to poor out-of-sample performance (Pavlidis et al. (2012)).

To facilitate better modeling of nonlinearities, semiparametric HAR-based hybrid models integrating Neural Networks have been developed. For example, Hillebrand

and Medeiros (2010) propose the Neural Network HAR model (NNHAR) to forecast RV with an additional logistic transformation term to capture the nonlinearities in RV components. Surprisingly, NNHAR only outperforms the standard log-linear HAR during low-volatility periods and such advantages are eliminated by bagging the forecasts of log-linear HAR model. This indicates that bagging may be more efficient than directly modeling the nonlinearities. M. Fernandes et al. (2014) extend the NNHAR model into the NNHARX model by including macroeconomic variables to forecast implied volatility (IV). The empirical results are mixed and NNHARX only beats the standard HAR in 22-day forecasting window. This indicates that the strong persistence of IV may predominates the nonlinearities over shorter forecasting horizons. Instead of using the logistic function, Psaradellis and Sermpinis (2016) pioneer to use Support Vector Machines (SVM) with Radial Basis Function (RBF) kernel for nonlinear transformation in HAR to forecast IV. Moreover, unlike the previous studies where the hyperparameters (i.e., the number of hidden nodes in the nonlinear transformation term) are selected according to authors' discretion, the value of those parameters are optimized and determined by walk-forward validation. The proposed approach consistently outperforms standard HAR and NNHAR, showing significant improvement in the out-of-sample forecasting of IV.

The puzzling results of semiparametric HAR-based hybrid models may reflect the following limitations of such approach: (1) This kind of hybrid models combines the HAR components with an extra nonlinear transformation term (i.e., the logistic transformation term). Such technique is called the cooperative modular combination and aims for better representing the different behaviors of time series (Sharkey (2002)). Regarding the application, the original HAR is first applied to the time series and then its residuals are modeled through nonlinear transformation function. However, this implicitly assumes that the residuals of HAR include prominent nonlinear patterns that can be modeled by Neural Networks (Zhang and Qi (2005)). Violation of such assumptions may deteriorate the performance of semiparametric HAR-based hybrid models. In addition, the simple additive combination may underestimate the relationship between the HAR components and the nonlinear term (Taskaya-Temizel and Casey (2005)); (2) In viewing of (1), the choice of nonlinear transformation function may also be problematic in such hybrid models. The NNHAR model which uses

the logistic function hardly improves the standard HAR while the model uses RBF kernel achieves significant improvement. This indicates that the nonlinear functions are chosen by authors rather than being reasonably optimized. Different nonlinear transformation functions process the input information in a very different way which may lead to undesirable outcomes; (3) The functioning of Neural Networks rely on the configuration of input layer, hidden layer and output layer. Although the hidden layer handles nonlinear transformation and creation of high-dimensional features, the design of input layer is found to be an important factor which would affect the performance of subsequent layers. The semiparametric HAR-based hybrid models only rely on the function of hidden layer (i.e., nonlinear transformation) and disregard other components of ANNs which may induce performance degeneration (Zhang and Berardi (2001), Taskaya-Temizel and Casey (2005)).

To overcome above limitations, we propose the nonparametric ANN-based hybrid models which takes various input features generated from HAR-type models. This approach differs from the semiparametric HAR-based hybrid models in the following aspects. First, the proposed model can well approximate a wide range of linear and nonlinear functions between the input features and the target output without assuming that nonlinear patterns only exist in the residuals. Second, our model consists of the input layer, hidden layer and output layer which fulfills the standard structure of ANNs. The hyperparameters of the model are reasonably optimized through validation rather than being chosen subjectively. Third, to predict the target output, the model enables learning not only from the past values but also from the future values. This allows bidirectional learning and a much longer window to capture the long-term dependence in RV.

The main contributions of this chapter are as follows:

(1) Two nonparametric ANN-based hybrid models are proposed to forecast RV on the basis of different Neural Network architectures. To allow different pattern-learning abilities, Deep Feedforward Neural Network (DFN) and Bidirectional Long Short-term Memory Network (BiLSTM) are used as the Neural Network basis in the proposed hybrid models. In addition, rather than simply adding a nonlinear transformation term to the original HAR model, we relax the assumptions used in the semiparametric HAR-based hybrid model by embracing a full Neural Network setting.

(2) Despite significant advantages of applying a full Neural Network setting, ANNs are often criticized by lacking explicit economic explanations. We improve the interpretability of the proposed hybrid models by feeding features generated by HAR-type models. The standard HAR and HAR-TCJ proposed by [Corsi et al. \(2010\)](#) successfully reproduce several stylized facts of RV (i.e., self-similarity, presence of jumps etc.). Moreover, the parameters of HAR-type models measure the impact of different volatility components on future RV which makes the HAR-type components more informative than past RV alone. By taking the HAR-type components as input features, our proposed hybrid models outperform all benchmarks in forecasting RV with the BiLSTM-based hybrid model being the best performer across all forecasting horizons in all subsamples.

The remainder of this chapter is structured as follows. Section 2 introduces the structure and design of the proposed model. Section 3 presents the data and experiment. Section 4 shows the out-of-sample prediction results. Finally, Section 5 concludes.

2.2 Models

2.2.1 HAR model

Our hybrid models are based on the HAR proposed by [Corsi \(2009\)](#). The HAR model successfully replicates main empirical features of financial returns such as long memory, fat tails and self-similarity. It has strong empirical performance in forecasting RV across various financial assets. In this study, we consider the RV estimator proposed by [Andersen and Bollerslev \(1998\)](#), which is equal to the square root of the sum of intraday squared returns:

$$RV_t = \sqrt{\sum_{j=1}^M r_{j,t}^2} \quad (2.1)$$

where $r_{j,t}$ stands for the intraday returns within each time interval in day t . To introduce the HAR model, we denote the average RV over the previous (h) days by:

$$RV_t^h = \frac{1}{h} \sum_{i=1}^h RV_{t-i+1} \quad (2.2)$$

Therefore, the weekly (5-day) and monthly (22-day) RV are the corresponding averages of daily RV, which is given by $RV_t^{(w)} = \frac{1}{5} \sum_{i=1}^5 RV_{t-i+1}$ and $RV_t^{(m)} = \frac{1}{22} \sum_{i=1}^{22} RV_{t-i+1}$ respectively. The logarithmic form of HAR is given by:

$$\log(RV_{t+1}) = \beta_0 + \beta_d \log(RV_t) + \beta_w \log(RV_t^{(w)}) + \beta_m \log(RV_t^{(m)}) + \epsilon_{t+1} \quad (2.3)$$

where ϵ_{t+1} is a sequence of independently and identically distributed (i.i.d) innovations with zero mean. Coefficient estimates, β_0 , β_d , β_w and β_m can be consistently obtained by a standard OLS regression. The HAR component, $\beta_d \log(RV_t)$, $\beta_w \log(RV_t^{(w)})$ and $\beta_m \log(RV_t^{(m)})$, reflect the magnitude of the heterogeneous reaction of different market participants to a given price change. The Autocorrelation Function (ACF) plot and Partial Autocorrelation Function (PACF) plot for SPY's RV are provided in Figure B.3.

Although HAR has very strong empirical performance in predicting RV, it fails to capture any nonlinearity in the RV dynamics. As a result, some researchers suggested to use smooth transition or threshold model to estimate RV ([McAleer and Medeiros \(2008\)](#)). However, this class of models generally requires substantial amount of data to identify the states and shows poor out-of-sample forecasting results ([Bucci \(2020\)](#)).

2.2.2 HAR-TCJ model

We also consider an extension of the HAR, the HAR-TCJ model proposed by [Corisi et al. \(2010\)](#). By implementing a more robust test to detect jumps, the HAR-TCJ builds on the HAR-CJ model in [Andersen, Bollerslev, and Diebold \(2007\)](#) and better incorporates jumps in the estimation of volatility models. The logarithmic form of HAR-TCJ is defined as:

$$\log(RV_{t+1}) = \beta_0 + \beta_{c,d} \log(T\hat{C}_t) + \beta_{c,w} \log(T\hat{C}_t^{(w)}) + \beta_{c,m} \log(T\hat{C}_t^{(m)}) + \beta_{j,d} \log(1 + T\hat{J}_t) + \epsilon_{t+1} \quad (2.4)$$

where $T\hat{C}_t^{(w)}$ and $T\hat{C}_t^{(m)}$ are the weekly and monthly averages of $T\hat{C}_t$ as in Eq.

The jump component is estimated based on the Threshold Bi-power variation

(TBPV) measure given by:

$$\hat{T}J_t = I_{\{C-Tz > \Phi_\alpha\}}(RV_t - TBPV_t)^+ \quad (2.5)$$

where $C - Tz$ is the test statistics of the jump test based on the corrected TBPV and Φ_α is the cumulative distribution function of the normal distribution at level $\alpha = 99.9\%$, see more details in [Corsi et al. \(2010\)](#). The TBPV measure is derived as:

$$TBPV_t = \mu^{-2} \sum_{j=2}^M |r_{t,j-1}| |r_{t,j}| I_{\{r_{t,j-1}^2 \leq v_{j-1}\}} I_{\{r_{t,j}^2 \leq v_j\}} \quad (2.6)$$

with $\mu = (2/\pi)^{0.5}$. The threshold v_t is equal to $c_v^2 \hat{V}_t$ with \hat{V}_t being the local variance estimator and $c_v = 3$ suggested by [Corsi et al. \(2010\)](#). The continuous part of variation is therefore given by:

$$TC_t = RV_t - \hat{T}J_t \quad (2.7)$$

The HAR-TCJ components, $\beta_{c,d} \log(TC_t)$, $\beta_{c,w} \log(TC_t^{(w)})$, $\beta_{c,m} \log(TC_t^{(m)})$ and $\beta_{j,d} \log(1 + \hat{T}J_t)$, provide more detailed features of realized volatility by accommodating the robust jump effects.

2.2.3 Neural Network models

ANNs are powerful nonparametric tools, mimicking the structure of the human brain and aiming for modeling and predicting the unobserved function underlying the observed data ([Arnerić et al. \(2014\)](#)). Empirical research suggests that ANNs are particularly suitable for forecasting Financial Time Series that exhibit nonlinear behaviors, like stock market volatility or stock market returns ([Maheu and McCurdy \(2002\)](#)), by learning and mapping nonlinear structure that linear model cannot process. In this way, Neural Network models can be implemented without any assumptions on the underlying data generating process.

The learning ability of ANNs can be improved by increasing the number of hidden layers and hidden nodes. However, the training algorithm can be very slow and the out-of-sample regularization can be poor when the entire network becomes too complicated. Some rules have been proposed in the literature to find the optimal number of hidden

layers and hidden nodes so that the good learning performance can be generalized to the testing sample (Sheela and Deepa (2013)). However, there is no uniform solution to this issue. Stinchcombe and White (1992) prove that Neural Network with a single hidden layer and sufficient hidden nodes is a universal approximator, meaning that such network is sufficient in approximating a wide range of linear and nonlinear functions. Following Bucci (2020), a single hidden layer Neural Network is assumed throughout this chapter, leaving the number of hidden nodes to be determined separately.

To allow different pattern learning behavior, two representative Neural Network models, DFN and BiLSTM, are used as Neural Network basis in the proposed hybrid models. DFN is one of the earliest developed ANNs and it consists of input layer, hidden layer and output layer. Fixed amount of information is input into DFN at time t to predict the target variable value at $t + 1$. Clearly, DFN is a short-memory model as it does not capture any possible long-range dependence in the data. The workings of BiLSTM, however, relies on memory cells which controls the amount of information flow through the input gate, forget gate and output gate. BiLSTM takes input at each time point until time t and only the important information will be processed into the next cell state. Therefore, BiLSTM captures both short-term and long-term memory in the data. These two Neural Network basis are further enhanced by the proposed hybrid models in the following section.

2.2.3.1 Deep Feedforward Neural Network model

DFN is a classical deep learning model. It assumes that information moves forward from the input layer to the output layer and uses the output of the previous layer as the input of the current layer. The calculation continues until the output layer is reached. The output of a typical DFN with single hidden layer is given by:

$$\hat{R}V_t = F(\beta_0 + \sum_{j=1}^q f(x_t w'_j) \beta_j), \quad (2.8)$$

where F is the activation function of the output layer, β_0 is the bias to be added to the final output, f is the activation function of hidden layer, $x_t = \{1, x_{1,t}, \dots, x_{i,t}\}$ is the $1 * (1 + 1)$ vector of features at time t , $w_j = \{w_{1,j}, \dots, w_{i+1,j}\}$ is the $1 * (i + 1)$ vector of weights, measuring the connection between the features and hidden node j , total

number of hidden nodes is q .

F and f can be chosen from a variety of functions in the empirical research. Given the nature of forecasting a real number (RV) in this study, we choose the identity function as the activation function of the output layer, that is, $F(a) = a$. In this sense, Eq.(2.8) can be written as:

$$\hat{RV}_t = \beta_0 + \sum_{j=1}^q f(x_t w'_j) \beta_j. \quad (2.9)$$

For the activation function of the hidden layer f , any continuous, nonlinear, differentiable and monotonic function can be applied. Common choices are sigmoid function and hyperbolic tangent function. However, these two functions suffer from the vanishing gradient problem and are generally slow in computing. Therefore, we choose the RELU function, that is, $f(a) = \max(0, a)$, to allow faster learning and better performance.

2.2.3.2 Bidirectional Long Short-term Memory Network model

DFN is perceived by researchers as static network given the fixed amount of information used to predict the target output variable at time t . Although DFN may be able to capture the temporal dependence contained in the data, it has no memory. The Long Short-term Memory (LSTM) Neural Network, a representative recurrent Neural Network model proposed by [Hochreiter and Schmidhuber \(1997\)](#), is proposed to overcome this problem by allowing information feedback. Instead of using fixed amount of information as does in DFN, LSTM utilizes all available input information up to the time t to predict the target output. It becomes extremely popular in predicting financial time series where nonlinear time dependence and long memory are prominent (see, e.g., [Schittenkopf et al. \(2000\)](#), [Tino et al. \(2001\)](#), [Bucci \(2020\)](#)). Compared with DFN, LSTM is characterized by using memory cells instead of hidden units to process information. Figure B.1 shows the structure of a LSTM memory cell.

LSTM controls information flow through the functioning of three gates on the cell state c_t : an input gate (i_t), a forget gate (f_t) and an output gate (o_t). Each gate performs different roles, as follows.

The input gate i_t controls the extent of information to be added to the cell, where

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (2.10)$$

The forget gate f_t controls the extend of information to be discarded from the cell, where

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (2.11)$$

The output gate o_t controls the extent of output information from the cell, where

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (2.12)$$

and σ_g is the gate activation function which is by default the sigmoid function. The response of the gate activation function is between 0 and 1, controlling the amount of information flow in that gate.

A nonlinear function, the hyperbolic tangent function (\tanh), is applied to generate a vector of candidate values, \tilde{c}_t , to update the cell state c_t as follows:

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (2.13)$$

The current cell state c_t is now updated by the operation of input gate and forget gate, as follows:

$$c_t = i_t \odot \tilde{c}_t + f_t \odot c_{t-1} \quad (2.14)$$

The output of LSTM memory cell at time t is then given by:

$$h_t = o_t \odot \tanh(c_t) \quad (2.15)$$

where \odot is the Hadamard production function.

Finally, the memory cell output h_t is projected to obtain the predicted output $\hat{R}V_t$, which is,

$$\hat{R}V_t = W_t h_t \quad (2.16)$$

where W_t is a projection matrix used to reduce the dimension of h_t .

However, for the highly time-dependent financial time series, future input informa-

tion coming up later than time t can also be helpful in predicting the target value at time t . The Bidirectional Long Short-term Memory (BiLSTM) Neural Network is therefore proposed by [Schuster and Paliwal \(1997\)](#) to include both the past and the future input information of a specific time frame to predict the target output. BiLSTM computes the forward output sequence \vec{h} using input information from time $t = 1$ to T and the backward output sequence \overleftarrow{h} using the information from time $t = T$ to 1. Fig presents the structure of BiLSTM Neural Network. Both the forward and the backward output sequence are calculated using the same LSTM updating equations specified in Eq.(2.10) - (2.15). The predicted output \hat{RV}_t is now given by:

$$RV_t = \sigma(\vec{h}, \overleftarrow{h}) \quad (2.17)$$

where σ is a function merging the forward and the backward output sequences. It can be a summation function, a multiplication function, a concatenating function or an average function. We choose the multiplication function for σ as it achieves the minimum validation loss discussed below.

As the above DFN, we use a single BiLSTM layer for the hidden layer and the number of hidden units is optimized separately. In addition, recurrent layer dropout regularization is applied to prevent overfitting. The BiLSTM is implemented using the "keras" package in R. Figure B.2 presents the basic structure of DFN, LSTM and BiLSTM.

[INSERT FIGURE B.2 ABOUT HERE]

2.2.4 Proposed hybrid models - BiLSTM with HAR-type models

Financial time series models such as GARCH-type models haven been integrated with different Neural Network models to predict financial volatility. Empirical results show that adding GARCH-type components significantly improve the volatility forecasts under the Neural Network structure, indicating those components are good feature representations of volatility. However, standard GARCH fails to reproduce the long-memory property which is an important stylized facts of financial returns. Extension of standard GARCH such as FIGARCH uses fractional difference operators to obtain

long-memory volatility but it is difficult to estimate and lacks explicit economic interpretation (Corsi (2009)). Such problems may negatively affect the interpretability of forecasting results generated by Neural Network models if those components are fed into Neural Networks as features.

HAR model is proposed by Corsi (2009) to address above issues of GARCH-type models and significantly improves volatility forecasting by successfully reproducing the main stylized facts of financial returns (i.e., self-similarity, long-memory and fat tails). This indicates that volatility is better represented under the HAR structure. It also maintains a simple structure and its parameters have clear economic interpretation. The standard HAR model is then extended by incorporating jump component in forecasting volatility. Corsi et al. (2010) propose a new jump test based on the Threshold Bipower variation and decompose the volatility into a continuous part and a jump part. Empirical results show that jumps have significant impact in forecasting future volatility, suggesting jumps serve to an important feature of financial volatility and volatility is better modelled by such decomposition.

As mentioned in the introduction, HAR model considers different volatility components realized over different time horizons and HAR-TCJ model looks into a more detailed decomposition of financial volatility. Their empirical results show significant improvement in forecasting volatility, indicating the characteristics of volatility are well featured by HAR-type components. Therefore, adding the HAR-type components into Neural Networks should improve the volatility forecasts compared to the Neural Networks without those input. To verify this hypothesis, we propose a hybrid model integrating HAR-type components into Neural Networks. We now explain in detail how information increases as HAR-type components are added.

In Eq.(2.3) of the standard HAR model, β_d measures the impact of volatility realized by actions of high-frequency traders (i.e., daily traders) on future volatility. β_w and β_m indicate the effect of volatility realized by behaviors of medium-frequency traders (i.e., weekly trades) and low-frequency traders (i.e., monthly traders) on future volatility respectively. By decomposing the volatility into a continuous part and jump part, the jump coefficient ($\beta_{j,d}$) of HAR-TCJ model in Eq.(2.4) provides extra information for the impact of jumps on future volatility. These parameters multiplied by the corresponding volatility component form the HAR-type components and they are more

informative than the past volatility alone. The HAR-type components are then added into Neural Networks as feature input. Neural networks transform these features into higher dimensional space in the hidden layer and learn the input-output map via the training data. DFN structure is applied to improve nonlinear pattern learning and BiLSTM is used to detect long-term dependence in the financial time-series data. Figure B.4 presents the architecture of the proposed hybrid model.

[INSERT FIGURE B.4 ABOUT HERE]

2.3 Data and experiment

In the empirical analysis, we choose one stock each from ten business sectors which gives us the price data of following ten stocks: Citigroup Inc. (C), Microsoft (MSFT), Freeport-McMoran (FCX), Pfizer (PFE), General Electric (GE), The Home Depot (HD), AT&T (T), Exxon-Mobil (XOM), Duke Energy (DUK) and Wal-Mart (WMT). We also include SPY which is the ETF of S&P 500 index to see results in more aggregate level. All data are obtained from Tick Data Inc. and the sampling frequency is 5-min. Our sample covers the period from Jan 03, 2000 to Dec 31, 2020. Descriptive statistics are provided in Table B.1. The experiment is to forecast RV over daily ($h = 1$) and weekly ($h = 5$) horizons. It is comprised of the following two steps.

Step 1: Feature construction and model building

- Data preparation: The daily, weekly and monthly RV in the standard HAR model are calculated using Eq. (2.1) - (2.2). The continuous part and jump part of RV in the HAR-TCJ model are computed by Eq.(2.5) - (2.7). Table B.1 provides the descriptive statistics of daily RV and daily \hat{TC} for each of the considered stocks.
- Data division: To examine the forecasting performance of different models under distinct market outlook, we divide our sample into three subsamples containing approximately equal observations: the pre-crisis sample (Jan 03, 2000 to Aug 31, 2007), the post-crisis sample (Sep 04, 2007 to April 30, 2014) and the last sample (May 01, 2014 to Dec 31, 2020). In addition, two thirds of each sample forms the training set and the remainder enters into the testing set ¹.

¹This split of training and testing example is conventional in the empirical literature and it is found

- Feature construction: To construct the HAR-type components, we estimate the parameters of the log-form HAR-type models using a rolling window of 1000 days². The estimated parameters are then multiplied by the corresponding volatility components which gives us the HAR-type components.
- Feature scaling: Given the huge difference between the value range of each feature input, we standardize all features to avoid big features dominating the features with smaller values.
- Model building: Single model and hybrid model are performed to forecast RV. Single model consists of HAR, HAR-TCJ, DFN and BiLSTM where Neural Networks only take past RV and returns as input. Hybrid model contains DFN-based and BiLSTM-based hybrid models where HAR-type components are integrated as input features of Neural Networks. Details are provided in Table B.2.

Step 2: Model training and testing sample forecasting

Although we apply the Neural Network models with single hidden layer in this chapter, it leaves us other tuning parameters to be determined before training (i.e., number of hidden nodes, dropout rate, batch size etc.). Tuning parameter selection is vital for the success of Neural Network models. For instance, increasing the number of hidden nodes improves the learning ability of Neural Networks characterized by the superior training sample performance. However, the results can be hardly generalized to the testing sample due to the problem of overfitting, leading to poor testing sample performance. We conduct the walk-forward validation to select the tuning parameters where the training sample is further divided into several small training and validation sample on a rolling window basis³. The optimal tuning parameters for single Neural Network models and hybrid models are selected separately according to the minimum average validation loss measured by Mean Squared Errors across all chosen stocks. The search space for each tuning parameter and the selected tuning parameter values are

that this split ratio subjects to less overfitting problem (Bucci (2020))

²We conduct the structural break test and find that 1000 days subject to at least one structural break for all chosen stocks.

³The selected rolling window length is 250 days. Each small training sample contains 500 observations and the subsequent 250 observations go into the small validation sample.

reported in Table B.3 - B.4.

The single and hybrid Neural Network models are then trained and optimized by gradient-based algorithm under the chosen tuning parameters. The testing sample performance is measured by Mean Squared Errors (MSE) and Diebold-Mariano (DM) test is performed to test the significance of performance difference of different models.

[INSERT TABLE B.2 ABOUT HERE]

[INSERT TABLES B.3 - B.4 ABOUT HERE]

2.4 Results

Table B.5 - B.7 present the out-of-sample forecasting performance of different models for the pre-crisis, post-crisis and last subsample respectively. We report the results by comparing and contrasting different models from a holistic perspective. We then look into the subsample results.

In general, at both the stock and SPY level, our proposed HAR-BiLSTM hybrid model consistently outperforms other benchmark models, achieving the minimum average forecasting loss across all forecasting horizons in all subsamples. HAR-DFN hybrid model ranks second in forecasting daily and weekly RV which also suggests the superiority of integrating Neural Networks with HAR-type models. DFN, as a short-memory Neural Network model, fails to build feedback loop between the input and hidden-layer output. BiLSTM, however, memorizes past input information for a long time and feed back (forward too as indicated by its name bidirectional) the current output through the functioning of input gate, forget gate and output gate, allowing long-memory modeling and therefore stronger pattern detection ability. In all subsamples, the DM test results show that on average the difference of forecasting performance between HAR-BiLSTM and HAR-DFN is not obvious in forecasting daily RV while becoming much more significant in forecasting weekly RV. This may suggest a more complex and longer-term dependent relationship between input variables and target output in longer forecasting horizon. Also, this reconfirms the empirical finding that long-memory Neural Network models have superior performance in capturing long-term dependence in the data (see e.g., [Kim and Won \(2018\)](#), [Bucci \(2020\)](#)). Moreover, the proposed hybrid models outperform the standard HAR and HAR-TCJ by a very

large extent in forecasting both daily and weekly RV. The performance difference is especially bigger in forecasting weekly RV. This indicates that strong nonlinearity exist in HAR-type components (daily, weekly and monthly RV, continuous components and jump component) which is beyond linearity modelling and requires nonlinear modeling algorithms.

Surprisingly, in all subsamples, single DFN and single BiLSTM, which takes past RV as input features, often underperform standard HAR and HAR-TCJ. However, such performance difference becomes much less in forecasting weekly RV. This might be attributed to the strong persistence of RV in short forecasting horizon. In other words, daily RV is highly persistent and such persistence is well modeled by the standard HAR. However, strong nonlinearity appears in longer forecasting horizon and Neural Network models become more powerful in capturing the nonlinear effects. In addition, superior forecasting performance of the proposed hybrid models indicates that HAR-type components are more informative input features than past RV. [Corsi \(2009\)](#) shows that the HAR model successfully reproduce the main stylized facts of RV such as self-similarity and long-memory. By inputting the HAR-type components into Neural Network basis, we actually provide: (1) main stylized facts of RV (self-similarity and long-memory); (2) components of RV (continuous part and jump part); (3) magnitude of volatility induced by different market participants (daily, weekly and monthly RV); (4) the impact of actions taken by high-frequency, medium-frequency and low-frequency traders on future RV (estimated parameters of HAR-type model). With more effective information, the performance of proposed hybrid models is significantly improved compared with single Neural Network models. This indicates that feature quality is a key factor which largely affect the performance of Neural Network models. This empirical finding may be meaningful in the following aspects: (1) components generated by classical Financial Time Series models may produce high-quality input features for Neural Network models and enhance their learning ability. This is achieved by successful and in-depth modeling of the statistical properties of target variables, offering Neural Network models a deeper and more micro vision; (2) Neural Network models, however, detect and model the nonlinearity that is beyond the scope of linear Financial Time Series models. This is particularly helpful when forecasting horizon is longer or large amount of exogenous variables are added.

On average, both HAR-DFN and HAR-BiLSTM perform better in the post-crisis period compared with other subsamples. This might be attributed to the unexpected events happened in the testing period of pre-crisis and last subsample. Specifically, the testing period of pre-crisis subsample is very close to the official beginning of the Great Recession in 2007 while that of last sample covers the burst of Covid-19 crisis in 2020. As those events are not included by the training samples, the abbreviate data behavior during the crisis periods may not be sufficiently learned by the hybrid models, leading to relatively higher loss in pre-crisis and last subsample. The testing period of post-crisis sample, however, is a relatively ordinary period compared with that of pre-crisis and the last subsample in which case the training sample performance is generalized well to the testing sample. Despite these, the proposed HAR-BiLSTM hybrid model still perform the best in all subsamples across all forecasting horizons, showing the superiority of long-memory Neural Network model.

[INSERT TABLES B.5 - B.7 ABOUT HERE]

2.5 Conclusion

This chapter proposes a new nonparametric Neural Network-based hybrid models to forecast realized volatility. Unlike semiparametric HAR-type hybrid models, where nonlinear transformation term is added to the linear part, we retain a full Neural Network setting which enables more powerful pattern detection and learning. Two representative Neural Network models, Deep Feedforward Neural Network (DFN) and Bi-directional Long Short-term Memory (BiLSTM) model, are used as Neural Network basis in this chapter. Compared with DFN, BiLSTM processes information both backwards and forwards which allows dynamic bidirectional learning.

By inputting HAR-type components, the proposed hybrid models consistently outperform other benchmark models in all subsamples across all forecasting horizons. The best performer is HAR-BiLSTM hybrid model, followed by the HAR-DFN. This suggests the superiority of long-term dependence detection model and the better learning ability of BiLSTM model. In addition, HAR-type components are stronger predictors of future RV than the past RV given the underperformance of single DFN and single BiLSTM. This finding may advocate the usage of HAR-type components

instead of raw RV or return data in the Neural Network models.

Moreover, on average, our proposed hybrid models achieve less forecasting loss in predicting weekly RV than predicting daily RV. This indicates that daily RV is highly persistent and such persistence is well modeled by the standard HAR. However, strong nonlinearity appears in longer forecasting horizon and Neural Network models become more powerful in capturing the nonlinear effects.

Appendix B

Chapter 2

B.1 Tables and Figures

Table B.1: Summary statistics

	MSFT	C	DUK	FCX	HD	PFE	T	WMT	XOM	GE	SPY
RV_t											
Mean	2.56	6.08	2.28	7.81	2.92	2.23	2.46	1.95	2.07	3.39	1.01
Std.	4.02	24.87	5.79	12.09	5.20	3.40	4.57	3.33	4.54	7.21	2.36
Skew.	6.06	19.31	15.66	6.18	7.69	7.28	9.42	7.06	12.43	8.98	9.79
Kurt.	59.27	567.11	375.27	59.84	97.08	88.96	194.58	87.41	250.98	133.38	146.79
$\hat{T}C_t$											
Mean	2.35	5.29	1.91	6.58	2.56	1.90	2.17	1.70	1.89	2.96	0.94
Std.	3.85	21.26	4.50	11.04	4.62	2.81	4.27	3.03	4.30	6.38	2.17
Skew.	6.28	23.62	11.91	7.03	8.01	7.06	10.48	7.79	13.65	10.11	10.31
Kurt.	64.13	896.82	222.39	76.74	108.07	88.51	243.52	110.79	302.83	179.15	173.07

Note: This table reports descriptive statistics of realized volatility and the continuous part of variation for SPY and other selected stocks. Descriptive statistics includes mean (Mean.), standard deviation (Std.), Skewness (Skew.) and Kurtosis (Kurt.).

Table B.2: Input variables of different models

Model	Return	Lagged RV	HAR Components	HAR-TCJ Components
Single SDN	Y	Y	N	N
Single BiLSTM	Y	Y	N	N
HAR-DFN	Y	N	Y	Y
HAR-BiLSTM	Y	N	Y	Y

Note: This table presents the input variables of single models and hybrid models applied in this study. Y indicates the corresponding variable is added to the model while N suggests the opposite.

Table B.3: Selected tuning parameter values for $h = 1$

Model	Parameters	Value	Search Space
Single DFN	Number of hidden nodes	4	[1, 2*Number of input features]; Increment = 1
	Hidden activation function	RELU	RELU, Tanh
	Batch size	20	[20, 32]; Increment = 1
	Dropout rate	0.2	[0.2, 0.5]; Increment = 0.1
	Number of epochs	Early callback	Maximum number of epochs is 150 with early callback to reduce processing time
Single BiLSTM	Number of hidden nodes	4	[1, 2*Number of input features]; Increment = 1
	Dropout rate	0.2	[0.2, 0.5]; Increment = 0.1
	Number of epochs	150	Maximum number of epochs is 150
	Merge mode	Multiplicative	Multiplicative, Average, Sum, Concatenate
HAR-DFN	Number of hidden nodes	8	[1, 2*Number of input features]; Increment = 1
	Hidden activation function	RELU	RELU, Tanh
	Batch size	26	[20, 32]; Increment = 1
	Dropout rate	0.2	[0.2, 0.5]; Increment = 0.1
	Number of epochs	Early callback	Maximum number of epochs is 150 with early callback to reduce processing time
HAR-BiLSTM	Number of hidden nodes	8	[1, 2*Number of input features]; Increment = 1
	Dropout rate	0.3	[0.2, 0.5]; Increment = 0.1
	Number of epochs	150	Maximum number of epochs is 150
	Merge mode	Multiplicative	Multiplicative, Average, Sum, Concatenate

Note: This table reports the tuning parameter values used in different models for $h = 1$. These parameters are selected according to the minimum average validation loss. The search space for each parameter is determined by empirical research convention.

Table B.4: Selected tuning parameter values for $h = 5$

Model	Parameters	Value	Search Space
Single DFN	Number of hidden nodes	8	[1, 2*Number of input features]; Increment = 1
	Hidden activation function	RELU	RELU, Tanh
	Batch size	32	[20, 32]; Increment = 1
	Dropout rate	0.3	[0.2, 0.5]; Increment = 0.1
	Number of epochs	Early callback	Maximum number of epochs is 150 with early callback to reduce processing time
Single BiLSTM	Number of hidden nodes	8	[1, 2*Number of input features]; Increment = 1
	Dropout rate	0.3	[0.2, 0.5]; Increment = 0.1
	Number of epochs	150	Maximum number of epochs is 150
	Merge mode	Multiplicative	Multiplicative, Average, Sum, Concatenate
HAR-DFN	Number of hidden nodes	16	[1, 2*Number of input features]; Increment = 1
	Hidden activation function	RELU	RELU, Tanh
	Batch size	32	[20, 32]; Increment = 1
	Dropout rate	0.3	[0.2, 0.5]; Increment = 0.1
	Number of epochs	Early callback	Maximum number of epochs is 150 with early callback to reduce processing time
HAR-BiLSTM	Number of hidden nodes	8	[1, 2*Number of input features]; Increment = 1
	Dropout rate	0.3	[0.2, 0.5]; Increment = 0.1
	Number of epochs	150	Maximum number of epochs is 150
	Merge mode	Multiplicative	Multiplicative, Average, Sum, Concatenate

Note: This table reports the tuning parameter values used in different models for $h = 5$. These parameters are selected according to the minimum average validation loss. The search space for each parameter is determined by empirical research convention.

Table B.5: Out-of-sample forecast losses: pre-crisis sample

	MSFT	C	DUK	FCX	HD	PFE	T	WMT	XOM	GE	Average	SPY
Horizon = 1												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
HAR-TCJ	1.007	0.993	0.978	0.990	1.007	0.983	0.996	0.991	0.996	0.980	0.992	1.004
DFN	1.706 ***	1.361 ***	1.134 ***	1.158 ***	1.255 ***	1.609 ***	1.705 ***	1.198 ***	1.270 ***	1.514 ***	1.391	1.280 ***
BiLSTM	1.542 ***	1.589 ***	1.331 ***	1.248 ***	1.212 ***	1.081 **	1.797 ***	1.124 ***	1.664 ***	1.498 ***	1.309	1.483 ***
HAR-DFN	0.967 *	0.968 *	0.970 *	0.993	1.005	0.979	1.071	0.998	1.001	0.903 **	0.986	0.881 ***
HAR-BiLSTM	0.963 *	0.875 ***	0.609 ***	0.954 **	0.955 **	0.858 ***	1.017	0.973	1.003	0.819 ***	0.903	0.879 ***
Horizon = 5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
HAR-TCJ	0.996	0.996	0.995	0.986	0.983	0.985	0.988	0.996	1.009	1.008	0.994	1.010
DFN	1.041 ***	1.185 ***	1.187 ***	1.013 ***	1.006	1.038 ***	1.136 ***	0.995	1.047 ***	1.020 ***	1.067	1.038 ***
BiLSTM	1.027 ***	1.045 ***	1.084 ***	1.008	1.000	1.001	1.119 ***	0.991	1.017 ***	1.001	1.029	1.019 ***
HAR-DFN	0.936 **	0.945 **	0.925 **	1.000	0.977	0.993	0.951 *	0.967 *	0.996	0.922 **	0.961	0.679 ***
HAR-BiLSTM	0.886 ***	0.887 ***	0.910 **	0.929 **	0.929 **	0.990	0.813 ***	0.959 *	0.849 ***	0.888 ***	0.904	0.690 ***

Note: The table reports the ratio between the MSE of different models and the MSE of the standard HAR model for the pre-crisis sample. Figures lower than one indicates the corresponding model achieves lower MSE than the standard HAR while figures higher than one suggests the opposite. Column Average reports the average MSE of the selected 10 individual stocks. The lowest loss ratio in each column of the Average and the SPY is highlighted in bold. *, ** and *** indicate that the relative forecasting differences between the corresponding model and the standard HAR are significant at 10%, 5% and 1% level using the Diebold–Mariano test (Newey–West heteroscedasticity consistent covariance matrix estimator). Pre-crisis sample covers the period from Jan 03, 2000 to Aug 31, 2007 with 1927 observations in total. Two thirds of the pre-crisis sample are the training sample and the remainder forms the testing sample.

Table B.6: Out-of-sample forecast losses: post-crisis sample

	MSFT	C	DUK	FCX	HD	PFE	T	WMT	XOM	GE	Average	SPY
Horizon = 1												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
HAR-TCJ	1.187 ***	1.145 ***	0.952 **	1.032 ***	1.481 ***	1.028 ***	1.064 ***	0.948 **	1.042 ***	1.058 ***	1.094	1.026 ***
DFN	1.023 **	1.020 **	1.199 ***	1.759 ***	1.089 ***	1.216 ***	1.063 ***	1.245 ***	1.108 ***	1.001	1.172	1.011
BiLSTM	1.165 ***	1.138 ***	1.233 ***	1.131 ***	1.029 ***	1.065 ***	1.118 ***	1.224 ***	1.131 ***	1.033 ***	1.127	1.025 ***
HAR-DFN	0.987	0.954 **	0.949 **	0.939 **	0.876 ***	0.897 ***	0.811 ***	0.972	0.998	0.741 ***	0.912	0.956 **
HAR-BiLSTM	0.979	0.896 ***	0.917 **	0.967 *	0.790 ***	0.817 ***	0.792 ***	0.825 ***	0.924 **	0.720 ***	0.863	0.899 ***
Horizon = 5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
HAR-TCJ	0.964 *	0.973	0.979	1.002	1.052 ***	0.984	1.011	0.988	0.813 ***	1.020 **	0.979	1.003
DFN	0.987	1.044 ***	1.055 ***	1.094 ***	1.052 ***	1.020 **	1.092 ***	1.041 ***	1.066 ***	1.085 ***	1.054	1.102 ***
BiLSTM	1.089 ***	0.996	1.050 ***	0.994	1.049 ***	0.982	1.023 **	1.019	1.038 ***	1.020 **	1.026	1.099 ***
HAR-DFN	0.993	0.825 ***	0.707 ***	0.886 ***	0.801 ***	0.743 ***	0.886 ***	0.934 **	0.995	0.706 ***	0.848	0.948 **
HAR-BiLSTM	0.960 *	0.768 ***	0.685 ***	0.713 ***	0.794 ***	0.636 ***	0.743 ***	0.652 ***	0.890 ***	0.704 ***	0.755	0.811 ***

Note: The table reports the ratio between the MSE of different models and the MSE of the standard HAR model for the post-crisis sample. Figures lower than one indicates the corresponding model achieves lower MSE than the standard HAR while figures higher than one suggests the opposite. Column Average reports the average MSE of the selected 10 individual stocks. The lowest loss ratio in each column of the Average and the SPY is highlighted in bold. *, ** and *** indicate that the relative forecasting differences between the corresponding model and the standard HAR are significant at 10%, 5% and 1% level using the Diebold–Mariano test (Newey–West heteroscedasticity consistent covariance matrix estimator). Post-crisis sample covers the period from Sep 04, 2007 to April 30, 2014 with 1676 observations in total. Two thirds of the post-crisis sample are the training sample and the remainder forms the testing sample.

Table B.7: Out-of-sample forecast losses: last sample

	MSFT	C	DUK	FCX	HD	PFE	T	WMT	XOM	GE	Average	SPY
Horizon = 1												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
HAR-TCJ	1.027 ***	1.025 ***	0.988	0.994	0.982	1.008	0.956 **	0.996	0.974	1.012	0.996	0.964
DFN	0.975	1.105 ***	1.021 ***	1.079 ***	1.302 ***	1.143 ***	0.950 **	0.992	1.072 ***	1.080 ***	1.072	1.126 ***
BiLSTM	1.037 ***	1.006	1.166 ***	1.003	1.225 ***	1.130 ***	0.944 **	0.999	1.011 ***	0.985	1.051	1.216 ***
HAR-DFN	0.960 *	0.932 **	0.987	0.974	0.989	0.985	0.919 **	0.942 **	0.965	0.922 **	0.958	0.929 **
HAR-BiLSTM	0.948 **	0.844 ***	0.953 **	0.927 **	0.937 **	0.975	0.910 **	0.936 **	0.957 **	0.888 ***	0.928	0.910 **
Horizon = 5												
HAR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
HAR-TCJ	0.989	0.999	1.000	1.001	0.989	1.002	1.000	1.008	1.000	0.987	0.998	1.005
DFN	1.019 ***	1.980 ***	1.021 ***	1.072 ***	1.125 ***	1.045 ***	1.010	1.014	1.082 ***	1.030 ***	1.140	0.999
BiLSTM	0.988	1.227 ***	1.040 ***	1.070 ***	1.008	1.009	1.002	0.956 **	1.047 ***	1.027 ***	1.037	1.023 ***
HAR-DFN	0.889 ***	0.856 ***	0.995	0.883 ***	0.725 ***	0.815 ***	0.894 ***	0.852 ***	0.982	0.877 ***	0.877	0.874 ***
HAR-BiLSTM	0.851 ***	0.841 ***	0.854 ***	0.864 ***	0.705 ***	0.803 ***	0.799 ***	0.743 ***	0.706 ***	0.734 ***	0.790	0.760 ***

Note: The table reports the ratio between the MSE of different models and the MSE of the standard HAR model for the last sample. Figures lower than one indicates the corresponding model achieves lower MSE than the standard HAR while figures higher than one suggests the opposite. Column Average reports the average MSE of the selected 10 individual stocks. The lowest loss ratio in each column of the Average and the SPY is highlighted in bold. *, ** and *** indicate that the relative forecasting differences between the corresponding model and the standard HAR are significant at 10%, 5% and 1% level using the Diebold–Mariano test (Newey–West heteroscedasticity consistent covariance matrix estimator). The last sample covers the period from May 01, 2014 to Dec 31, 2020 with 1681 observations in total. Two thirds of the last sample are the training sample and the remainder forms the testing sample.

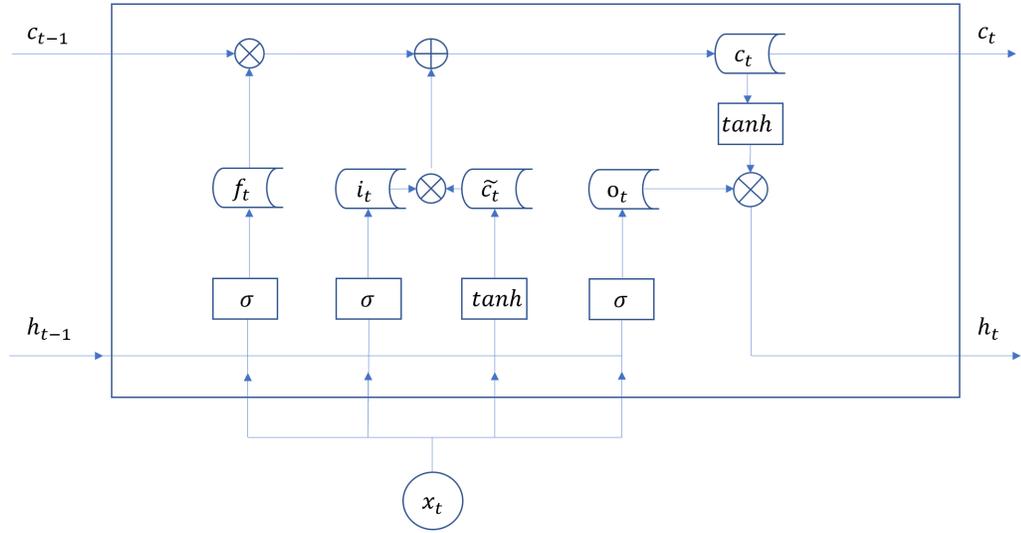
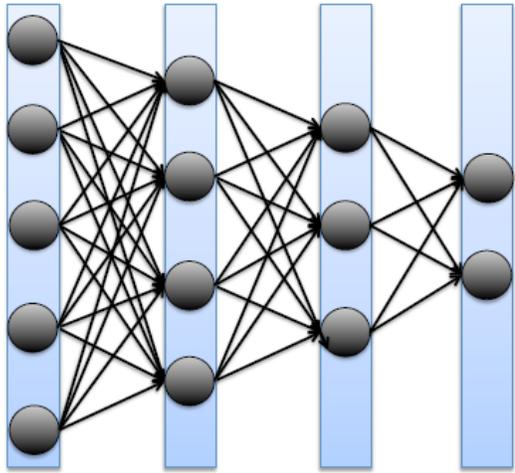
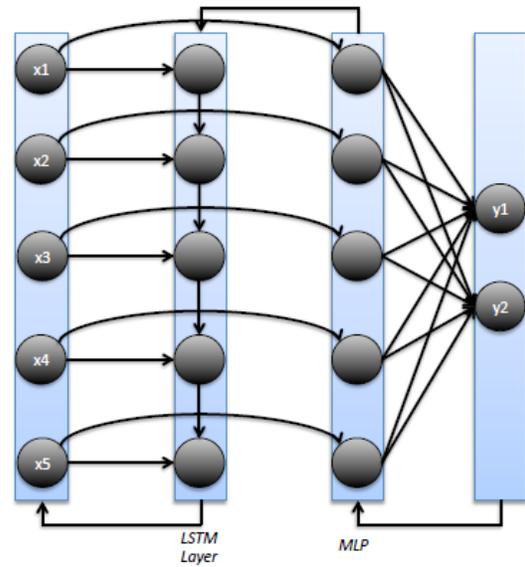


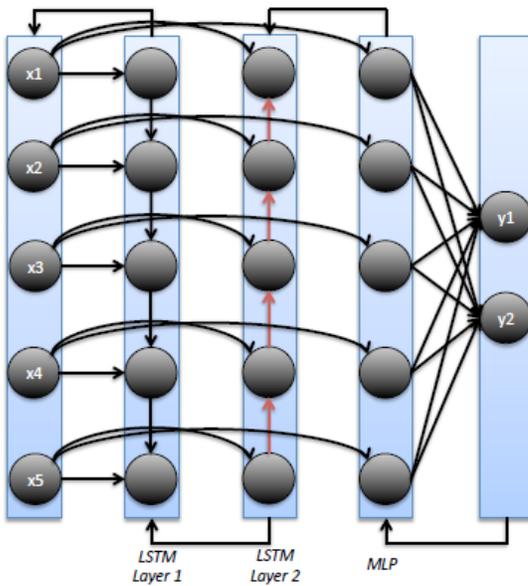
Figure B.1: LSTM memory cell



(a) Feedforward Neural Network



(b) Long Short-term Memory Network



(c) Bidirectional Long Short-term Memory Network

Figure B.2: Structure of Neural Network models

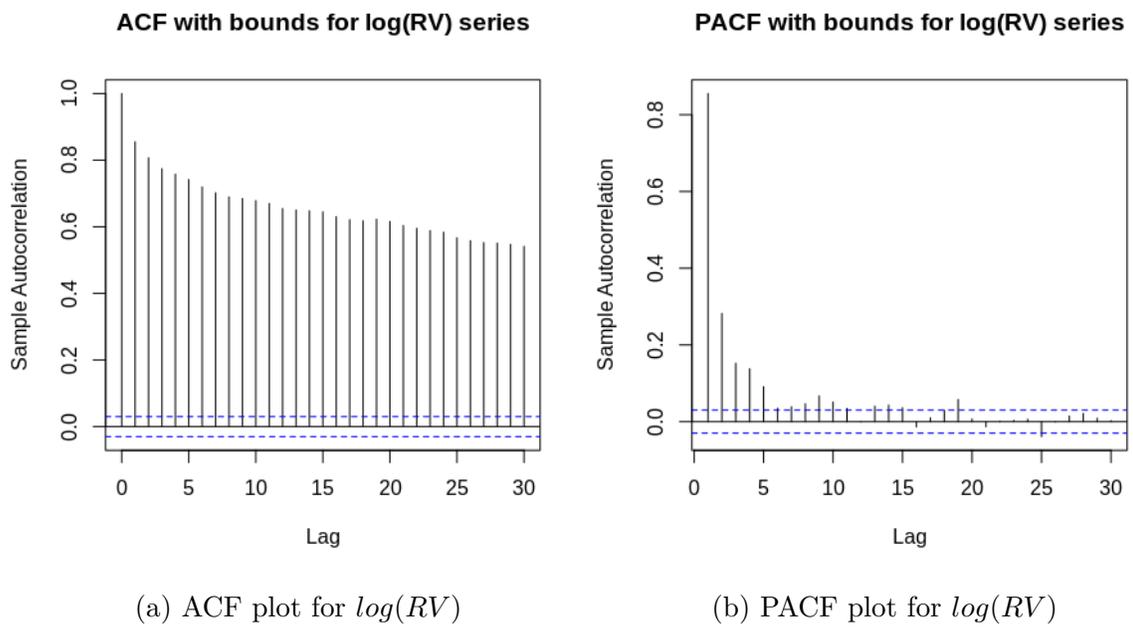


Figure B.3: ACF and PACF plot of SPY's RV

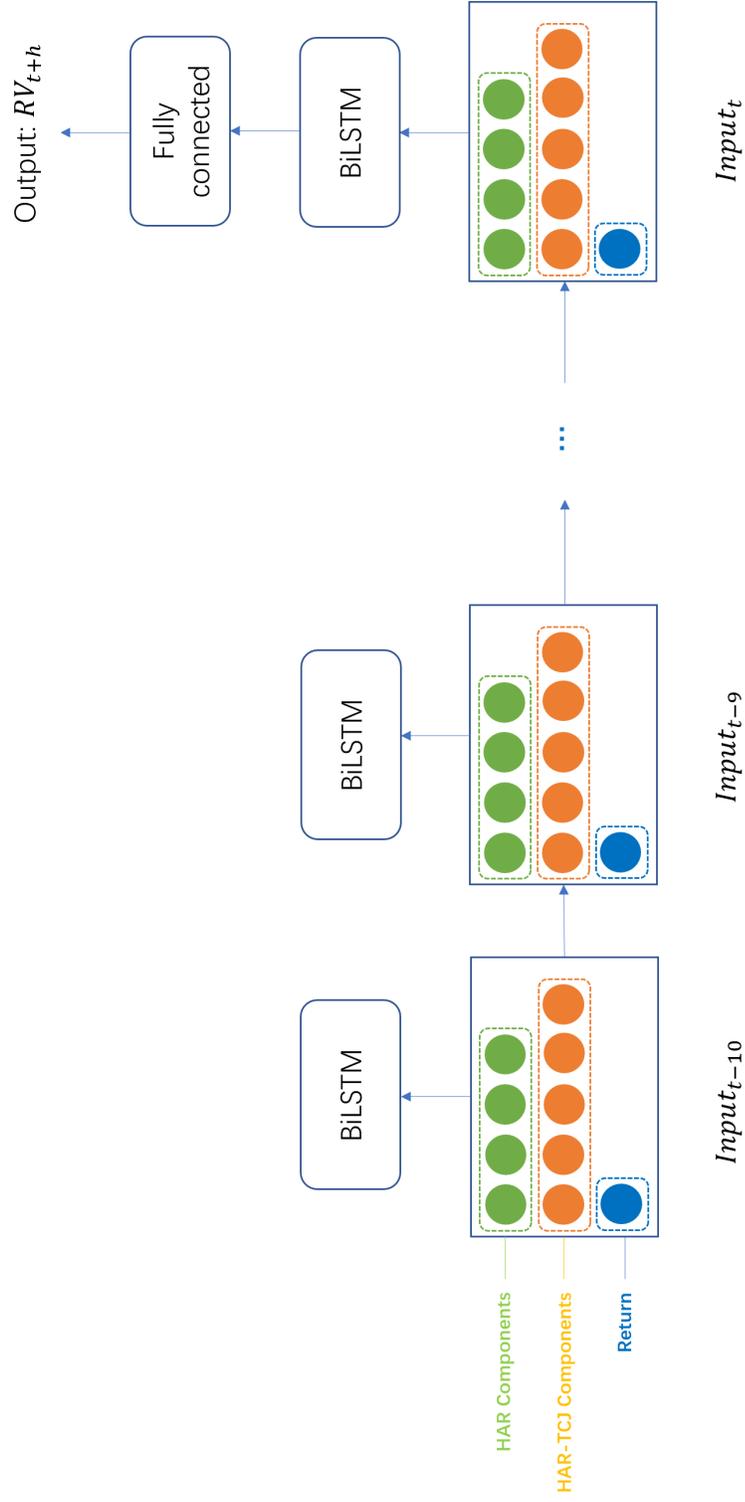


Figure B.4: Structure of the proposed HAR-BiLSTM hybrid model

Note: This figure shows the structure of the proposed HAR-BiLSTM model. To predict the target output RV_{t+h} , HAR-type components are integrated into BiLSTM network as input variables from $t-10$ to t . A fully connected dense layer is added to the BiLSTM network at time t to project results into correct dimension.

Chapter 3

When MIDAS Meets LASSO: Forecasting Tail Risk Using Effective Macroeconomic Variables

3.1 Introduction

In the context of recent credit and financial crises, appropriate risk measures raise greater interest for banks and other financial institutions. Risk measures have become an essential tool in supporting asset management decisions for banks and other financial institutions especially under market turmoil. Value at Risk (VaR) and Expected Shortfall (ES) are two prevailing risk measures that currently dominate financial regulatory framework. VaR is the maximum loss that would occur within a certain time period, for a predefined confidence level. It can be defined mathematically as:

$$VaR_t^\alpha \equiv \inf\{y_t \in \mathbb{R} | F_Y(y_t | \mathcal{F}_{t-1}) \geq \alpha\},$$

where $F_Y(\cdot | \mathcal{F}_{t-1})$ is the cumulative distribution function of asset returns y_t over a horizon given the information set \mathcal{F}_{t-1} , and $\alpha \in (0, 1)$ is a given significance level. As a quantile, VaR can be expressed directly in terms of the inverse cumulative distribution function: $VaR_t^\alpha = F_Y^{-1}(\alpha | \mathcal{F}_{t-1})$. As a risk measure, its conceptual simplicity and computational ease made VaR very popular among financial practitioners.

Despite its favorable properties, VaR has inherent drawbacks. First, VaR ignores the shape and structure of the tail; i.e., it does not tell us the loss level that would occur with a probability lower than the selected confidence level. Second, VaR is not a “coherent” risk measure and hence it does not consider the benefits of diversification (Artzner et al. (1999)). As a result, investors and risk managers may encounter greater loss which is beyond the VaR level Yamai and Yoshida (2005). After the financial crisis of 2007-2008, Basel Committee on Banking Supervision (2019) propose to use the ES risk measure, which is defined as the conditional expectation of returns that are below VaR. Unlike VaR, ES is a coherent risk measure and more informative by considering the tail shape of the loss distribution. Therefore, ES has been used as an alternative risk measure, complementing the VaR measure. It is defined as:

$$ES_t^\alpha \equiv \mathbb{E}[y_t | y_t \leq VaR_t^\alpha, \mathcal{F}_{t-1}].$$

However, little work has been done in modeling ES. This is partly because ES is not “elicitable” which means that no loss function exists for which ES is the solution that minimizes the loss.¹ Many studies contribute to the problem of elicibility (see, e.g., Engle and Manganelli (2004), Taylor (2008), Zhu and Galbraith (2011), Fissler and Ziegel (2016), Du and Escanciano (2017) and Patton et al. (2019)). Until the work of Fissler and Ziegel (2016), ES is found to be jointly elicitable with VaR and a set of suitable scoring functions are proposed.

With the development of methodologies, forecasting efficiency of risk measures can be significantly improved by incorporating information from the intraday data or high-frequency data into parametric models (Giot and Laurent (2004), Hansen et al. (2012), Louzis et al. (2014)) and semiparametric models (Clements et al. (2008), Meng and Taylor (2020), Lazar and Xue (2020), Gerlach and C. Wang (2020)). Specifically, realized volatility, proposed by Andersen and Bollerslev (1998) and Alizadeh et al. (2002), is a near unbiased return variation measure and is one of the most widely exploited intraday variables given its close relationship with risk measures. The commonly used realized volatility is selected with 5-minute or 10-minute frequency.

¹A risk measure is elicitable if the correct forecast of the measure is the unique minimizer of the expectation of at least one scoring function (Fissler and Ziegel (2016)). Such scoring functions are called strictly consistent for the risk measure.

Adding realized volatility is found to outperform earlier models in forecasting VaR and ES. However, these studies only exploit the effectiveness of high-frequency variables in predicting future tail risk and completely disregard the important role of macroeconomic information which is usually recorded at a lower frequency than tail risk (e.g., quarterly). A few attempting work have found some evidence supporting the value of low-frequency variables in tail risk forecasting (see, e.g., [Massacci \(2017\)](#), [Candila et al. \(2020\)](#), [Le \(2020\)](#), [Xu et al. \(2021\)](#)). However, three challenges of incorporating macroeconomic information into the tail risk modeling remain to be solved: (1) the common data frequency disalignment between macroeconomic variables and risk measures ([Ghysels et al. \(2004\)](#), [Engle, Ghysels, et al. \(2013\)](#), [Candila et al. \(2020\)](#)); (2) the need of a framework than can accommodate large amount of variables while utilizing only the most important variables among abundant candidates ([Gu et al. \(2020\)](#), [L. Chen et al. \(2020\)](#)); (3) the need of a suitable loss function that can accommodate variable selection and parameter estimation in the proposed model ([Patton et al. \(2019\)](#), [Taylor \(2019\)](#)). In view of this, it is necessary to calibrate a model with clear loss function that integrates different but effective information when modeling the tail dynamics.

We address the first challenge through the channel of volatility and we propose a new quantile-based model with an effective variable screening procedure to overcome the rest challenges. Stock market volatility is an important driving factor of tail risk and is found to be closely related to low-frequency macroeconomic and financial indicators through the Mixed Data Sampling (MIDAS) regression model proposed by [Ghysels et al. \(2004\)](#). Based on their work, many extended studies have shown that the dynamic of return volatility is characterized by multiple components capturing information at different time horizons. For example, [Engle, Ghysels, et al. \(2013\)](#) propose the GARCH-MIDAS model to extract two components of volatility, the short-term component that follows a GARCH(1,1) process and the long-term component that relates to macroeconomic variables. Their study finds significant macro-volatility relationship by directly incorporating low-frequency macroeconomic variables into the long-term volatility component. Their findings have been strengthened by many other studies and the GARCH-MIDAS model has become the most popular model in investigating the relationships between aggregate financial volatility and macroeconomic & financial variables (see, e.g., [Asgharian et al. \(2013\)](#), [Conrad, Loch, and Daniel \(2014\)](#), [Conrad](#)

and Loch (2015), Pan et al. (2017), Su et al. (2017), Conrad and Kleen (2019)). Following Engle, Ghysels, et al. (2013), we decompose the return volatility into a short-term and a long-term component where an extensive set of macroeconomic & financial indicators are embedded. However, we extend the GARCH-MIDAS model to allow the direct forecast of tail risk.

A new quantile-based model is therefore proposed for jointly estimating VaR and ES, characterizing the relationships between risk measures and macroeconomic & financial indicators. Quantile-based modeling avoids distributional assumptions on returns and allows the dynamics of the quantiles to vary for each probability level Engle and Manganeli (2004). This approach has been focused on VaR and produces superior VaR forecasts (see, e.g., Şener et al. (2012)). Taylor (2019) extends this approach and provides an apparent way of producing ES forecasts in a quantile setting. By assuming the AL density on the pair of VaR and ES, he shows that the negative of AL likelihood function belongs to the function set proposed by Fissler and Ziegel (2016) and it is therefore strictly consistent for the joint evaluation of VaR and ES. Following Taylor (2019), we assume the AL density on VaR and ES to obtain a strictly consistent scoring function. We differ from Taylor’s approach by integrating a large set of low-frequency variables through the channel of volatility, allowing the usage of new information to support the tail risk forecasting.

However, with a large number of variables being considered, the number of estimated parameters also increases, which increases the model complexity and reduces estimation efficiency. Fang et al. (2020) apply Adaptive-Lasso proposed by Zou (2006) to select variables that display the strongest signal in predicting long-term volatility component under the GARCH-MIDAS framework. They find that the GARCH-MIDAS model with variable selection outperforms all benchmark models except the model with realized volatility. The limitation of their approach is the selected variables do not vary for every out-of-sample estimation and it is unreasonable to assume those variables have the same contribution in each rolling window of the testing sample. To address this problem, we propose an innovative dynamic variable selection process where the variables are selected on a rolling-window basis. It helps us to determine the most important variables and allows us to visualize the evolution of supporting variables in predicting VaR and ES.

The main contribution of this chapter is to address two main research gaps in tail risk forecasting:

(1) We propose the ES-CAViaR-MIDAS model to integrate low-frequency variables into the high-frequency tail risk forecasting, allowing the usage of richer information. As stated above, the lack of a suitable model that deals with frequency disalignment makes it difficult to utilize low-frequency information such as macroeconomic indicators in forecasting tail risk. The proposed model addresses this problem through the channel of volatility and retains the semiparametric setting to avoid distributional assumptions on returns.

(2) We embed a dynamic variable selection device to select and utilize the most important low-frequency variables in forecasting future tail risk. Instead of putting all candidate variables into the model or subjectively picking variables, we let the data speak for itself by selecting the most informative variables that maximize the AL likelihood function with Adaptive-Lasso penalty. In addition, the selection process is on a quarterly rolling basis, allowing the impact of low-frequency variables on the tail risk to be visualized. The negative relationship between the selected macroeconomic variables and tail risk provides novel evidence of countercyclical tail risk. The empirical results show that our proposed model with the selected variables outperforms all benchmark models especially the model with only realized volatility in forecasting future tail risk at all chosen confidence levels.

The remainder of this chapter is structured as follows. Section 2 briefly introduces the proposed model and the procedure of variable selection. Section 3 presents the data used in our empirical study. Section 4 shows the dynamic variable selection results and out-of-sample performance. Finally, Section 5 concludes.

3.2 Models

3.2.1 The ES-CAViaR-MIDAS model

Let $r_{i,t}$ be an asset's return on the i th day of the t th quarter, where $i = 1, \dots, N_t$, $t = 1, \dots, T$. The setting allows different numbers of trading days per quarter. Following

Engle, Ghysels, et al. (2013), we specify the model of $r_{i,t}$ as follows:

$$r_{i,t} = \mu + \sqrt{\tau_t g_{i,t}} \epsilon_{i,t}, \quad (3.1)$$

where $\epsilon_{i,t}$ is identically and independently distributed (i.i.d.) and follows an unknown distribution $F(\epsilon_{i,t})$ with zero mean and unit variance. μ denotes the mean of asset return. τ_t denotes the long-term volatility component in quarter t , and $g_{i,t}$ denotes the short-term volatility component on the i th day in the t th quarter.

In the GARCH-MIDAS model proposed by Engle, Ghysels, et al. (2013), the short-term volatility component follows a GARCH(1,1) process:

$$g_{i,t} = (1 - \beta_1 - \beta_2) + \beta_1 \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta_2 g_{i-1,t}, \quad (3.2)$$

The long-term volatility component with single low-frequency variable is given by:

$$\log(\tau_t) = m + \theta \sum_{k=1}^K \psi_k(w_1, w_2) X_{t-k}, \quad (3.3)$$

where X_t is the exogenous variable in quarter t , K is the lagged order, and $\log(\tau_t)$ is considered rather than τ_t in order to ensure non-negative long-term volatility. Here, $\psi_k(w_1, w_2)$ is the Beta Weighting scheme, which is defined as:

$$\psi_k(w_1, w_2) = \frac{(k/(K+1))^{w_1-1} \cdot (1 - k/(K+1))^{w_2-1}}{\sum_{s=1}^K (s/(K+1))^{w_1-1} \cdot (1 - s/(K+1))^{w_2-1}}, \quad (3.4)$$

where $\psi_k(w_1, w_2)$ measures the weight for the k th lag of explanatory variable X_{t-k} and it is determined by two parameters w_1 and w_2 . For $k = 1, \dots, K$, we have $\psi_k \geq 0$ and $\sum_{k=1}^K \psi_k = 1$.

Given a probability level α for quantile, we characterize the quantile expression of (3.1) as:

$$\begin{aligned} v_{i,t}(r_{i,t}; \alpha | \Psi_{t-1}) - \mu &= a \sqrt{\tau_t g_{i,t}}, \text{ where } a = F_\alpha^{-1}(\epsilon_{i,t}), \\ e_{i,t}(r_{i,t}; \alpha | \Psi_{t-1}) - \mu &= b \sqrt{\tau_t g_{i,t}}, \text{ where } b = \mathbb{E}[\epsilon_{i,t} | \epsilon_{i,t} \leq a], \end{aligned} \quad (3.5)$$

where Ψ_{t-1} is the information set at time $t-1$, $F_\alpha^{-1}(\epsilon_{i,t})$ is the quantile of $\epsilon_{i,t}$ at level α . $v_{i,t}(r_{i,t}; \alpha | \Psi_{t-1})$ and $e_{i,t}(r_{i,t}; \alpha | \Psi_{t-1})$ denote VaR and ES based on the information

up to day $i - 1$ of quarter t at the α confidence level respectively. We use $v_{i,t}$ and $e_{i,t}$ from now for notation convenience.

Unlike Engle, Ghysels, et al. (2013), we consider the Threshold GARCH (1,1) process proposed by Zakoian (1994) for the short-term volatility component to better account for the well-known leverage effect. We now have:

$$g_{i,t}^{\frac{1}{2}} = \beta_0 + \beta_1 \frac{(r_{i-1,t} - \mu)^+}{\tau_t^{\frac{1}{2}}} + \beta_2 \frac{(r_{i-1,t} - \mu)^-}{\tau_t^{\frac{1}{2}}} + \beta_3 g_{i-1,t}^{\frac{1}{2}}, \quad (3.6)$$

where $(r_{i-1,t} - \mu)^+ = \max\{r_{i-1,t} - \mu, 0\}$ and $(r_{i-1,t} - \mu)^- = \min\{r_{i-1,t} - \mu, 0\}$.

$\tau_t^{\frac{1}{2}}$ is the long-term component of the volatility in the t th month, whose specification is shown as:

$$\tau_t^{\frac{1}{2}} = \exp\left(m + \theta \sum_{k=1}^K \psi_k(w_1, w_2) X_{t-k}\right). \quad (3.7)$$

After combining Eq.(3.4) - Eq.(3.7), we now propose a model for the joint forecasting of VaR and ES with a single low-frequency variable:

$$\begin{cases} v_{i,t} - \mu = a\sqrt{\tau_t} g_{i,t} \\ e_{i,t} - \mu = b\sqrt{\tau_t} g_{i,t} \\ g_{i,t}^{\frac{1}{2}} = \eta_0 + \beta_1 \frac{(r_{i-1,t} - \mu)^+}{\tau_t^{\frac{1}{2}}} + \beta_2 \frac{(r_{i-1,t} - \mu)^-}{\tau_t^{\frac{1}{2}}} + \beta_3 g_{i-1,t}^{\frac{1}{2}} \\ \tau_t^{\frac{1}{2}} = \exp\left(m + \theta \sum_{k=1}^K \psi_k(w_1, w_2) X_{t-k}\right) \\ \psi_k(w_1, w_2) = \frac{(k/(K+1))^{w_1-1} \cdot (1-k/(K+1))^{w_2-1}}{\sum_{s=1}^K (s/(K+1))^{w_1-1} \cdot (1-s/(K+1))^{w_2-1}}. \end{cases} \quad (3.8)$$

Substituting Eq.(3.6) and Eq.(3.7) into Eq.(3.5), the proposed model in Eq.(3.8) can be written as:

$$\begin{cases} v_{i,t} - \mu = \beta_0 + \beta_1 (r_{i-1,t} - \mu)^+ + \beta_2 (r_{i-1,t} - \mu)^- + \beta_3 (v_{i-1,t} - \mu) \\ e_{i,t} - \mu = \gamma_0 (v_{i,t} - \mu), \end{cases} \quad (3.9)$$

where $\beta_0 = a\eta_0 \exp\left(m + \theta \sum_{k=1}^K \psi_k(w_1, w_2) X_{t-k}\right)$ and $\gamma_0 = b/a$. The model in Eq.(3.9) has a similar framework with the CAViaR Asymmetric Slope (AS) model with ES

component proposed by [Taylor \(2019\)](#). Thus, we name Eq.(3.9) as the ‘‘ES-CAViaR-MIDAS’’ model.²

Although the parameters of the proposed ES-CAViaR-MIDAS model can be obtained by the maximum likelihood estimation with an assumed distribution for innovations $\epsilon_{i,t}$, it is unrealistic to assume a consistent distribution for innovations. To avoid this, we follow [Taylor \(2019\)](#) to assume the AL density on the pair of (VaR, ES) rather than on the innovations. Our proposed model falls into a semiparametric framework where parameters are estimated by maximizing the AL likelihood function.³

A few studies that investigate the relationships between the stock market tail risk and low-frequency macroeconomic variables focus on the effect of single macroeconomic variable on tail risk, while some variables may have joint effects ([Candila et al. \(2020\)](#), [Le \(2020\)](#), [Xu et al. \(2021\)](#)). Therefore, a clear and feasible framework is needed to integrate large number of low-frequency variables into the tail risk forecasting. In our study, we generalize the model Eq.(3.9) by considering an extensive set of low-frequency variables:

$$\begin{cases} v_{i,t} - \mu = \beta_0 + \beta_1(r_{i-1,t} - \mu)^+ + \beta_2(r_{i-1,t} - \mu)^- + \beta_3(v_{i-1,t} - \mu) \\ e_{i,t} - \mu = \gamma_0(v_{i,t} - \mu) \\ \beta_0 = \exp \left(m + \sum_{j=1}^J \theta_j \sum_{k=1}^K \psi_k(w_{j,1}, w_{j,2}) X_{j,t-k} \right), \end{cases} \quad (3.10)$$

where J denotes the number of explanatory variables ($J = 20$ in this study), and θ_j measures the impact of the j th explanatory variable on the long-term volatility.

The proposed model is estimated by maximizing the AL likelihood function:

$$LLF(\Phi) = \sum_{t=1}^T \sum_{i=1}^{N_t} \left[\log \left(\frac{\alpha - 1}{e_{i,t}} \right) + \frac{(r_{i,t} - v_{i,t})(\alpha - \mathbf{1}\{r_{i,t} \leq v_{i,t}\})}{\alpha e_{i,t}} \right]. \quad (3.11)$$

²This model can be extended with alternative CAViaR frameworks for VaR forecasting, which are introduced by [Engle and Manganelli \(2004\)](#) and [Taylor \(2019\)](#), including the Symmetric Absolute Value (SAV), Adaptive and Indirect GARCH(1, 1). In this chapter, we only consider the CAViaR-AS model.

³[Patton et al. \(2019\)](#) propose a set of semiparametric models for VaR and ES jointly, in which parameters are estimated by minimizing the FZ loss function by [Fissler and Ziegel \(2016\)](#). [Taylor \(2019\)](#) shows that the result by maximizing the AL density function is consistent with the one by minimizing the FZ loss function with the assumption of zero-mean returns.

3.2.2 ES-CAViaR-MIDAS with variable selection

As the number of parameters rises with the number variables included, difficulties may arise in extracting those with the strongest signals from those who are less important. We therefore conduct variable screening by incorporating the Adaptive-Lasso penalty proposed by [Zou \(2006\)](#). The penalized likelihood function is:

$$PLLF_\lambda(\Phi) = \sum_{t=1}^T \sum_{i=1}^{N_t} \left[\log \left(\frac{\alpha - 1}{e_{i,t}} \right) + \frac{(r_{i,t} - v_{i,t})(\alpha - \mathbf{1}\{r_{i,t} \leq v_{i,t}\})}{\alpha e_{i,t}} \right] - \lambda \sum_{j=1}^J \hat{w}_j |\theta_j|, \quad (3.12)$$

where $\lambda > 0$ is the tuning parameter regulating shrinkage power in the Adaptive-Lasso penalty, $PLLF_\lambda(\Phi)$ denotes the penalized likelihood function for a given λ , and \hat{w}_j is the adaptive weight for each explanatory variable X_j .

To obtain the adaptive weights, which is calculated as $\hat{w}_j = \frac{1}{|\hat{\theta}_j|^\eta}$, we estimate the ES-CAViaR-MIDAS model for all J variables to obtain the estimates $\hat{\theta}_j$. Following [Zou \(2006\)](#), we set $\eta = 2$ to achieve higher probability of obtaining the true model.

3.2.3 Tuning parameter selection

Choosing the value of the tuning parameter λ in the penalized likelihood function is vital for the success of variable screening process. A small λ increases the number of selected variables while taking the risk of including noisy variables that are far from the underlying true model. On the contrary, a large enough λ rules out all variables although some of them display strong signal. Among the many techniques for tuning parameter selection, Cross-validation (CV) and Information Criteria (IC) are familiar options. Considering the nature of time series data in our empirical study and the large number of covariates, we apply Generalized Information Criteria (GIC) to determine the tuning parameter.

With high-dimensional penalized likelihood, [Fan and Tang \(2013\)](#) recommend that GIC is used to select the tuning parameter. Its advantage draws from a trade-off between model fitting and model complexity, which is achieved by involving two components. The first evaluates the goodness of fit which rises with the number of explanatory variables. The second penalizes model complexity as associated with the

number of variables. The GIC for a given λ is calculated as follows:

$$GIC_\lambda = \frac{1}{N} \left\{ 2[LLF(\hat{\Phi}) - PLLF_\lambda(\hat{\Phi}_\lambda)] + a(N, p)|\hat{\theta}_\lambda| \right\} \quad (3.13)$$

where N is the number of observations. $LLF(\hat{\Phi})$ is the maximum value of the likelihood function including all candidate variables, and $PLLF_\lambda(\hat{\Phi}_\lambda)$ is the maximum value of penalized likelihood function for a given λ with variable selection. $2[LLF(\hat{\Phi}) - PLLF_\lambda(\hat{\Phi}_\lambda)]$ is the scaled deviation between the original model including all variables and the model with variable selection. $|\hat{\theta}_\lambda|$ is the l_1 -norm of parameter vector θ for a given λ . $a(N, p)$ is a positive value regulating the trade off between model fitting and model complexity, which depends on the number of observations N and the number of estimated parameters p . Following Fang et al. (2020), we set $a(N, p) = \log\{\log(N)\} \cdot \log(p)$. For the empirical analysis, the optimum tuning parameter λ^* is selected according to the minimum GIC_λ over the range $[0, \lambda_{max}]$.

3.2.4 Estimation framework

Two problems arise with the estimation stage: dynamic selection and identification. With regard to the first, variable screening is usually static offering limited support to out-of-sample forecasting. It is also unreasonable to assume the same variables would be selected at different time and the selected variables would make the same contribution in predicting the target variable over time. To address this issue, we conduct a dynamic variable selection process where the optimum tuning parameter λ^* and important variables are selected on a rolling-window basis.

With the identification problem, it arises when variables are dropped from the model if their respective parameters are shrunk to zero during the screening process. This would lead to an identification problem as the Beta Weighting parameters of the dropped variables will not enter in the penalized likelihood function and therefore will not be identified. To avoid this problem, we fix the Beta Weighting parameters when estimating the rest of the parameters. Specifically, suppose the penalized likelihood function for the ES-CAViaR-MIDAS model depends on two sets of parameters $\Phi_1 = (\mu, \beta_1, \beta_2, \beta_3, \theta_1, \theta_2, \dots, \theta_{20})$ and $\Phi_2 = (w_{1,1}, w_{1,2}, w_{2,1}, w_{2,2}, \dots, w_{20,1}, w_{20,2})$. We fix the Beta Weighting parameters $\Phi_2 = \hat{\Phi}_2$ before estimating Φ_1 where $\hat{\Phi}_2$ is sequentially

obtained by estimating the ES-CAViaR-MIDAS model for each candidate variable.

The two-stage procedure is proposed to address the respective problems in a single framework, which we illustrate as follows with added details given in Figure C.1.

Step 1: Obtain the Beta Weighting parameter estimates and calculate the adaptive weight

- Estimate the ES-CAViaR-MIDAS model for each of the J variables by maximizing the AL likelihood function as in Eq.(3.11) under linear constraints, and obtain Beta Weighting parameter estimates $\Phi_2 = \hat{\Phi}_2$.
- Set $\Phi_2 = \hat{\Phi}_2$. Estimate the ES-CAViaR-MIDAS model with all J variables by maximizing the AL likelihood function as in Eq.(3.11) under linear constraints, and obtain the parameter estimates $\hat{\theta}_j$. Calculate the adaptive weight as $\hat{w}_j = \frac{1}{|\hat{\theta}_j|^n}$.

Step 2: Dynamically choose the optimum tuning parameter and select variables

- The entire out-of-sample period is divided into 66 rolling windows. For each rolling window, we choose the optimum tuning parameter and select variables.
- Set $\Phi_2 = \hat{\Phi}_2$ and $w_j = \hat{w}_j$. For each rolling window, we estimate the ES-CAViaR-MIDAS model with variable selection by maximizing the penalized AL likelihood function as in Eq.(3.12), with the tuning parameter λ on a grid of $[0, \lambda_{max}]$. Obtain the parameter estimates $\hat{\Phi}_1$ and calculate GIC for each value of λ .
- Choose the optimum tuning parameter λ^* according to the minimum GIC and obtain the selected variables.
- Repeat the above process until the last window is reached.

[INSERT FIGURE C.1 ABOUT HERE]

3.3 Data and empirical study

In the empirical analysis, we investigate the S&P500, US macroeconomic and financial data from 1969Q1 to 2021Q2. The S&P500 index data is obtained from Yahoo

Finance and We consider the daily stock market returns which are calculated as the natural logarithm of S&P500 index prices. With macroeconomic and financial variables collected quarterly, it is necessary to align the frequency of the variables, that are recorded at daily/monthly frequencies, by taking quarterly averages.

Substantial data revision can be frequent for macroeconomic and financial variables [Conrad and Loch \(2015\)](#). As a result, using revised instead of real-time data may mislead the forecast evaluations. The importance of using real-time data has long been documented in the empirical studies. For example, [Stark \(2010\)](#) shows that the forecast accuracy of the Survey of Professional Forecasters declines as the data are revised over time, indicating the negative impact of using revised data. [Croushore \(2011\)](#) argues that forecasters generally produce forecasts based on existing methodologies and cannot be expected to predict future changes in methodology. Thus, forecast evaluations should focus on early release of data and ideally the real-time data.

In this chapter, we obtain the first release (real-time) macroeconomic and financial data from the Real-time Data Research Center (RDRC) of the Federal Reserve Bank of Philadelphia. Other variables are collected from the FRED database at the Federal Reserve Bank of St Louis, Tradingeconomics.com, the Survey of Consumers from University of Michigan (SCUM), the Federal Reserve Bank of Chicago (FRBC) and the personal website of French and Manela. Descriptive statistics and the variable correlation matrix are reported in Table C.1 and Table C.2 respectively. ⁴

3.3.1 Macroeconomic variables

The following papers provide a context for our macroeconomic variables: [Paye \(2012\)](#), [Engle, Ghysels, et al. \(2013\)](#), [Asgharian et al. \(2013\)](#), [Conrad and Loch \(2015\)](#), [Pan et al. \(2017\)](#), [Su et al. \(2017\)](#), [Conrad and Kleen \(2019\)](#) and [Fang et al. \(2020\)](#). Macroeconomic variables in our study are real GDP growth rate, industrial production growth rate, unemployment rate, housing starts, post-tax nominal corporate profits, real personal consumption, CPI, PPI, the Chicago Fed national activity index (CFNAI), the new orders index of the Institute of Supply Management, monetary base, consumer sentiment index of the University of Michigan, real GDP volatility and CPI volatility.

CFNAI is an index designed to gauge overall economic activity and related

⁴Tradingeconomics.com is a database offering economic and financial data.

inflationary pressure. It can be seen as a proxy of business cycles. The New Orders Index measures the changes in employment, production, inventories, supplier deliveries and new orders. It provides an idea of future economic growth and is a leading economic indicator. In addition, macroeconomic volatility is found to be important determinant of stock market volatility (Schwert (1989), Liljeblom and Stenius (1997), Engle, Ghysels, et al. (2013)). The macroeconomic volatility is measured by volatility of first release GDP growth rate and volatility of CPI. Following Engle, Ghysels, et al. (2013), we use GARCH(1,1) model to estimate quarterly macroeconomic volatility.

Following Conrad and Loch (2015) and Fang et al. (2020), we consider CFNAI in levels and take annualized quarterly percentage change as $100((X_t - X_{t-1})^4 - 1)$ for other variables.

3.3.2 Financial variables

We include six financial variables: equity market returns (MKT), short-term reversal factor (STR), default spread, term spread, realized volatility (RV) and implied volatility (IV).

Equity market returns (MKT) considered in Fama and French (1992) capture the leverage effect, showing the negative relationship between stock returns and volatility (Black (1976), Christiansen et al. (2012), Nonejad (2017)). From Nagel (2012), the expected return of reversal strategy rises predictably and dramatically during periods of market turmoil, thereby indicating that the short-term reversal factor (STR) is related to stock market volatility. For the spread terms, the following proxies are used: the yield spread between BAA and AAA rated bonds for default spread; the yield spread between the 10-year treasury bond and the 3-month treasury bill for term spread. Following Andersen, Bollerslev, and Meddahi (2011), we also consider realized volatility in forecasting volatility. The quarterly volatility is calculated as:

$$RV_t = \sum_{i=1}^{N_t} r_{i,t}^2 \quad (3.14)$$

The implied volatility is a proxy of financial market uncertainty. CBOE VIX and VXO are two implied volatility indices which are used to measure the expected market volatility implied by stock index option prices. Many studies find that options-based

volatility is more informative for forecasting purposes than time-series volatility models based on returns of stock market index (Martens and Zein (2004), Becker et al. (2009), D. Fernandes et al. (2014)). Therefore, implied volatility is considered as an explanatory variable in the long-term volatility component. However, the data horizons of VIX and VXO do not match that of the other variables given their availability from 1990 and 1986 respectively. Manela and Moreira (2017) propose a news-based implied volatility index (NVIX) to capture investors' perception of future financial market uncertainty. It is an estimated VXO index based on the data from the front-page articles of the Wall Street Journal. Su et al. (2017) find that NVIX is a source of financial aggregate volatility. Thus, we use NVIX as a proxy of implied volatility before 1986 and switch to VXO between 1986 and 1990. CBOE VIX is used thereafter in the empirical analysis.

[INSERT TABLES C.1-C.2 ABOUT HERE]

3.3.3 Principal Component Analysis (PCA) for variables

Dimensionality reduction techniques such as Principal Component Analysis (PCA) are alternatives of feature selection methods in big data analysis. LASSO, as a feature selection method, focuses on dropping uninformative variables, while PCA creates a lower-dimensional representation of the original features (Jolliffe and Cadima (2016)). PCA is widely used to reduce the dimensionality of big datasets, increasing interpretability but minimizing the information loss at the same time. It does so by creating principle components that maximizes the data variation and therefore retaining as much the statistical information as possible. To examine the power of lower-dimensional features represented by the principle components, we conduct the following PCA analysis.

We conduct PCA for the 20 macroeconomic and financial variables. We standardize all the variables for implementation. From the PCA, we find that the first, second and third principal component respectively accounts for 28.12%, 13.91% and 11.22% of the variation in the 20 variables. Moreover, the first three components only explain 53.25% of the variation in the data in total.⁵ The results indicate that PCA is insufficient in representing the original features. Therefore, the Adaptive-Lasso is used to select

⁵More detailed results can be found in Table C.3.

informative variables for forecasting, instead of using the principal components directly.⁶

[INSERT TABLE C.3 ABOUT HERE]

3.4 Empirical analysis

We evaluate one-day-ahead VaR and ES forecasts for the daily log returns of S&P 500 index using four confidence levels: 1%, 2.5%, 5% and 10%. For each model and confidence level, one-day ahead VaR and ES forecasts in one quarter are generated by the parameter estimates based on the data in the previous 36 quarters. After each estimation, we move the rolling window one quarter forward. The out-of-sample period for each model is from 2005Q1 to 2021Q2.

We backtest the VaR and ES forecasts of the proposed model with the selected variables and compare its performance with that of the following benchmark models: the ES-CAViaR-MIDAS model that incorporates one variable at a time in the long-term volatility component (Conrad and Loch (2015), Fang et al. (2020)), the ES-CAViaR-MIDAS model without any low-frequency variables, the ES-CAViaR-MIDAS model with all 20 macroeconomic & financial variables, and the ES-CAViaR-MIDAS model that incorporates the first three principal components (PC1, PC2, PC3) separately and jointly. In addition to the model integrating all selected variables, we generate the combination forecasts based on the forecast result of the ES-CAViaR-MIDAS model with each of the selected variable using the simple average combination and loss function-based combination (Conrad and Kleen (2019), Happersberger et al. (2020)). Description of these models are provided in Table C.4.

[INSERT TABLE C.4 ABOUT HERE]

3.4.1 Dynamic variable selection using Adaptive-Lasso

Using a dynamic variable selection process, whereby the optimum tuning parameter and corresponding variables are selected on a rolling quarterly basis, allows the impact of

⁶But we consider incorporating the first three principal components into the proposed model as benchmarks.

low-frequency variables on the tail risk to be visualized and provides a new perspective to add the most important information variables in predicting VaR and ES.

We first standardize the macroeconomic and financial variables. To select the tuning parameter λ , we then review a $[0, 30]$ range in increments of 0.1, calculating the GIC for the respective value of λ . The λ with the minimum GIC in each window is regarded as the optimum tuning parameter.

The third panel of Figures C.2 - C.5 present the results of dynamic variable selection process at four confidence levels 1%, 2.5%, 5% and 10%. The results are reported by time and then by confidence levels.

Across the out-of-sample period, RV, term spread and housing starts mostly serve to predict VaR and ES across at all chosen confidence levels. Past market volatility, the yield spread between long-term and short-term government bonds and the volume of housing starts are predominant predictors of future tail risk. As indicated by the heat map, those three variables have a consistently strong but opposite impact on future VaR and ES. For RV, the green tabs in the heat map indicate its positive correlation with VaR and ES. That is, the higher the RV, the greater the future VaR and ES. Several studies have found that more accurate risk forecasts can be generated by incorporating RV into the dynamic volatility forecasting models (see, e.g., [Giot and Laurent \(2004\)](#), [Clements et al. \(2008\)](#), [Hansen et al. \(2012\)](#), [Louzis et al. \(2014\)](#), [Lazar and Xue \(2020\)](#)). Different from the previous studies, we allow the data speak for itself rather than assuming statistical models which include RV but may not fit the data well. Our study shows that the RV is consistently selected as one of the strongest predictors of future tail risk by the Adaptive-Lasso. However, term spread and housing starts, as indicated by their red tabs, are negatively associated with future tail risk. The larger the term spread and housing starts, the lower the future tail risk. To the best of our knowledge, our study provides novel evidence for the negative relationship between housing starts and financial market tail risk, indicating the countercyclical pattern of tail risk. Inspired by [Conrad, Loch, and Daniel \(2014\)](#) and [Fang et al. \(2020\)](#), the relevance of term spread and housing starts for tail risk prediction drives from the following: (1) These two variables are powerful predictors of future changes in economic activity. Since either a smaller term spread or yield curve inversion is mostly followed by recessions, these are associated with increased uncertainty and financial market risk. More housing starts,

however, implies future economic growth which can be rationalized by the empirical observation of low mortgage interest rate that facilitate economic upturns and safer investment projects (see, e.g., [Estrella and Trubin \(2006\)](#), [Leamer \(2007\)](#), [Kydland et al. \(2012\)](#)); (2) Term spread is closely related to the investors' expectations on investment risk. If short-term interest rates are expected to fall, this raises the price of long-term securities so lowering their yields relative to the short-term securities and indicating higher risk ([Wheelock, Wohar, et al. \(2009\)](#)). However, more housing construction is associated with the expansion of the credit market, which drives economic growth and investment opportunities with lower risk. During the great recession period between 2007 and 2009, in addition to the three strongest predictors, corporate profits is also selected at all chosen confidence levels. The red tabs of corporate profits suggests its negative association with tail risk. Corporate profits has strong impact on firm valuation and is a key determinant of financial investment decisions. Firms with higher corporate profits, especially during turbulent period, are relatively safer investment targets perceived by investors ([Froot et al. \(1993\)](#)). During the COVID-19 crisis period, RV, term spread and housing starts dominate other variables in forecasting stock market tail risk. Although RV, term spread and housing starts are three strongest predictors of tail risk, the importance of other low-frequency variables should not be neglected.

Across the confidence levels, as is shown in the second panel of Figures C.2 - C.5, the average number of selected variables is the highest at 1% level while is the lowest at 10% level. This indicates that extreme VaR and ES may rely on extra information in addition to RV, term spread and housing starts. For example, at 1% confidence level, other low-frequency variables such as unemployment rate, corporate profits and MKT are also consistently selected for most of the rolling windows. However, fewer variables are selected as confidence level increases.

[INSERT FIGURES C.2 - C.5 ABOUT HERE]

3.4.2 Backtesting approaches

We implement six backtesting approaches to compare the performance of out-of-sample forecasts generated by the proposed model with that by other benchmarks. We now introduce six prevailing backtesting methods for VaR and ES via assessing the quantile score, the unconditional coverage (UC) test, the Dynamic Quantile (DQ) regression,

the bootstrap test for ES and the Dynamic Expected Shortfall (DES) regression.

3.4.2.1 Unconditional coverage (UC) test

Familiar procedures for evaluating the performance of VaR forecasts are based on VaR failures, i.e.,

$$I_t = \mathbf{1}\{r_t \leq v_t^\alpha\}.$$

One commonly used VaR backtesting method, known as the unconditional coverage (UC) test, is proposed by [Kupiec \(1995\)](#), and uses the proportion of failures as its main tool. In this test, the hit proportion is defined as the percentage of the returns that are below the estimated VaR. The difference between the hit proportion and theoretical value α is then examined. The decision on the null hypothesis depends on the Likelihood Ratio (LR) test that is applied:

$$H_{UC}^{VaR} : \mathbb{E}_{t-1}[I_t] = \alpha.$$

3.4.2.2 Dynamic quantile (DQ) regression

Not only is the UC test statistically weak for small samples, it is criticized for ignoring the clustering of failures ([Nieto and Ruiz \(2016\)](#)). To address these drawbacks, the conditional coverage (CC) test is considered where the null hypothesis is:

$$H_{CC}^{VaR} : \mathbb{E}_{t-1}[I_t | I_{t-1}] = \alpha.$$

We employ the dynamic quantile (DQ) test proposed by [Engle and Manganelli \(2004\)](#) to implement the CC test. The DQ test is valid under the misspecification when ignoring the conditionally correlated probabilities and can be extended to examine other explanatory variables. The DQ test examines both whether the hit variable defined as $Hit_{v,t} = \mathbf{1}\{r_t \leq v_t\} - \alpha$, follows an i.i.d. Bernoulli distribution with probability level α and whether it is independent of the VaR estimator; the expected value of $Hit_{v,t}$ is 0. Furthermore, given the definition of the quantile function, the conditional expectation of v_t given any information known at $t - 1$ must also be 0. The implication is that the hit function cannot be correlated with other lagged variables. Similarly, the $Hit_{v,t}$ must not be autocorrelated. If $Hit_{v,t}$ satisfies the above conditions, there will be no

autocorrelation in the hits and no measurement error. We include one lag of $Hit_{v,t}$ in the regression function of the test. Consider the following DQ regression:

$$Hit_{v,t} = a_0 + a_1 Hit_{v,t-1} + a_2 v_{t-1} + u_{v,t}, \quad (3.15)$$

where $\mathbf{a} = [a_0, a_1, a_2]$ is the set of parameters of the regression. We test whether all parameters in the set \mathbf{a} are zero. The DQ test statistic follows an asymptotic $\chi^2(3)$ distribution under the null hypothesis.

3.4.2.3 Quantile score function

An alternative way to assess the out-of-sample VaR forecasts is to use the proposed quantile score function. Given its use in quantile regression, a reasonable choice for the score function is the linear piecewise loss function for VaR ([Giacomini and Komunjer \(2005\)](#)), which can be modified as the quantile score function expressed as:

$$S(v_t, r_t; \alpha) = (r_t - v_t)(\alpha - I_t). \quad (3.16)$$

As VaR is an elicitable risk measure, the quantile score is strictly consistent for VaR. By this backtesting method, the best model is that which generates the lowest score from the above score function.

3.4.2.4 Bootstrap test for ES

ES is not elicitable ([Gneiting \(2011\)](#)) given the absence of a suitable loss function for ES. In backtesting the ES, [McNeil and Frey \(2000\)](#) propose a bootstrap test which focuses upon discrepancies between the observed return and the ES forecast, for the periods where the return exceeds the VaR forecast. Under the null hypothesis, the standardized discrepancies should have zero unconditional and conditional expectations. Given the typically small sample of discrepancies, a test of zero conditional expectation is generally not performed. By implication, the dynamic properties of the ES estimates are not then evaluated.

To address the problem of sample size, [McNeil and Frey \(2000\)](#) employ a bootstrap test. This avoids the need for any distributional assumption and test for zero

unconditional mean of the VaR exceptions. As this test focuses on observations exceeding the VaR forecasts, the assessment of ES forecasts is not independent of those forecasts. This problem, along with the nonelicitability of ES, prompts consideration of a scoring function for jointly evaluating ES and VaR forecasts.

3.4.2.5 Dynamic Expected Shortfall (DES) regression

We follow the backtesting method of [Patton et al. \(2019\)](#) to evaluate the ES estimates individually, using a dynamic ES (DES) regression test:

$$\lambda_{e,t}^s = b_0 + b_1 \lambda_{e,t-1}^s + b_2 e_{t-1} + u_{e,t}, \quad (3.17)$$

where $\lambda_{e,t}^s$ is the standardized version of $\lambda_{e,t}$, which is defined as:

$$\lambda_{e,t}^s = \frac{\lambda_{e,t}}{e_t} = \frac{1}{\alpha} \mathbf{1}\{r_t \leq v_t\} \frac{r_t}{e_t} - 1,$$

$\mathbf{b} = [b_0, b_1, b_2]$ is the set of parameters of the regression. Based on the null hypothesis, we test whether all parameters in set \mathbf{b} are zero. The main intuition for this test is the same as that for the Dynamic quantile (DQ) regression discussed above.

3.4.2.6 FZ loss function for (VaR, ES)

To compare the VaR and ES forecasts jointly, a loss function proposed by [Fissler and Ziegel \(2016\)](#) is employed. The authors discuss how VaR and ES are jointly elicitable and present a group of loss functions for risk measure estimation and backtesting. We follow the choice of [Patton et al. \(2019\)](#) for the loss function FZ0, which is defined as:

$$L_{FZ0}(Y, v, e; \alpha) = -\frac{1}{\alpha e} \mathbf{1}\{Y \leq v\} (v - Y) + \frac{v}{e} + \log(-e) - 1. \quad (3.18)$$

To compare the performance of each model using the FZ0 loss function, we calculate the average loss value $L_{FZ0} = \frac{1}{T} \sum_{t=1}^T L_{FZ0,t}$ for different α values.

3.4.3 Backtesting results

The out-of-sample forecasting period is of 2005Q1 to 2021Q2, encompasses both the global financial crisis 2007 to 2009 and the COVID-19 recession starting early 2020. To compare the forecasting performance of the proposed model with the benchmarks, we present the backtesting results based on the above approaches in Table C.5. Table C.6 displays the performance rankings for the joint forecasting of VaR and ES based on the FZ loss function. We first report the overall findings, followed by the detailed results for each backtesting approach.

For most of the applied back tests and confidence levels, our proposed ES-CAViaR-MIDAS model with selected variables and the combination forecasts consistently outperform other benchmark models. Across all confidence levels, the loss-based forecast combination using the forecasts generated by each of the selected variable achieves the minimum loss in forecasting VaR and ES jointly. The proposed ES-CAViaR-MIDAS model with selected variables ranks second in joint forecasting VaR and ES. This suggests that the low-frequency information is valuable in forecasting high-frequency tail risk and that its power is further strengthened by the dynamic variable selection process. The detailed backtesting results are discussed as follows.

Panel A in Table C.5 gives average score of VaR forecasts generated by each model at different confidence levels. The lowest three values for each confidence level are presented in **bold**. In general, the loss function-based combination forecasts using forecasts generated by each selected variable perform the best across different confidence levels. In addition, for $\alpha = 1\%$ and $\alpha = 2.5\%$, ES-CAViaR-MIDAS model with the variables selected by our proposed approach performs well compared with other benchmarks, delivering lower score for the VaR forecasts. The simple average combination method is also highly ranked in some cases, e.g., for $\alpha = 5\%$. However, the out-of-sample performance of the ES-CAViaR-MIDAS that incorporates all 20 macroeconomic & financial variables has higher average loss. This indicates that a model without shrinkage and regularization suffers from the overfitting problem.

The hitting proportions and their corresponding p -values are presented in Panel B of Table C.5. We first count the number of VaR rejections for each model, then perform the UC backtest for all forecasts at four confidence levels. The hitting proportions with p -values larger than 10% are presented in **bold**, and hitting proportions with p -values

larger than 5% are presented in *italics*. For $\alpha = 5\%$ and $\alpha = 10\%$, the proposed ES-CAViaR-MIDAS model with selected variables and two forecast combinations achieve fewer UC test rejections compared to the rejections of other benchmarks. In more extreme cases, i.e., $\alpha = 1\%$ and $\alpha = 2.5\%$, all models fail to pass the UC test, having p -values larger than 5%, even for the original ES-CAViaR-MIDAS model without any low-frequency variables. The misspecification of the original model might be one of the causes of the bias.⁷

Panel C of Table C.5 reports p -values of the DQ test of VaR forecasts at four confidence levels. p -values greater than 10% are presented in **bold** indicating no evidence against the optimality at 10% significance level. p -values above 5% are in *italics*. For $\alpha = 5\%$ and $\alpha = 10\%$, all models pass the DQ test. The proposed ES-CAViaR-MIDAS model with selected variables achieves the largest p -value at $\alpha = 10\%$. When we consider the 2.5% VaR forecasting, only a few models pass the DQ test. However, significant improvements are achieved by adding the selected variables in the ES-CAViaR-MIDAS model which has the highest p -value in this case. For the most extreme case, i.e., $\alpha = 1\%$, integrating all the selected variables into the ES-CAViaR-MIDAS model does not pass the DQ test, although the loss-based combination forecasts are able to pass the DQ test significantly. Overall, adding the selected variables generally helps models to pass the DQ test.

To backtest the performance of the models for VaR and ES jointly, we apply the FZ loss function to compute the average loss for all models. Panel D of Table C.5 presents the average loss at different confidence levels. The lowest three average losses in each column are presented in **bold**. Similar to the results in Panel A, the loss-based forecast combination method consistently has the lowest average loss for all confidence levels. In more extreme cases, i.e., $\alpha = 1\%$ and $\alpha = 2.5\%$, the ES-CAViaR-MIDAS model with all selected variables always ranks Top3 models. In addition, models incorporating RV perform better than models including other variables in this backtesting scheme. Overall, dynamic variable selection improves the joint forecasting of VaR and ES.

To evaluate the ES forecasts, we employ the bootstrap test for the zero unconditional and conditional expectation of the standardized discrepancies, between the observed returns and the ES forecasts, for the periods where the returns exceed the VaR forecasts.

⁷In further study, we would consider other semiparametric models integrating the MIDAS framework, then assess the out-of-sample performance.

Results are shown in Panel E of Table C.5. In this panel, **bold** indicates the ES forecasts at each confidence level that passes the bootstrap test with p -values above 10%; and where the p -values above 5% they are presented in *italics*. Overall, the ES-CAViaR-MIDAS model with selected variables and the combination approaches perform best, having the largest p -values.

In Panel F of Table C.5, we show the p -values of the DES test on the ES forecasts at four confidence levels. Also, p -values above 10% are indicated in **bold**, indicating no evidence against the optimality at 10% significance level. The p -values above 5% are in *italics*. In like fashion to the results of the DQ test, the incorporation of machine learning selected variables in the ES-CAViaR-MIDAS model helps to pass the DES test significantly in the cases of $\alpha = 5\%$ and $\alpha = 10\%$. It is worth to mention that at the confidence level of 2.5%, the p -value for the ES-CAViaR-MIDAS model with our proposed selection method is the highest among others. For the most extreme case of $\alpha = 1\%$, although the model with all selected variables does not pass the test significantly, the loss-based combination forecasts using forecasts generated by each of the selected variable has the highest p -value.

[INSERT TABLE C.5 ABOUT HERE]

The first panel of Table C.6 presents the rankings of out-of-sample performance based on the values of FZ loss function for four confidence levels. The best model ranks 1 while the worst ranks 29 given there are 29 competing models in total. Columns 1-4 report the rankings for each confidence level. Columns 5-6 present the average loss for four confidence levels and the rankings of the average loss respectively. In general, the loss-based combination forecasts using forecasts generated by each selected variable ranks first, followed by the the ES-CAViaR-MIDAS model with selected variables. The only exception is at 10% confidence level where the loss-based combination forecasts ranks second. This indicates that the selected variables improve the joint forecasting of VaR and ES, showing the value of effective low-frequency information.

[INSERT TABLE C.6 ABOUT HERE]

3.4.4 Robustness checks

3.4.4.1 Restricted Beta Weighting schemes

Under the restricted Beta Weighting scheme, w_1 is set to 1 to generate a decaying pattern of weights rather than the hump-shaped weights in the previous analysis. The use of restricted Beta Weighting scheme may lead to new results. In this section, we fix $w_1 = 1$ to see how this constraint changes the results with all other settings being the same as before. The Beta Weighting scheme now becomes:

$$\psi_k(w_2) = \frac{(1 - k/(K + 1))^{w_2 - 1}}{\sum_{s=1}^K (1 - s/(K + 1))^{w_2 - 1}}, \quad (3.19)$$

The third panel of Figures C.6 - C.9 present the results of dynamic variable selection process under the restricted Beta Weighting scheme at four confidence levels 1%, 2.5%, 5% and 10%. In general, similar results are obtained under the restricted Beta Weighting scheme. Industrial production, housing starts, term spread and RV are predominant predictors of future tail risk. For industrial production and housing starts, their red tabs in the heat map indicate that they are negatively correlated with future VaR and ES, providing evidence for the countercyclical pattern of tail risk. For term spread and RV, their positive relationship with future tail risk is shown by the green tabs in the heat map. During the great recession period between 2007 and 2009, in addition to those strongest predictors, corporate profits is also consistently selected and negatively correlated with tail risk. During the COVID-19 crisis period, industrial production, housing starts, term spread and RV dominate other low-frequency variables in forecasting tail risk. Moreover, as is shown in the second panel of Figures C.6 - C.9, the average number of selected variables is the highest at 1% confidence level while is the lowest at 10% level. For example, at 1% level, additional to the strongest predictors, other variables such as unemployment rate and MKT are also consistently selected for most of the rolling windows. However, fewer variables are selected when confidence level increases. This indicates that more information may be needed to predict more extreme VaR and ES.

[INSERT FIGURES C.6 - C.9 ABOUT HERE]

The second panel of Table C.6 displays the the performance rankings for the joint

forecasting of VaR and ES under the restricted Beta Weighting scheme based on the FZ loss function. The loss-based combination forecasts using forecasts generated by each selected variable consistently ranks first, followed by the the ES-CAViaR-MIDAS model with selected variables. Model without variable selection ranks 24th among 29 competing models which indicates that simply adding the low-frequency information deteriorates the model performance while the effective variables selected by our proposed process improve the joint forecasting of VaR and ES.

To better visualize the model performance, Table C.7 presents the backtesting results under the restricted Beta Weighting scheme based on the same backtesting approaches in the previous section. For most of the applied backtests and confidence levels, our proposed ES-CAViaR-MIDAS model with selected variables and the combination forecasts consistently outperform other benchmark models. For the joint forecasting of VaR and ES, the loss-based forecast combination using the forecasts generated by each of the selected variable achieves the minimum loss at all confidence levels. The proposed ES-CAViaR-MIDAS model with selected variables ranks second. For the individual forecasting of VaR and ES, the corresponding backtesting results also show the superiority of our proposed model by achieving either lower losses or higher p -values.

[INSERT TABLE C.7 ABOUT HERE]

3.5 Conclusion

Accurate VaR and ES forecasts enhance statistical analysis relevant to decisions taken by financial and risk managers, regulators and market participants. In our new framework, the integration of low-frequency signals (of macroeconomic and financial indicators) into high-frequency tail risk forecasting (using the proposed ES-CAViaR-MIDAS model) improves the accuracy of those forecasts. However, simply adding all low-frequency variables are found to deteriorate tail risk forecasts. To select the most informative low-frequency variables, we propose an innovative approach that maximizes the penalized Asymmetric Laplace (AL) likelihood function with an Adaptive-Lasso penalty. By integrating the selected variables, our proposed model achieves the minimum loss in the joint forecasting of VaR and ES.

Three variables, namely, realized volatility, term spread and housing starts are consistently selected for most of the rolling windows and serve to the strongest predictors of future tail risk. To the best of our knowledge, the empirical relationships between housing starts and tail risk is a unique and innovative result. In the impact of housing starts on VaR and ES, we identify new evidence of the countercyclical risk measures. In addition to the strongest predictors, the value of other low-frequency variables should not be ignored.

In conclusion, our study provides novel evidence for the value of low-frequency macroeconomic & financial variables in the high-frequency tail risk forecasting. The proposed new framework enable market participants and regulators utilizing broader information in practical applications.

Appendix C

Chapter 3

C.1 Tables and Figures

Table C.1: Summary statistics

Variable	Obs.	Min.	Med.	Max.	Mean	Std.	Skew.	Kurt.	Database
<i>Stock market data</i>									
S&P 500 returns	13240	-22.90	0.05	10.96	0.03	1.08	-1.02	25.21	DataStream
<i>Macroeconomic data</i>									
Real GDP	210	-32.90	2.53	33.08	2.38	4.41	-1.15	30.07	RDRC
Industrial production	210	-25.71	3.03	18.81	2.09	6.23	-1.22	4.19	RDRC
Unemployment rate	210	-4.20	-0.07	9.20	0.01	0.80	6.67	87.17	RDRC
Housing starts	210	-76.50	4.20	304.56	8.48	46.89	2.28	10.72	RDRC
Corporate profits	210	-88.01	8.59	407.35	12.20	40.24	5.53	48.98	FRED
Personal consumption	210	-34.61	3.07	40.70	2.92	4.70	-0.20	38.60	RDRC
CPI	210	-10.84	3.42	16.76	4.00	3.53	0.61	2.57	FRED
PPI	210	-37.79	3.41	31.60	3.94	7.98	-0.10	4.81	FRED
CFNAI	210	-3.62	0.07	2.01	-0.03	0.89	-1.59	4.44	FRBC
New orders	210	27.27	56.08	71.90	55.10	7.52	-0.81	1.27	Trading
Moneytary base	210	-19.31	6.48	605.97	11.97	46.93	10.68	128.23	FRED
Consumer sentiment	210	-22.57	-0.10	16.27	-0.03	5.31	-0.27	1.83	SCUM
Real GDP volatility	210	2.78	3.99	31.16	4.64	2.94	6.55	53.73	RDRC ^a
CPI volatility	210	2.56	3.96	13.79	4.72	2.22	1.87	3.40	RDRC ^a
<i>Financial data</i>									
MKT	210	-9.72	1.00	7.29	0.57	2.94	-0.68	0.94	French
STR	210	-8.66	0.49	7.66	0.47	1.89	-0.15	3.98	French
Default spread	210	0.56	0.96	3.02	1.08	0.43	1.81	4.24	FRED ^a
Term spread	210	-1.43	1.69	3.80	1.62	1.21	-0.38	-0.57	RDRC
RV	210	8.14	42.88	1143.58	73.48	123.40	6.14	44.16	DataStream ^a
IV	210	10.31	20.43	58.58	20.33	6.11	2.10	9.80	FRED & Manela

Note: This table reports descriptive statistics for daily stock market returns and quarterly macroeconomic & financial variables. Descriptive statistics includes number of observations (Obs.), minimum (Min.), maximum (Max.), mean (Mean.), standard deviation (Std.), Skewness (Skew.) and Kurtosis (Kurt.).

^a Indicates that the variable is computed by authors using the data from stated database

Table C.2: Variable correlation matrix

	GDP	IP	Unemp	HS	CP	PC	CPI	PPI	CFNAI	NO	Money	Sentiment	GDP vol	CPI vol	MKT	STR	DS	TS	RV	IV	
GDP	1.000																				
IP	0.546	1.000																			
Unemp	-0.833	-0.319	1.000																		
HS	0.420	0.234	-0.322	1.000																	
CP	0.328	0.134	-0.219	0.204	1.000																
PC	0.855	0.335	-0.762	0.442	0.375	1.000															
CPI	-0.017	-0.085	-0.034	-0.068	0.027	-0.013	1.000														
PPI	0.187	0.149	-0.215	0.021	0.111	0.157	0.728	1.000													
CFNAI	0.690	0.795	-0.583	0.282	0.099	0.507	0.018	0.212	1.000												
NO	0.600	0.736	-0.476	0.194	0.206	0.393	-0.076	0.245	0.835	1.000											
Money	-0.328	-0.050	0.396	-0.183	-0.113	-0.327	-0.301	-0.360	-0.366	-0.277	1.000										
Sentiment	0.317	0.206	-0.334	0.239	0.131	0.337	-0.123	-0.132	0.283	0.274	-0.196	1.000									
GDP vol	0.375	0.276	-0.350	0.290	0.283	0.417	0.058	0.197	0.213	0.285	-0.028	0.151	1.000								
CPI vol	-0.260	-0.345	0.166	-0.081	-0.005	-0.139	0.671	0.289	-0.297	-0.392	0.012	-0.062	0.046	1.000							
MKT	-0.011	0.011	0.088	0.229	-0.011	0.003	-0.130	-0.095	0.005	0.066	-0.117	0.251	0.143	-0.104	1.000						
STR	-0.141	-0.104	0.167	0.007	0.007	-0.084	0.084	0.061	-0.048	-0.043	-0.157	-0.026	0.019	0.077	0.364	1.000					
DS	-0.246	-0.339	0.256	0.037	0.106	-0.143	0.032	-0.228	-0.489	-0.467	0.318	0.059	0.022	0.430	0.056	0.026	1.000				
TS	0.109	0.100	-0.045	0.190	0.132	0.058	-0.388	-0.242	0.067	0.242	0.098	0.202	-0.049	-0.366	0.138	0.044	0.182	1.000			
RV	-0.272	-0.282	0.256	-0.155	0.020	-0.282	-0.197	-0.244	-0.458	-0.372	0.483	-0.221	-0.037	0.079	-0.408	-0.141	0.372	0.011	1.000		
IV	-0.175	-0.189	0.237	-0.095	0.135	-0.156	-0.107	-0.212	-0.333	-0.316	0.488	-0.204	0.111	0.171	-0.265	-0.030	0.503	0.075	0.826	1.000	

Table C.3: Principal Component Analysis for 20 macroeconomic and financial variables

Component	Explanatory power	Cumulative sum
1	28.12%	28.12%
2	13.91%	42.03%
3	11.22%	53.25%
4	9.34%	62.59%
5	6.06%	68.65%
6	4.72%	73.37%
7	4.58%	77.96%
8	3.73%	81.69%
9	3.67%	85.36%
10	2.99%	88.35%
11	2.45%	90.79%
12	2.04%	92.84%
13	1.88%	94.72%
14	1.44%	96.16%
15	1.11%	97.27%
16	0.83%	98.10%
17	0.63%	98.74%
18	0.61%	99.34%
19	0.38%	99.72%
20	0.28%	100.00%

Table C.4: Description of competing models

Model name	Description
Variable name	ES-CAViaR-MIDAS model that incorporates one variable at a time
None	ES-CAViaR-MIDAS model without any low-frequency variables
All	ES-CAViaR-MIDAS model with all 20 macroeconomic & financial variables
PC1	ES-CAViaR-MIDAS model with the first principal component PC1 only
PC2	ES-CAViaR-MIDAS model with the second principal component PC2 only
PC3	ES-CAViaR-MIDAS model with the third principal component PC3 only
PC1-3	ES-CAViaR-MIDAS model with the first three principal components
Lasso	ES-CAViaR-MIDAS model with selected variables
Combine1	Simple average combination forecasts using selected variables
Combine2	Loss-based combination forecasts using selected variables

Table C.5: Backtesting results for estimated risk measures (VaR, ES) of the S&P 500 index

	Panel A: Average loss (VaR)				Panel B: Hit proportion (VaR)				Panel C: DQ p-values (VaR)			
	1%	2.5%	5%	10%	1%	2.5%	5%	10%	1%	2.5%	5%	10%
Real GDP	0.036	0.075	0.125	0.201	1.49%	3.44%	5.60%	10.10%	0.088	0.052	0.567	0.296
IP	0.037	0.075	0.125	0.200	1.64%	3.37%	5.68%	9.98%	0.021	0.022	0.279	0.238
dunemp	0.036	0.076	0.125	0.201	1.56%	3.49%	5.51%	10.03%	0.057	0.049	0.633	0.406
dhous	0.036	0.074	0.124	0.201	1.59%	3.49%	5.51%	10.00%	0.024	0.118	0.134	0.323
deprof	0.036	0.076	0.126	0.200	1.52%	3.49%	5.75%	10.03%	0.091	0.007	0.050	0.274
drcons	0.036	0.075	0.125	0.201	1.59%	3.37%	5.65%	10.10%	0.039	0.060	0.735	0.250
CPI	0.036	0.075	0.125	0.200	1.59%	3.32%	5.44%	9.88%	0.057	0.071	0.609	0.285
PPI	0.036	0.076	0.125	0.200	1.61%	3.51%	5.70%	10.05%	0.026	0.027	0.482	0.310
CFNAI	0.037	0.075	0.125	0.200	1.71%	3.46%	5.80%	10.13%	0.014	0.095	0.380	0.198
new orders	0.036	0.076	0.125	0.201	1.59%	3.34%	5.60%	10.10%	0.031	0.023	0.142	0.185
monetary base	0.036	0.075	0.125	0.200	1.59%	3.44%	5.46%	10.08%	0.045	0.056	0.318	0.238
consumer sentiment	0.036	0.075	0.125	0.201	1.56%	3.46%	5.51%	10.05%	0.063	0.014	0.363	0.297
gdp_vol	0.036	0.075	0.125	0.200	1.56%	3.44%	5.72%	10.08%	0.066	0.094	0.623	0.223
cpi_vol	0.036	0.076	0.125	0.200	1.54%	3.37%	5.60%	10.03%	0.071	0.035	0.511	0.226
MKT	0.036	0.076	0.126	0.200	1.68%	3.44%	5.56%	9.93%	0.015	0.010	0.497	0.265
STR	0.036	0.075	0.125	0.200	1.52%	3.42%	5.46%	9.93%	0.070	0.065	0.580	0.265
Dspread	0.036	0.076	0.125	0.200	1.47%	3.34%	5.46%	9.93%	0.142	0.010	0.408	0.210
Tspread	0.036	0.076	0.125	0.201	1.54%	3.27%	5.39%	10.13%	0.049	0.035	0.053	0.186
RV	0.036	0.076	0.126	0.202	1.56%	3.22%	5.46%	9.91%	0.069	0.172	0.500	0.421
IV	0.037	0.075	0.125	0.201	1.68%	3.46%	5.82%	10.29%	0.016	0.019	0.046	0.332
None	0.036	0.076	0.126	0.201	1.71%	3.61%	5.77%	10.00%	0.014	0.009	0.049	0.274
All	0.038	0.082	0.132	0.201	1.73%	3.27%	6.13%	10.27%	0.017	0.045	0.118	0.480
PC1	0.037	0.077	0.126	0.202	1.59%	3.54%	5.58%	10.10%	0.053	0.009	0.113	0.457
PC2	0.036	0.076	0.125	0.200	1.59%	3.49%	5.53%	9.93%	0.037	0.020	0.438	0.260
PC3	0.036	0.076	0.126	0.201	1.54%	3.34%	5.65%	10.05%	0.076	0.050	0.373	0.374
PC1-3	0.037	0.076	0.126	0.201	1.64%	3.42%	5.60%	10.20%	0.026	0.030	0.254	0.305
Lasso	0.036	0.074	0.125	0.201	1.68%	3.15%	5.60%	10.13%	0.010	0.219	0.351	0.507
Combine1	0.036	0.075	0.124	0.201	1.64%	3.37%	5.58%	10.00%	0.012	0.045	0.184	0.212
Combine2	0.034	0.072	0.122	0.200	1.47%	3.03%	5.51%	10.03%	0.091	0.173	0.376	0.150

(Continued on the next page)

(Continued) Backtesting results for estimated risk measures (VaR, ES) of the S&P 500 index

	Panel D: Average loss (VaR, ES)			Panel E: Bootstrap p-values (ES)			Panel F: DES p-values (ES)					
	1%	2.5%	5%	10%	1%	2.5%	5%	10%	1%	2.5%	5%	10%
Real GDP	1.206	0.991	0.788	0.551	0.992	0.306	0.144	0.132	0.165	<i>0.089</i>	0.521	0.508
IP	1.208	0.995	0.785	0.550	0.996	0.503	0.325	0.144	<i>0.056</i>	0.037	0.306	0.525
dunemp	1.194	0.990	0.783	0.550	1.000	0.133	0.158	0.226	0.124	<i>0.072</i>	0.512	0.670
dhous	1.197	0.986	0.779	0.550	1.000	0.199	0.282	0.158	<i>0.070</i>	0.141	0.197	0.609
dcprof	1.199	1.016	0.806	0.550	0.998	0.115	0.150	<i>0.094</i>	0.161	0.010	0.040	0.515
drcons	1.196	0.989	0.786	0.551	0.994	0.308	0.195	0.152	<i>0.085</i>	<i>0.097</i>	0.670	0.516
CPI	1.206	0.998	0.791	0.550	0.998	0.231	0.166	0.118	0.105	0.120	0.640	0.626
PPI	1.194	1.001	0.791	0.549	1.000	0.106	<i>0.069</i>	0.236	<i>0.073</i>	0.047	0.397	0.594
CFNAI	1.201	0.977	0.778	0.550	0.999	0.690	0.372	<i>0.092</i>	0.043	0.152	0.455	0.387
new orders	1.192	0.995	0.786	0.553	0.999	0.225	0.335	0.166	<i>0.080</i>	0.036	0.171	0.522
monetary base	1.199	0.980	0.789	0.550	1.000	0.107	0.267	0.202	0.110	0.107	0.303	0.554
consumer sentiment	1.200	0.984	0.782	0.552	0.998	0.161	0.225	0.183	0.129	0.032	0.331	0.574
gdp_vol	1.203	0.994	0.788	0.551	0.997	0.157	0.131	0.180	0.117	0.148	0.617	0.499
cpi_vol	1.210	1.005	0.791	0.551	0.998	0.154	0.134	0.153	0.140	<i>0.055</i>	0.502	0.562
MKT	1.203	1.015	0.796	0.552	1.000	0.197	<i>0.074</i>	0.170	0.049	0.018	0.377	0.613
STR	1.198	0.992	0.787	0.551	0.999	0.110	0.135	0.172	0.150	0.105	0.575	0.631
Dspread	1.187	1.003	0.787	0.549	0.998	0.117	<i>0.085</i>	0.144	0.225	0.018	0.344	0.495
Tspread	1.189	0.992	0.788	0.554	0.999	0.198	0.372	0.238	0.136	<i>0.058</i>	0.132	0.389
RV	1.184	0.968	0.772	0.548	1.000	0.325	0.494	0.177	0.134	0.222	0.524	0.693
IV	1.194	0.970	0.775	0.542	1.000	0.698	0.389	0.024	0.046	0.039	<i>0.065</i>	0.348
None	1.217	1.004	0.794	0.553	1.000	0.334	0.332	0.048	0.037	0.022	<i>0.082</i>	0.542
All	1.228	1.053	0.842	0.549	1.000	0.844	0.333	<i>0.097</i>	0.045	0.017	0.049	0.544
PC1	1.212	0.998	0.797	0.553	0.989	0.428	0.118	<i>0.093</i>	0.103	0.017	0.131	0.589
PC2	1.200	0.996	0.790	0.550	0.999	0.179	0.136	<i>0.086</i>	<i>0.094</i>	0.041	0.381	0.574
PC3	1.200	0.998	0.794	0.554	0.984	0.251	<i>0.088</i>	0.189	0.127	<i>0.088</i>	0.293	0.618
PC1-3	1.218	0.991	0.795	0.553	0.971	0.294	0.127	0.258	<i>0.054</i>	<i>0.058</i>	0.249	0.467
Lasso	1.174	0.938	0.768	0.548	0.991	0.767	0.533	<i>0.054</i>	0.031	0.226	0.334	0.607
Combine1	1.189	0.966	0.772	0.547	1.000	0.430	0.576	0.041	0.100	<i>0.097</i>	0.252	0.421
Combine2	1.124	0.914	0.737	0.545	1.000	<i>0.083</i>	0.563	<i>0.062</i>	0.257	0.312	0.468	0.370

Note: Panel A and Panel D present the average loss using the quantile score function for VaR and the FZ0 loss function for (VaR, ES) with confidence levels: 1%, 2.5%, 5% and 10%. The first three lowest average loss in each column are highlighted in bold. Panel B show the hit proportion for the VaR forecasts. Panel C, E and F present p -values for the DQ regression test for VaR, bootstrap test and DES regression test for ES, respectively. p -values that are greater than 0.05 (indicating no evidence against optimality at the 0.1 level) are in bold, and values between 0.05 and 0.1 are in italics.

Table C.6: Out-of-sample performance rankings for various levels of α

	Average loss: (VaR, ES)					Average loss: (VaR, ES) with fixed w_1						
	1%	2.5%	5%	10%	Average	Rank	1%	2.5%	5%	10%	Average	Rank
Real GDP	22	13	15	19	17.3	18	16	11	16	13	14.0	12
IP	24	17	10	16	16.8	17	19	12	14	14	14.8	13
dunemp	10	11	9	12	10.5	8	8	10	9	15	10.5	9
dhous	12	9	7	11	9.8	6	5	7	8	10	7.5	7
deprof	15	28	28	14	21.3	22	28	29	24	18	24.8	28
drcons	11	10	11	20	13.0	11	3	8	10	8	7.3	6
CPI	23	22	20	13	19.5	21	11	20	21	20	18.0	16
PPI	8	23	22	6	14.8	14	17	21	22	19	19.8	21
CFNAI	19	6	6	10	10.3	7	6	4	6	6	5.5	5
new orders	7	18	12	27	16.0	15	22	13	15	9	14.8	14
moneyary base	14	7	18	9	12.0	9	25	17	17	25	21.0	22
consumer sentiment	17	8	8	22	13.8	12	13	9	7	7	9.0	8
gdp_vol	20	16	17	21	18.5	20	18	22	28	27	23.8	25
cpi_vol	25	26	21	18	22.5	24	10	28	19	17	18.5	19
MKT	21	27	26	23	24.3	27	27	26	23	21	24.3	26
STR	13	14	13	17	14.3	13	9	14	12	12	11.8	11
Dspread	4	24	14	8	12.5	10	20	18	11	11	15.0	15
Tspread	6	15	16	29	16.5	16	12	15	18	28	18.3	17
RV	3	4	3	4	3.5	3	7	5	3	2	4.3	4
IV	9	5	5	1	5.0	5	21	16	5	3	11.3	10
None	27	25	24	25	25.3	29	29	6	13	29	19.3	20
All	29	29	29	7	23.5	26	24	24	26	23	24.3	27
PC1	26	20	27	26	24.8	28	23	25	29	24	25.3	29
PC2	16	19	19	15	17.3	19	14	23	20	16	18.3	18
PC3	18	21	23	28	22.5	25	15	27	25	26	23.3	23
PC1-3	28	12	25	24	22.3	23	26	19	27	22	23.5	24
Lasso	2	2	2	5	2.8	2	2	2	2	4	2.5	2
Combine1	5	3	4	3	3.8	4	4	3	4	5	4.0	3
Combine2	1	1	1	2	1.3	1	1	1	1	1	1.0	1

Note: This table presents the rankings of out-of-sample performance based on the values of FZ loss function for four confidence levels. The best model ranks 1 while the worst ranks 29 given there are 29 competing models in total. Columns 1-4 and Columns 7-10 report the rankings for each confidence level. Columns 5-6 and Columns 11-12 present the average loss for four confidence levels and the rankings of the average loss respectively.

Table C.7: Backtesting results for estimated risk measures (VaR, ES) of the S&P 500 index with fixed w_1

	Panel A: Average loss (VaR)				Panel B: Hit proportion (VaR)				Panel C: DQ p-values (VaR)			
	1%	2.5%	5%	10%	1%	2.5%	5%	10%	1%	2.5%	5%	10%
Real GDP	0.036	0.075	0.125	0.201	1.71%	3.27%	5.39%	9.91%	0.013	0.073	0.552	0.269
IP	0.037	0.075	0.125	0.201	1.71%	3.37%	5.58%	10.00%	0.013	0.040	0.218	0.295
dunemp	0.036	0.076	0.125	0.203	1.66%	3.42%	5.60%	10.17%	0.027	0.035	0.474	0.808
dhous	0.036	0.074	0.124	0.200	1.56%	3.27%	5.56%	10.15%	0.056	0.074	0.341	0.147
deprof	0.037	0.076	0.125	0.200	1.76%	3.51%	5.89%	10.34%	0.012	0.018	0.083	0.334
drcons	0.035	0.075	0.125	0.200	1.64%	3.25%	5.29%	9.88%	0.017	0.085	0.662	0.344
CPI	0.036	0.076	0.125	0.200	1.59%	3.22%	5.29%	9.86%	0.049	0.092	0.752	0.225
PPI	0.036	0.075	0.125	0.200	1.56%	3.54%	5.65%	10.15%	0.045	0.015	0.375	0.360
CFNAI	0.036	0.074	0.125	0.200	1.52%	3.25%	5.32%	9.98%	0.064	0.092	0.500	0.241
new orders	0.036	0.075	0.125	0.200	1.68%	3.42%	5.63%	10.15%	0.013	0.035	0.189	0.187
monetary base	0.036	0.076	0.125	0.202	1.59%	3.15%	5.51%	9.96%	0.060	0.182	0.431	0.506
consumer sentiment	0.036	0.075	0.125	0.200	1.61%	3.42%	5.48%	10.05%	0.014	0.028	0.502	0.205
gdp_vol	0.036	0.076	0.126	0.201	1.52%	3.13%	5.36%	9.86%	0.078	0.135	0.394	0.250
cpi_vol	0.036	0.076	0.125	0.200	1.54%	3.39%	5.65%	10.03%	0.067	0.035	0.334	0.244
MKT	0.037	0.076	0.125	0.201	1.64%	3.42%	5.51%	9.98%	0.035	0.025	0.480	0.292
STR	0.036	0.075	0.125	0.200	1.54%	3.27%	5.34%	9.84%	0.049	0.066	0.565	0.366
Dspread	0.036	0.075	0.125	0.200	1.61%	3.37%	5.53%	9.96%	0.039	0.029	0.366	0.161
Tspread	0.036	0.075	0.125	0.201	1.54%	3.32%	5.46%	9.81%	0.046	0.049	0.023	0.126
RV	0.037	0.076	0.126	0.203	1.61%	3.13%	5.48%	9.81%	0.035	0.256	0.525	0.744
IV	0.036	0.076	0.126	0.202	1.59%	3.46%	5.96%	10.53%	0.038	0.028	0.044	0.287
All	0.044	0.078	0.128	0.258	1.30%	2.93%	5.34%	9.84%	0.009	0.552	0.478	0.004
None	0.036	0.076	0.126	0.201	1.71%	3.56%	5.77%	10.00%	0.014	0.012	0.049	0.274
PC1	0.037	0.077	0.126	0.202	1.59%	3.54%	5.58%	10.10%	0.053	0.009	0.113	0.457
PC2	0.036	0.076	0.125	0.200	1.59%	3.49%	5.53%	9.93%	0.037	0.020	0.438	0.260
PC3	0.036	0.076	0.126	0.201	1.54%	3.34%	5.65%	10.05%	0.076	0.050	0.373	0.374
PC1-3	0.037	0.076	0.126	0.201	1.64%	3.42%	5.60%	10.20%	0.026	0.030	0.254	0.305
Lasso	0.037	0.074	0.126	0.201	1.59%	3.13%	5.65%	10.49%	0.019	0.185	0.162	0.096
Combine1	0.036	0.075	0.125	0.200	1.54%	3.42%	5.44%	10.03%	0.038	0.030	0.144	0.150
Combine2	0.033	0.071	0.122	0.199	1.47%	3.10%	5.60%	10.08%	0.079	0.055	0.153	0.110

(Continued on the next page)

(Continued) Backtesting results for estimated risk measures (VaR, ES) of the S&P 500 index with fixed w_1

	Panel D: Average loss (VaR, ES)			Panel E: Bootstrap p-values (ES)			Panel F: DES p-values (ES)					
	1%	2.5%	5%	10%	1%	2.5%	5%	10%	1%	2.5%	5%	10%
Real GDP	1.201	0.984	0.787	0.549	1.000	0.164	0.109	0.150	0.045	0.120	0.512	0.633
IP	1.205	0.985	0.786	0.550	1.000	0.193	0.387	0.116	0.035	<i>0.067</i>	0.270	0.617
dunemp	1.193	0.982	0.782	0.550	1.000	0.152	0.218	0.158	<i>0.068</i>	<i>0.065</i>	0.406	0.798
dhous	1.183	0.979	0.780	0.548	1.000	0.229	0.115	<i>0.075</i>	0.109	0.111	0.298	0.438
dcprof	1.226	1.005	0.794	0.551	1.000	0.379	0.279	<i>0.080</i>	0.029	0.025	<i>0.089</i>	0.404
drcons	1.174	0.979	0.783	0.547	1.000	0.130	0.177	0.215	<i>0.073</i>	0.146	0.648	0.701
CPI	1.198	0.993	0.791	0.552	1.000	0.107	<i>0.080</i>	<i>0.084</i>	0.120	0.164	0.675	0.651
PPI	1.204	0.993	0.791	0.551	1.000	<i>0.080</i>	0.120	0.047	0.107	0.036	0.290	0.548
CFNAI	1.184	0.972	0.779	0.546	1.000	<i>0.081</i>	0.288	0.248	0.182	0.166	0.553	0.566
new orders	1.211	0.985	0.786	0.547	1.000	0.265	0.242	0.124	0.036	<i>0.052</i>	0.207	0.469
monetary base	1.217	0.987	0.788	0.553	1.000	0.248	0.218	<i>0.092</i>	0.119	0.233	0.399	0.771
consumer sentiment	1.199	0.981	0.779	0.546	0.997	0.148	<i>0.097</i>	0.125	0.044	<i>0.057</i>	0.441	0.475
gdp_vol	1.205	0.995	0.797	0.554	0.998	0.302	0.282	0.162	0.169	0.198	0.487	0.674
cpi_vol	1.196	0.999	0.789	0.551	0.997	0.288	0.163	<i>0.085</i>	0.132	<i>0.055</i>	0.286	0.557
MKT	1.220	0.998	0.793	0.553	0.984	0.333	<i>0.098</i>	<i>0.068</i>	<i>0.073</i>	0.044	0.399	0.585
STR	1.194	0.986	0.784	0.549	1.000	0.169	0.174	<i>0.085</i>	0.111	0.118	0.583	0.769
Dspread	1.205	0.989	0.784	0.548	0.993	0.165	0.211	<i>0.078</i>	<i>0.087</i>	<i>0.052</i>	0.360	0.479
Tspread	1.198	0.986	0.788	0.554	1.000	0.338	0.435	<i>0.071</i>	0.126	<i>0.086</i>	<i>0.096</i>	0.364
RV	1.185	0.973	0.772	0.545	1.000	0.674	0.506	0.038	<i>0.098</i>	0.270	0.537	0.931
IV	1.209	0.987	0.778	0.545	0.926	0.634	0.317	0.009	<i>0.057</i>	0.038	<i>0.057</i>	0.270
All	1.292	0.978	0.785	0.603	0.999	0.969	0.397	0.293	0.010	0.539	0.485	0.010
None	1.216	0.998	0.794	0.553	0.999	0.321	0.338	0.025	0.037	0.022	<i>0.082</i>	0.542
PC1	1.212	0.998	0.797	0.553	0.994	0.427	0.113	0.031	0.103	0.017	0.131	0.589
PC2	1.200	0.996	0.790	0.550	1.000	0.170	0.125	0.035	<i>0.094</i>	0.041	0.381	0.574
PC3	1.200	0.998	0.794	0.554	0.974	0.246	<i>0.088</i>	<i>0.076</i>	0.127	<i>0.088</i>	0.293	0.618
PC1-3	1.218	0.991	0.795	0.553	0.973	0.285	0.116	0.130	<i>0.054</i>	<i>0.058</i>	0.249	0.467
Lasso	1.157	0.940	0.770	0.545	0.997	0.524	0.371	0.023	0.032	0.231	0.153	0.120
Combine1	1.177	0.971	0.777	0.546	1.000	0.133	0.485	<i>0.087</i>	0.106	<i>0.060</i>	0.233	0.389
Combine2	1.081	0.907	0.740	0.532	1.000	0.004	0.895	0.255	0.304	0.130	0.313	0.315

Note: Panel A and Panel D present the average loss using the quantile score function for VaR and the FZ0 loss function for (VaR, ES) with confidence levels: 1%, 2.5%, 5% and 10%. The first three lowest average loss in each column are highlighted in bold. Panel B show the hit proportion for the VaR forecasts. Panel C, E and F present p -values for the DQ regression test for VaR, bootstrap test and DES regression test for ES, respectively. p -values that are greater than 0.05 (indicating no evidence against optimality at the 0.1 level) are in bold, and values between 0.05 and 0.1 are in italics.

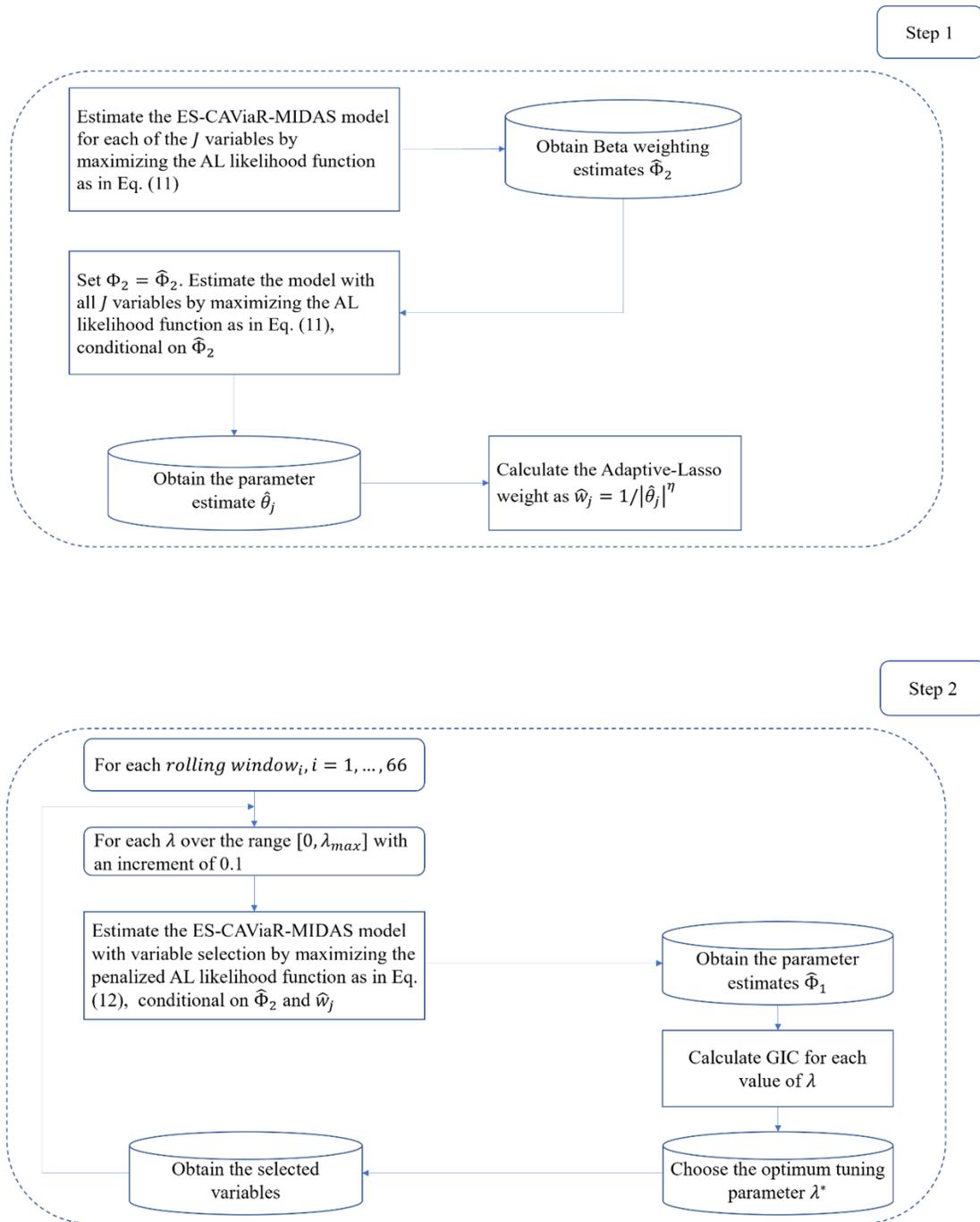


Figure C.1: Estimation framework

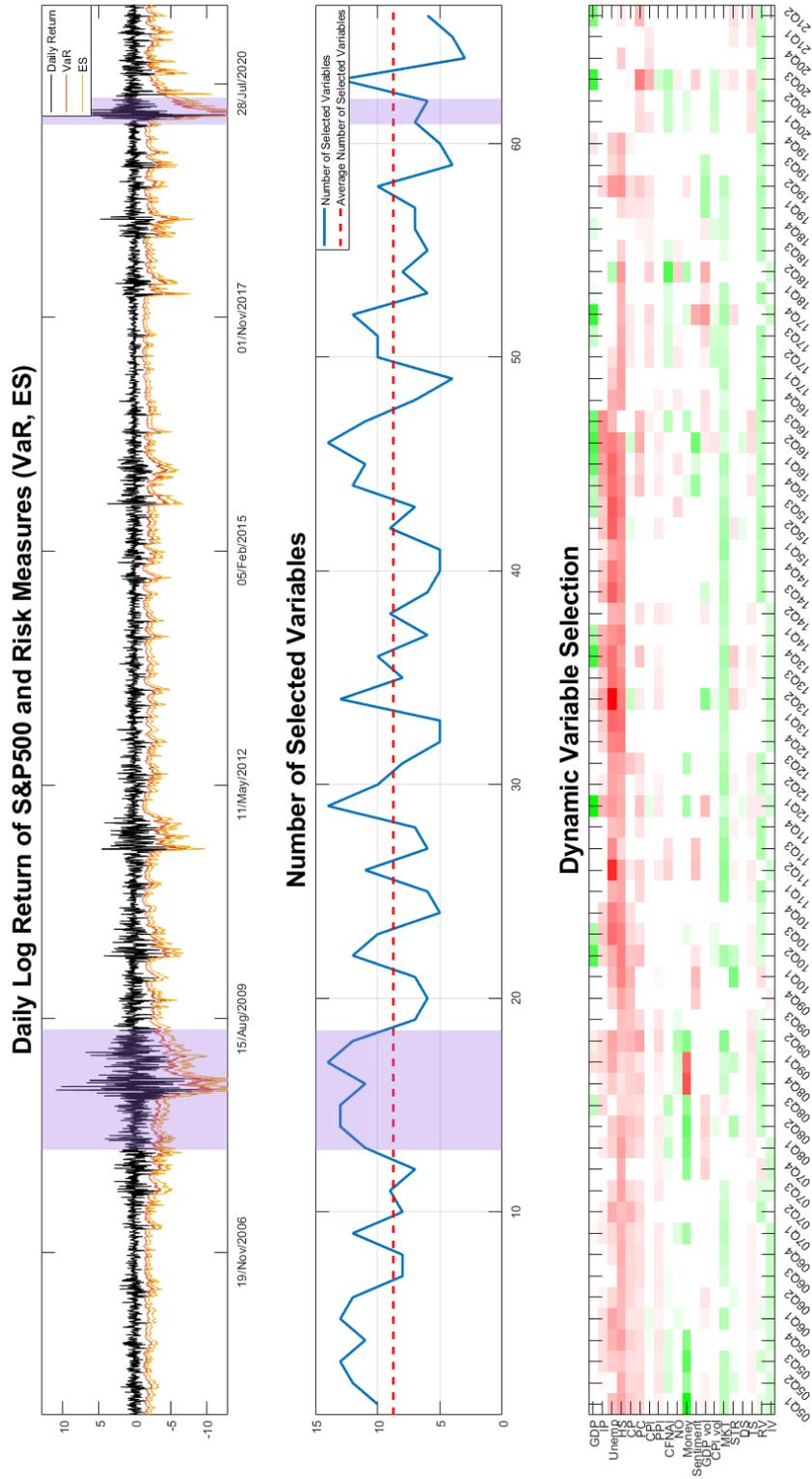


Figure C.2: Dynamic variable selection at $\alpha = 1\%$

Note: This figure shows the dynamic variable selection results at 1% confidence level. The color tab in this figure represents that the corresponding variable is selected by the Adaptive-Lasso. Green tab indicates a positive relationship between the corresponding variable and future tail risk while red tab suggests the opposite.

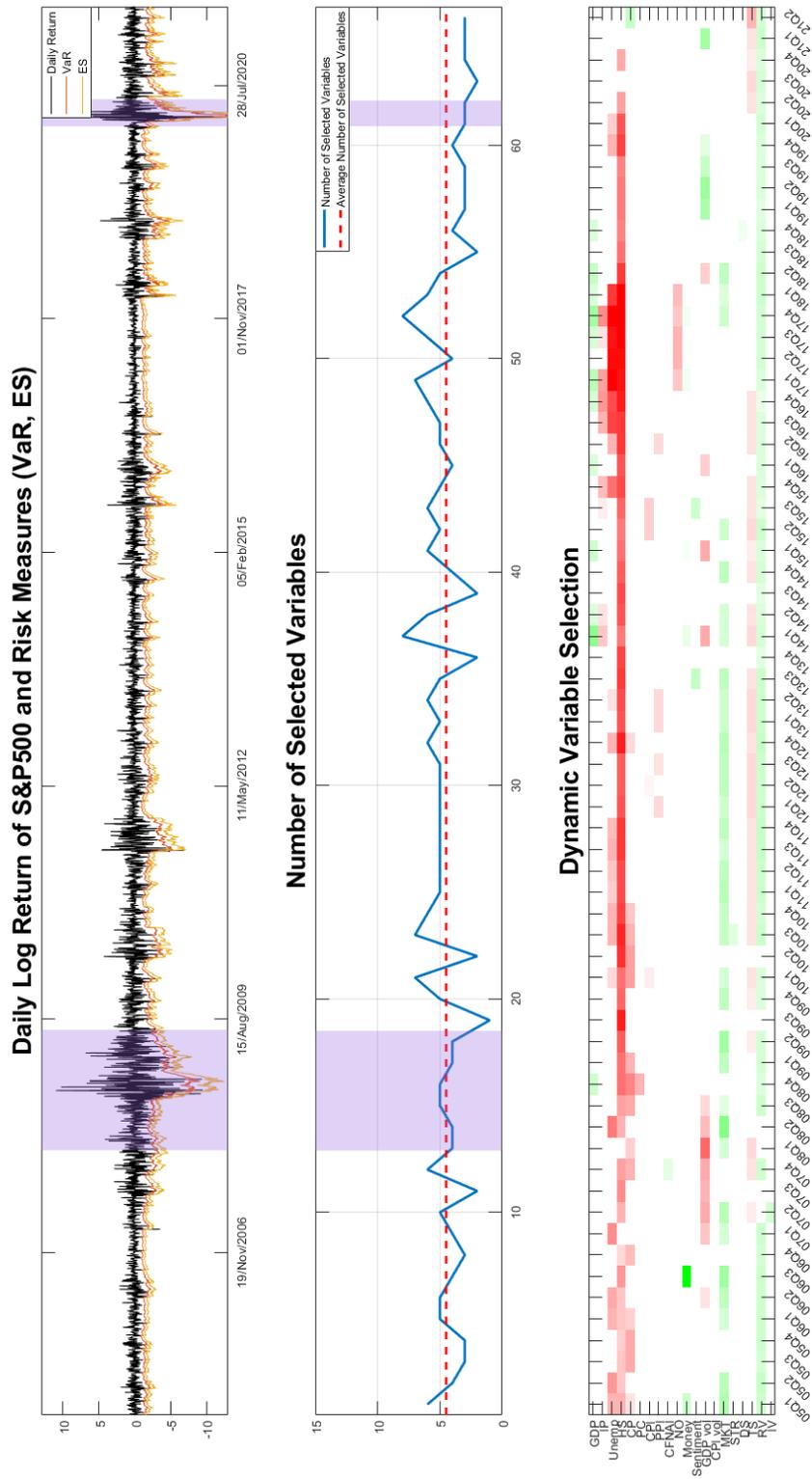


Figure C.3: Dynamic variable selection at $\alpha = 2.5\%$

Note: This figure shows the dynamic variable selection results at 2.5% confidence level. The color tab in this figure represents that the corresponding variable is selected by the Adaptive-Lasso. Green tab indicates a positive relationship between the corresponding variable and future tail risk while red tab suggests the opposite.

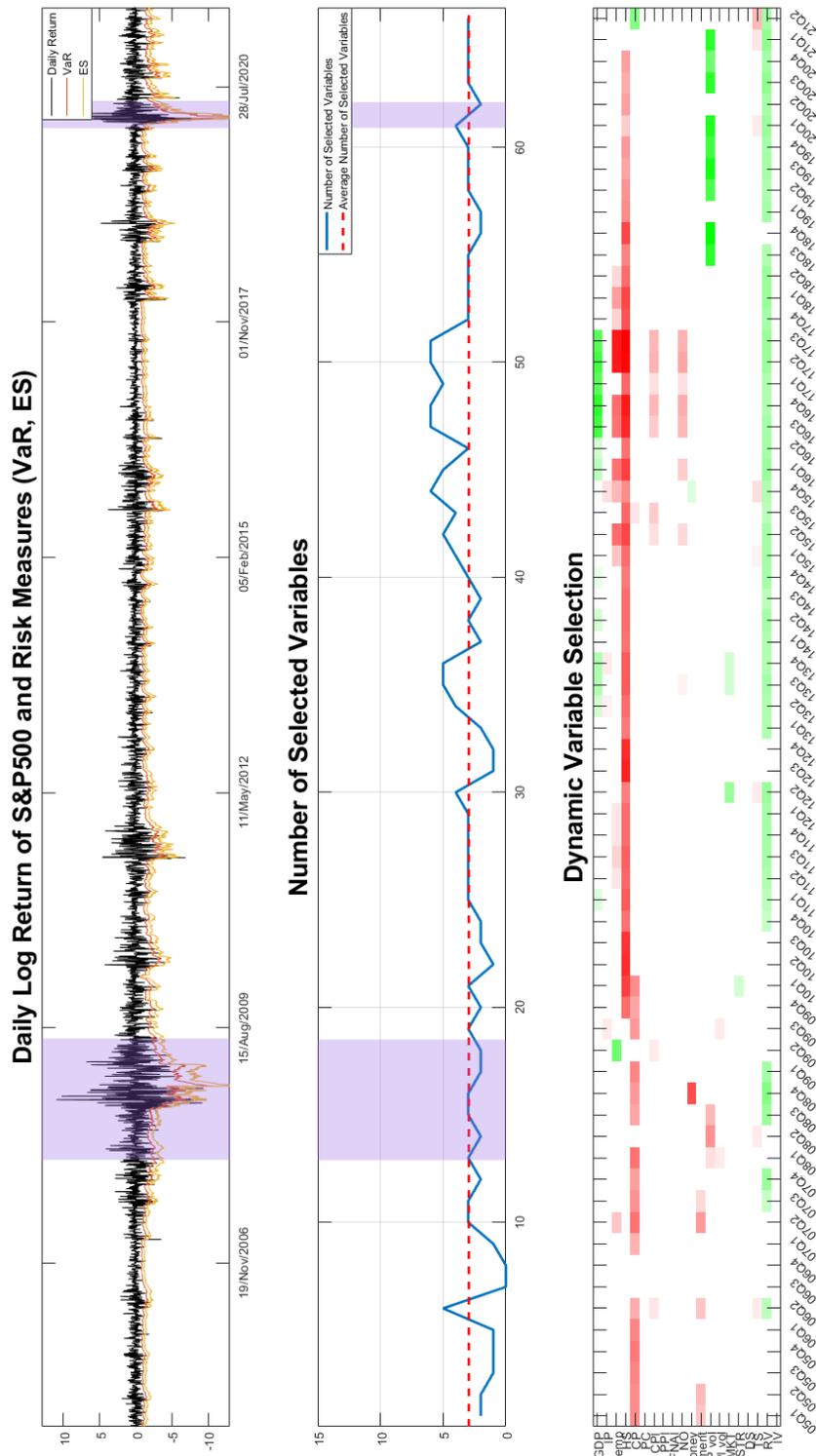


Figure C.4: Dynamic variable selection at $\alpha = 5\%$

Note: This figure shows the dynamic variable selection results at 5% confidence level. The color tab in this figure represents that the corresponding variable is selected by the Adaptive-Lasso. Green tab indicates a positive relationship between the corresponding variable and future tail risk while red tab suggests the opposite.

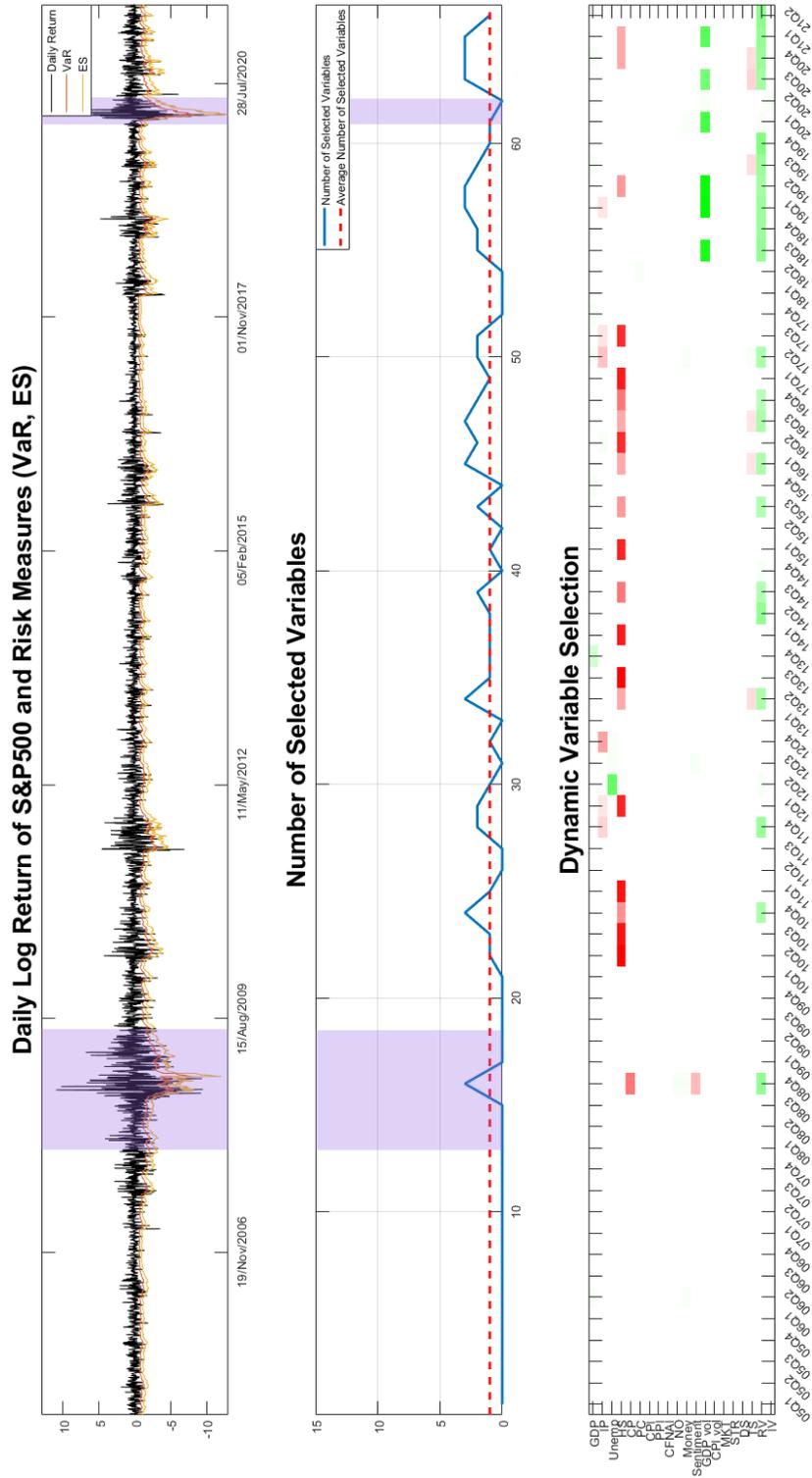


Figure C.5: Dynamic variable selection at $\alpha = 10\%$

Note: This figure shows the dynamic variable selection results at 10% confidence level. The color tab in this figure represents that the corresponding variable is selected by the Adaptive-Lasso. Green tab indicates a positive relationship between the corresponding variable and future tail risk while red tab suggests the opposite.

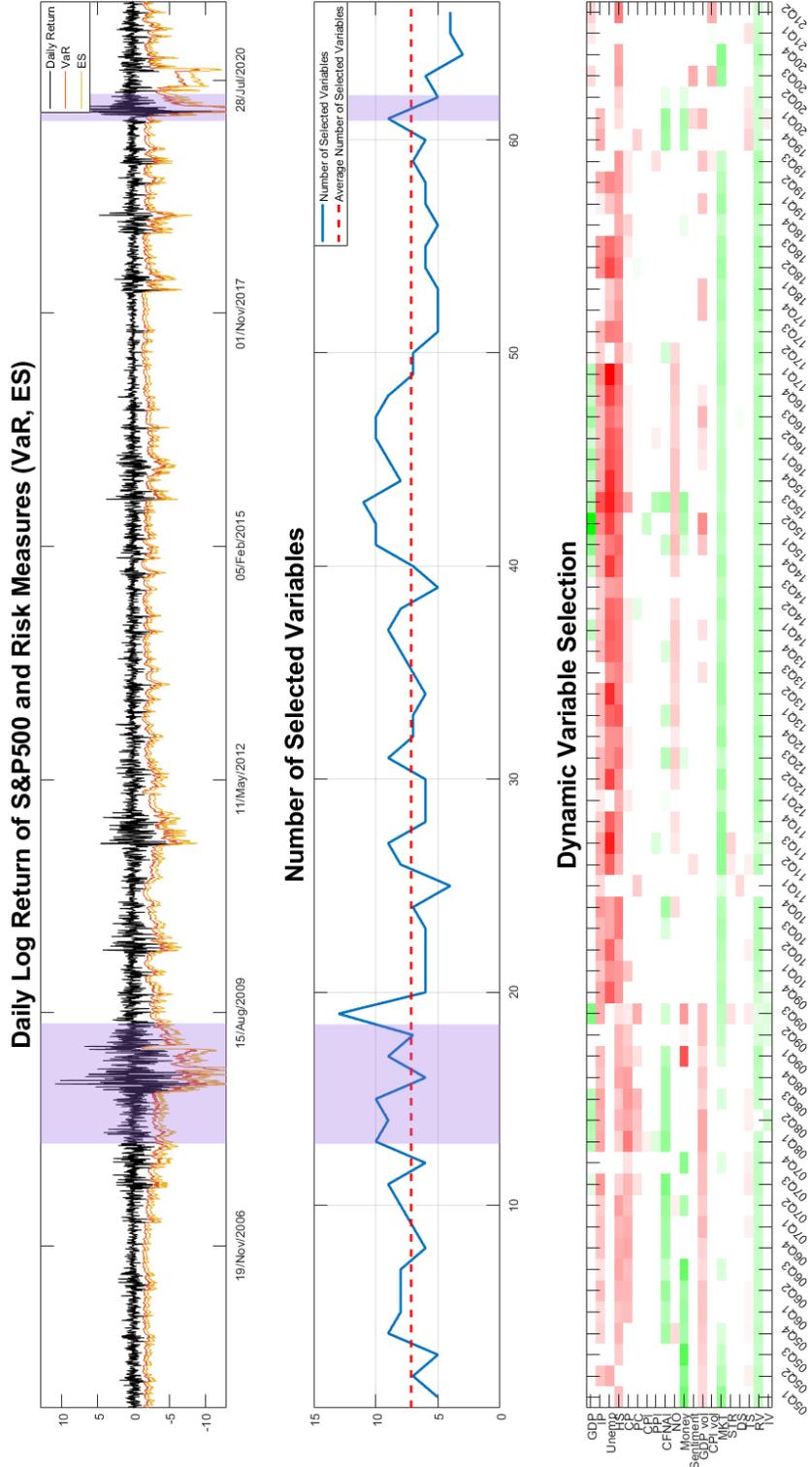


Figure C.6: Dynamic variable selection at $\alpha = 1\%$ with fixed w_1

Note: This figure shows the dynamic variable selection results under the restricted Beta weighting scheme at 1% confidence level. The color tab in this figure represents that the corresponding variable is selected by the Adaptive-Lasso. Green tab indicates a positive relationship between the corresponding variable and future tail risk while red tab suggests the opposite.

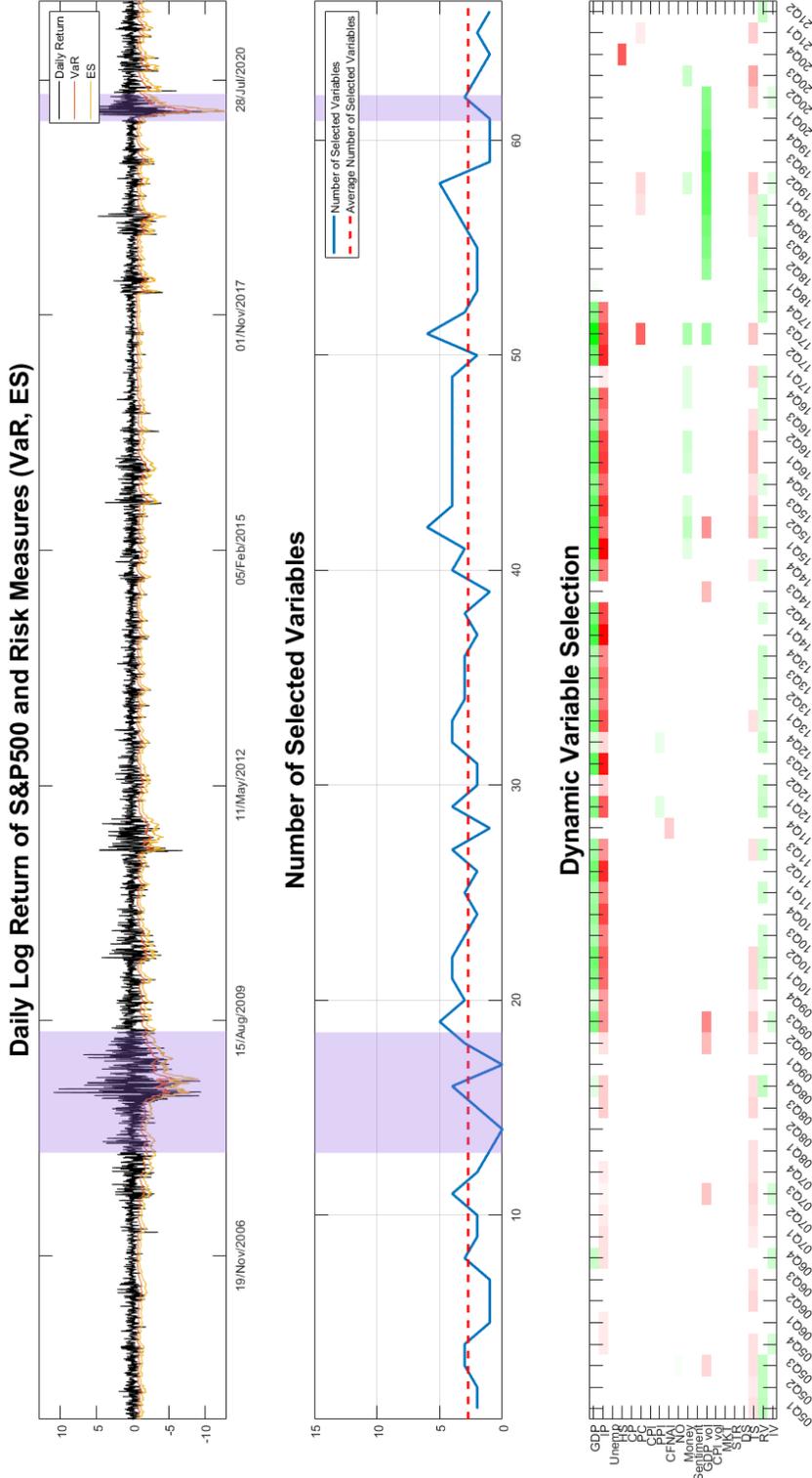


Figure C.9: Dynamic variable selection at $\alpha = 10\%$ with fixed w_1

Note: This figure shows the dynamic variable selection results under the restricted Beta weighting scheme at 10% confidence level. The color tab in this figure represents that the corresponding variable is selected by the Adaptive-Lasso. Green tab indicates a positive relationship between the corresponding variable and future tail risk while red tab suggests the opposite.

Chapter 4

Concluding Remarks and Further Developments

This thesis proposes new models and frameworks to forecast realized volatility and stock market tail risk by integrating Machine Learning algorithms with Financial Time Series models. More information can be exploited in the proposed hybrid framework than in the traditional Financial Time Series models. Moreover, the interpretability of the proposed hybrid models is enhanced by taking informative features generated by Financial Time Series models.

In Chapter 1, we examine the impact of periodicity on volatility-volume relation. It has been found in the empirical literature that periodicity is an important stylized fact of financial returns and failure to account this may distort the underlying volatility process. However, market activity, such as trading volume and number of trades, also displays prominent periodicity which is much less exploited. By joint modeling the periodicity factor in return and in trading volume, we investigate the volatility-volume relation in a more micro way. The mixture model results show that the trading volume is strongly correlated with volatility in early trading day while such correlation becomes weaker during the rest of the day. This indicates that the intraday volatility-volume relation may be time-varying and failure to account this might explain the contradictory results in volatility-volume literature. Moreover, by filtering periodicity from trading volume, both the explanatory power and forecasting accuracy of trading volume on return volatility are significantly improved under both linear regression

and Random Forest framework. This advocates the usage of filtered market activity measures instead of the raw measures. Furthermore, the forecasting performance of Random Forest is much better than linear regression model, which may suggest that the relationship between volatility and volume is highly nonlinear and requires high-dimensional Machine Learning algorithms to extract. Due to data availability, Chapter 1 is limited in investigating only 5-min data for three stocks. We could use more data frequency and more asset class to broaden the research.

In Chapter 2, we propose hybrid models integrating Neural Networks (NN) with Heterogeneous Autoregressive-type (HAR) models to forecast realized volatility. The proposed hybrid models contribute by overcoming the limitations of existing Semiparametric HAR-based hybrid models which attempts to model nonlinearity and by improving the realized volatility forecasting. The general form of Semiparametric HAR-based hybrid models is the original HAR components with nonlinearly transformed residuals. This approach implicitly assumes that residuals contain prominent nonlinearity that can be modeled by NN and violation of such assumption may lead to undesirable results. In addition, the simple additive combination in Semiparametric HAR-based hybrid models may underestimate the relationship between the linear and nonlinear components. Besides, the hyperparameters, such as number of hidden units and nonlinear transformation function, are subjectively chosen by the researcher applying Semiparametric HAR-based hybrid models which may be problematic. To overcome these limitations, we propose the Nonparametric NN-based hybrid models which takes various input features generated from HAR-type models. The proposed model approximates the input-output map through efficient learning and its superior performance is achieved by taking high-quality HAR-type components. All the hyperparameters in the proposed hybrid models are reasonably optimized through walk-forward validation. The empirical results show that NN-based hybrid models achieve the best out-of-sample forecasting performance in all subsamples across all forecasting horizons. The single NN models which only takes past realized volatility as input variables often underperform the standard HAR. However, such performance difference is significantly smaller in longer forecasting horizon (i.e., weekly). This indicates that daily realized volatility is highly persistent and HAR-type components are more informative features than past realized volatility under the NN architectures.

Chapter 2 results are limited to stocks which could be enriched by considering other asset class such as exchange rate.

In Chapter 3, a new framework for the joint estimation and forecasting of Value at Risk (VaR) and Expected Shortfall (ES) is proposed, which incorporates low-frequency macroeconomic and financial indicators into the quantile-based MIDAS model. We contribute by proposing a powerful tool utilizing effective low-frequency information in the high-frequency tail risk forecasting. The proposed framework can accommodate any amount of low-frequency variables without identification problem and the embedded variable screening process ensures only the effective low-frequency information can enter into the estimation. The empirical results show that three variables (namely, realized volatility, term spread and housing starts) are consistently selected for most of the rolling windows and serve to the strongest predictors of future tail risk. In addition to the strongest predictors, the value of other low-frequency variables should not be neglected as is revealed by the superior forecasting performance achieved by incorporating the selected variables on a rolling-window basis. Moreover, the average number of selected variables increases in predicting more extreme tail risk which indicates that more extreme VaR and ES may rely on additional information. The out-of-sample backtesting results show that our method passes most backtests with relatively higher p -values and achieves the minimum loss in the joint forecasting of VaR and ES. Chapter 3 is limited in examining the stock market tail risk and could be extended to banking data which focuses on the banks tail risk.

In the future, this thesis can be possibly further developed in the following ways. First, for Chapter 1, we may try to build rigorous statistical model and provides theoretical support for the joint modeling of periodicity factor in both return and trading volume. The existing Mixture of Distribution Hypothesis focuses on modeling the joint distribution of return and trading volume conditional on same information flow variable. This offers an intuitively appealing explanation for the strongly positive correlation between return volatility and trading volume. We may further develop this hypothesis in a more micro way by focusing on the common periodicity factor in both return and trading volume. Moreover, the study can be enriched by including different sampling frequency and other Machine Learning algorithms. Second, for Chapter 2, we may include Semiparametric HAR-based hybrid models as additional

benchmarks. Given the fundamental difference between the Semiparametric HAR-based hybrid models and our proposed Neural Network-based hybrid models, it is worth to examine their performance in more forecasting horizons and on different asset classes. Third, for Chapter 3, we may conduct simulations to justify the rationality of our proposed model. Two types of simulation will serve different purposes. The first simulation focuses on the quality of proposed data generating process by simulating data according to proposed process and estimating the simulated data. If our proposed model is reasonable, it should sufficiently recover the true parameters. The second simulation attempts to examine the power of proposed variable screening process. Ideally, it should correctly select important variables from the simulated data. Moreover, we may consider constructing portfolios with lower tail risk by weighting assets according to their individual tail risk estimated by the proposed methodology.

In conclusion, this thesis focuses on integrating Machine Learning algorithms with Financial Time Series models. We find strong empirical evidence that the Machine Learning algorithms well complement Financial Time Series models in complex relationship detection, variable selection and nonlinearity modeling. Financial Time Series models, however, may provide high-quality input features for Machine Learning algorithms and enhance performance of those algorithms. The thesis has rich development potential which will be realized in the future.

References

- Admati, A. R. and P. Pfleiderer (1988). “A theory of intraday patterns: Volume and price variability”. *The review of financial studies*, 1(1), pp. 3–40.
- Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002). “Range-based estimation of stochastic volatility models”. *Journal of Finance*, 57(3), pp. 1047–1091.
- Andersen, T. G. (1996). “Return volatility and trading volume: An information flow interpretation of stochastic volatility”. *The Journal of Finance*, 51(1), pp. 169–204.
- Andersen, T. G. and T. Bollerslev (1997). “Intraday periodicity and volatility persistence in financial markets”. *Journal of empirical finance*, 4(2-3), pp. 115–158.
- Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007). “Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility”. *The review of economics and statistics*, 89(4), pp. 701–720.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). “The distribution of realized stock return volatility”. *Journal of financial economics*, 61(1), pp. 43–76.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). “Modeling and forecasting realized volatility”. *Econometrica*, 71(2), pp. 579–625.
- Andersen, T. G. and T. Bollerslev (1998). “Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts”. *International Economic Review*, 39(4), pp. 885–905.
- Andersen, T. G., T. Bollerslev, and N. Meddahi (2011). “Realized volatility forecasting and market microstructure noise”. *Journal of Econometrics*, 160(1), pp. 220–234.

- Arnerić, J., T. Poklepović, and Z. Aljinović (2014). “GARCH based artificial neural networks in forecasting conditional variance of stock returns”. *Croatian Operational Research Review*, pp. 329–343.
- Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). “Coherent measures of risk”. *Mathematical finance*, 9(3), pp. 203–228.
- Asgharian, H., F. Javed, and A. J. Hou (2013). “The importance of the macroeconomic variables in forecasting stock return variance: A GARCH-MIDAS approach”. *Journal of Forecasting*, 32(7), pp. 600–612.
- Ausin, M. C. and P. Galeano (2007). “Bayesian estimation of the Gaussian mixture GARCH model”. *Computational Statistics & Data Analysis*, 51(5), pp. 2636–2652.
- Avramov, D., T. Chordia, and A. Goyal (2006). “The impact of trades on daily volatility”. *The Review of Financial Studies*, 19(4), pp. 1241–1277.
- Barclay, M. J. and J. B. Warner (1993). “Stealth trading and volatility: Which trades move prices?” *Journal of financial Economics*, 34(3), pp. 281–305.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). “Econometric analysis of realized volatility and its use in estimating stochastic volatility models”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), pp. 253–280.
- Basel Committee on Banking Supervision (2019). *Minimum capital requirements for market risk*. Available at <https://www.bis.org/bcbs/publ/d457.pdf>. Bank for International Settlements. URL: <http://https://www.bis.org/bcbs/publ/d457.pdf>.
- Becker, R., A. Clements, and A. McClelland (2009). “The jump component of S&P 500 volatility and the VIX index”. *Journal of Banking and Finance*, 33(6), pp. 1033–1038.
- Belgiu, M. and L. Drăguț (2016). “Random forest in remote sensing: A review of applications and future directions”. *ISPRS journal of photogrammetry and remote sensing*, 114, pp. 24–31.

- Black, F. (1976). “Studies in stock price volatility changes”. In: *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.
- Bollerslev, T. (1986). “Generalized autoregressive conditional heteroskedasticity”. *Journal of econometrics*, 31(3), pp. 307–327.
- Boudt, K., C. Croux, and S. Laurent (2011). “Robust estimation of intraweek periodicity in volatility and jump detection”. *Journal of Empirical Finance*, 18(2), pp. 353–367.
- Breiman, L. (2001). “Random forests”. *Machine learning*, 45(1), pp. 5–32.
- Bucci, A. (2020). “Realized volatility forecasting with neural networks”. *Journal of Financial Econometrics*, 18(3), pp. 502–531.
- Candila, V., G. M. Gallo, and L. Petrella (2020). “Using mixed-frequency and realized measures in quantile regression”. Available at SSRN 3722927.
- Chan, C. C. and W. M. Fong (2006). “Realized volatility and transactions”. *Journal of Banking & Finance*, 30(7), pp. 2063–2085.
- Chan, K. and W.-M. Fong (2000). “Trade size, order imbalance, and the volatility–volume relation”. *Journal of Financial Economics*, 57(2), pp. 247–273.
- Chen, J.-F., W.-L. Chen, C.-P. Huang, S.-H. Huang, and A.-P. Chen (2016). “Financial time-series data analysis using deep convolutional neural networks”. In: *2016 7th International conference on cloud computing and big data (CCBD)*. IEEE, pp. 87–92.
- Chen, L., M. Pelger, and J. Zhu (2020). “Deep learning in asset pricing”. Available at SSRN 3350138.
- Chen, W., H. Xu, L. Jia, and Y. Gao (2021). “Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants”. *International Journal of Forecasting*, 37(1), pp. 28–43.
- Christiansen, C., M. Schmeling, and A. Schrimpf (2012). “A comprehensive look at financial volatility prediction by economic variables”. *Journal of Applied Econometrics*, 27(6), pp. 956–977.

- Clark, P. K. (1973). “A subordinated stochastic process model with finite variance for speculative prices”. *Econometrica: journal of the Econometric Society*, pp. 135–155.
- Clements, M. P., A. B. Galvão, and J. H. Kim (2008). “Quantile forecasts of daily exchange rate returns from forecasts of realized volatility”. *Journal of Empirical Finance*, 15(4), pp. 729–750.
- Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990). “STL: A seasonal-trend decomposition”. *J. Off. Stat*, 6(1), pp. 3–73.
- Conrad, C. and O. Kleen (2019). “Two are better than one: Volatility forecasting using multiplicative component GARCH-MIDAS models”. *Journal of Applied Econometrics*, 35(1), pp. 19–45.
- Conrad, C. and K. Loch (2015). “Anticipating long-term stock market volatility”. *Journal of Applied Econometrics*, 30(7), pp. 1090–1114.
- Conrad, C., K. Loch, and R. Daniel (2014). “On the macroeconomic determinants of long-term volatilities and correlations in U.S. stock and crude oil markets”. *Journal of Empirical Finance*, 29, pp. 26–40.
- Corsi, F. (2009). “A simple approximate long-memory model of realized volatility”. *Journal of Financial Econometrics*, 7, pp. 174–196. DOI: <http://dx.doi.org/10.1016/j.jeconom.2010.07.008>..
- Corsi, F., D. Pirino, and R. Reno (2010). “Threshold bipower variation and the impact of jumps on volatility forecasting”. *Journal of Econometrics*, 159(2), pp. 276–288.
- Creamer, G. G. and Y. Freund (2004). “Predicting performance and quantifying corporate governance risk for latin american adrs and banks”. *Financial Engineering and Applications, MIT, Cambridge*.
- Crouch, R. L. (1970). “The volume of transactions and price changes on the New York Stock Exchange”. *Financial Analysts Journal*, 26(4), pp. 104–109.
- Croushore, D. (2011). “Frontiers of real-time data analysis”. *Journal of Economic Literature*, 49(1), pp. 72–100.

- Dash, P. K., M. Nayak, M. R. Senapati, and I. W. Lee (2007). “Mining for similarities in time series data using wavelet-based feature vectors and neural networks”. *Engineering Applications of Artificial Intelligence*, 20(2), pp. 185–201.
- Demirer, R., K. Gkillas, R. Gupta, and C. Pierdzioch (2021). “Risk aversion and the predictability of crude oil market volatility: A forecasting experiment with random forests”. *Journal of the Operational Research Society*, pp. 1–13.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). “Maximum likelihood from incomplete data via the EM algorithm”. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), pp. 1–22.
- Doering, J., M. Fairbank, and S. Markose (2017). “Convolutional neural networks applied to high-frequency market microstructure forecasting”. In: *2017 9th computer science and electronic engineering (ceec)*. IEEE, pp. 31–36.
- Donaldson, R. G. and M. Kamstra (1997). “An artificial neural network-GARCH model for international stock return volatility”. *Journal of Empirical Finance*, 4(1), pp. 17–46.
- Du, Z. and J. C. Escanciano (2017). “Backtesting expected shortfall: Accounting for tail risk”. *Management Science*, 63(4), pp. 940–958.
- Duda, R. O., P. E. Hart, and D. G. Stork (1973). *Pattern classification and scene analysis*. Vol. 3. Wiley New York.
- Easley, D., N. M. Kiefer, and M. O’Hara (1997). “The information content of the trading process”. *Journal of Empirical Finance*, 4(2-3), pp. 159–186.
- Engle, R. F. (1982). “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation”. *Econometrica: Journal of the econometric society*, pp. 987–1007.
- Engle, R. F., E. Ghysels, and B. Sohn (2013). “Stock market volatility and macroeconomic fundamentals”. *Review of Economics and Statistics*, 95(3), pp. 776–797.
- Engle, R. F. and S. Manganelli (2004). “CAViaR: Conditional autoregressive value at risk by regression quantiles”. *Journal of Business & Economic Statistics*, 22(4), pp. 367–381.

- Epps, T. W. and M. L. Epps (1976). “The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis”. *Econometrica: Journal of the Econometric Society*, pp. 305–321.
- Estrella, A. and M. Trubin (2006). “The yield curve as a leading indicator: Some practical issues”. *Current issues in Economics and Finance*, 12(5).
- Fama, E. F. and K. R. French (1992). “The cross-section of expected stock returns”. *Journal of Finance*, 47(2), pp. 427–465.
- Fan, Y. and C. Y. Tang (2013). “Tuning parameter selection in high dimensional penalized likelihood”. *Journal of the Royal Statistical Society: SERIES B: Statistical Methodology*, pp. 531–552.
- Fang, T., T.-H. Lee, and Z. Su (2020). “Predicting the long-term stock market volatility: A GARCH-MIDAS model with variable selection”. *Journal of Empirical Finance*, 58, pp. 36–49.
- Faraway, J. and C. Chatfield (1998). “Time series forecasting with neural networks: a comparative study using the air line data”. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(2), pp. 231–250.
- Fernandes, D., J. G. Lynch Jr, and R. G. Netemeyer (2014). “Financial literacy, financial education and downstream financial behaviors”. *Management Science*, 60(8), pp. 1861–2109.
- Fernandes, M., M. C. Medeiros, and M. Scharth (2014). “Modeling and predicting the CBOE market volatility index”. *Journal of Banking & Finance*, 40, pp. 1–10.
- Fissler, T. and J. F. Ziegel (2016). “Higher order elicibility and Osband’s principle”. *Annals of Statistics*, 44(4), pp. 1680–1707.
- Fotopoulos, S. B. (2017). “Symmetric Gaussian mixture distributions with GGC scales”. *Journal of Multivariate Analysis*, 160, pp. 185–194.
- Frey, R. and A. J. McNeil (2002). “VaR and expected shortfall in portfolios of dependent credit risks: conceptual and practical insights”. *Journal of banking & finance*, 26(7), pp. 1317–1334.

- Froot, K. A., D. S. Scharfstein, and J. C. Stein (1993). “Risk management: Coordinating corporate investment and financing policies”. *the Journal of Finance*, 48(5), pp. 1629–1658.
- Gerlach, R. and C. Wang (2020). “Semi-parametric dynamic asymmetric Laplace models for tail risk forecasting, incorporating realized measures”. *International Journal of Forecasting*, 36(2), pp. 489–506.
- Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). “The MIDAS touch: Mixed data sampling regression models”. preprint on webpage at <https://rady.ucsd.edu/faculty/directory/valkanov/pub/docs/midas-touch.pdf>.
- Giacomini, R. and I. Komunjer (2005). “Evaluation and combination of conditional quantile forecasts”. *Journal of Business & Economic Statistics*, 23(4), pp. 416–431.
- Giot, P. and S. Laurent (2004). “Modelling daily value-at-risk using realized volatility and ARCH type models”. *Journal of Empirical Finance*, 11(3), pp. 379–398.
- Giot, P., S. Laurent, and M. Petitjean (2010). “Trading activity, realized volatility and jumps”. *Journal of Empirical Finance*, 17(1), pp. 168–175.
- Glosten, L. R. and L. E. Harris (1988). “Estimating the components of the bid/ask spread”. *Journal of financial Economics*, 21(1), pp. 123–142.
- Glosten, L. R. and P. R. Milgrom (1985). “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders”. *Journal of financial economics*, 14(1), pp. 71–100.
- Gneiting, T. (2011). “Making and evaluating point forecasts”. *Journal of the American Statistical Association*, 106(494), pp. 746–762.
- Gradojevic, N., D. Kukulj, R. Adcock, and V. Djakovic (2021). “Forecasting Bitcoin with technical analysis: A not-so-random forest?” *International Journal of Forecasting*.
- Granger, C. W. and O. Morgenstern (1963). “Spectral analysis of New York stock market prices 1”. *Kyklos*, 16(1), pp. 1–27.
- Gu, S., B. Kelly, and D. Xiu (2020). “Empirical asset pricing via machine learning”. *The Review of Financial Studies*, 33(5), pp. 2223–2273.

- Hajizadeh, E., A. Seifi, M. F. Zarandi, and I. Turksen (2012). “A hybrid modeling approach for forecasting the volatility of S&P 500 index return”. *Expert Systems with Applications*, 39(1), pp. 431–436.
- Hamid, S. A. and Z. Iqbal (2004). “Using neural networks for forecasting volatility of S&P 500 Index futures prices”. *Journal of Business Research*, 57(10), pp. 1116–1125.
- Hansen, P. R., Z. Huang, and H. H. Shek (2012). “Realized garch: a joint model for returns and realized measures of volatility”. *Journal of Applied Econometrics*, 27(6), pp. 877–906.
- Happersberger, D., H. Lohre, and I. Nolte (2020). “Estimating portfolio risk for tail risk protection strategies”. *European Financial Management*, 26(4), pp. 1107–1146.
- Harris, L. (1989). “S&P 500 cash stock price volatilities”. *The Journal of Finance*, 44(5), pp. 1155–1175.
- Hasbrouck, J. (1999). “The dynamics of discrete bid and ask quotes”. *The Journal of finance*, 54(6), pp. 2109–2142.
- Hewamalage, H., C. Bergmeir, and K. Bandara (2021). “Recurrent neural networks for time series forecasting: Current status and future directions”. *International Journal of Forecasting*, 37(1), pp. 388–427.
- Hillebrand, E. and M. C. Medeiros (2010). “The benefits of bagging for forecast models of realized volatility”. *Econometric Reviews*, 29(5-6), pp. 571–593.
- Hochreiter, S. and J. Schmidhuber (1997). “Long short-term memory”. *Neural computation*, 9(8), pp. 1735–1780.
- Horvath, B., A. Muguruza, and M. Tomas (2021). “Deep learning volatility: a deep neural network perspective on pricing and calibration in (rough) volatility models”. *Quantitative Finance*, 21(1), pp. 11–27.
- Huang, R. D. and R. W. Masulis (2003). “Trading activity and stock price volatility: evidence from the London Stock Exchange”. *Journal of Empirical Finance*, 10(3), pp. 249–269.
- Izzeldin, M. (2007). “Trading volume and the number of trades: a comparative study using high frequency data”.

- Jolliffe, I. T. and J. Cadima (2016). “Principal component analysis: a review and recent developments”. *Philosophical Transactions of the Royal Society A*, 374(2065). ISSN: 1364-503X. DOI: doi.org/10.1098/rsta.2015.0202.
- Jones, C. M., G. Kaul, and M. L. Lipson (1994). “Transactions, volume, and volatility”. *The Review of Financial Studies*, 7(4), pp. 631–651.
- Khan, M. A. I. (2011). “Financial volatility forecasting by nonlinear support vector machine heterogeneous autoregressive model: evidence from Nikkei 225 stock index”. *International Journal of Economics and Finance*, 3(4), p. 138.
- Kim, H. Y. and C. H. Won (2018). “Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models”. *Expert Systems with Applications*, 103, pp. 25–37.
- Kristjanpoller, W., A. Fadic, and M. C. Minutolo (2014). “Volatility forecast using hybrid neural network models”. *Expert Systems with Applications*, 41(5), pp. 2437–2442.
- Kupiec, P. H. (1995). “Techniques for Verifying the Accuracy of Risk Measurement Models”. *Journal of Derivatives*, 3(2), pp. 73–84. ISSN: 1074-1240. DOI: [10.3905/jod.1995.407942](https://doi.org/10.3905/jod.1995.407942).
- Kydland, F. E., P. Rupert, and R. Sustek (2012). “Housing dynamics over the business cycle”. *Departmental working papers. Department of Economics, UCSB, UC Santa Barbara*.
- Kyle, A. S. (1985). “Continuous auctions and insider trading”. *Econometrica: Journal of the Econometric Society*, pp. 1315–1335.
- Lai, Y., C. W. Chen, and R. Gerlach (2009). “Optimal dynamic hedging via copula-threshold-GARCH models”. *Mathematics and Computers in Simulation*, 79(8), pp. 2609–2624.
- Lamoureux, C. G. and W. D. Lastrapes (1990). “Heteroskedasticity in stock return data: Volume versus GARCH effects”. *The journal of finance*, 45(1), pp. 221–229.

- Lazar, E. and X. Xue (2020). “Forecasting risk measures using intraday data in a generalized autoregressive score framework”. *International Journal of Forecasting*, 36(3), pp. 1057–1072.
- Le, T. H. (2020). “Forecasting value at risk and expected shortfall with mixed data sampling”. *International Journal of Forecasting*, 36(4), pp. 1362–1379.
- Leamer, E. E. (2007). “Housing is the business cycle”. *National Bureau of Economic Research Cambridge, Mass., USA*.
- Liesenfeld, R. (2001). “A generalized bivariate mixture model for stock price volatility and trading volume”. *Journal of econometrics*, 104(1), pp. 141–178.
- Liljeblom, E. and M. Stenius (1997). “Macroeconomic volatility and stock market volatility: Empirical evidence on Finnish data”. *Applied Financial Economics*, 7(4), pp. 419–426.
- Louzis, D. P., S. Xanthopoulos-Sisinis, and A. P. Refenes (2014). “Realized volatility models and alternative Value-at-Risk prediction strategies”. *Economic Modelling*, 40, pp. 101–116.
- Luu, J. C. and M. Martens (2003). “Testing the mixture-of-distributions hypothesis using “realized” volatility”. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 23(7), pp. 661–679.
- Maciel, L., F. Gomide, and R. Ballini (2016). “Evolving fuzzy-GARCH approach for financial volatility modeling and forecasting”. *Computational Economics*, 48(3), pp. 379–398.
- Maheu, J. M. and T. H. McCurdy (2002). “Nonlinear features of realized FX volatility”. *Review of Economics and Statistics*, 84(4), pp. 668–681.
- Manela, A. and A. Moreira (2017). “News implied volatility and disaster concerns”. *Journal of Financial Economics*, 123(1), pp. 137–162.
- Martens, M. and J. Zein (2004). “Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility”. *Journal of Futures Markets*, 24(11), pp. 1005–1028.

- Massacci, D. (2017). “Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness”. *Management Science*, 63(9), pp. 2773–3145.
- McAleer, M. and M. C. Medeiros (2008). “A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries”. *Journal of Econometrics*, 147(1), pp. 104–119.
- McInish, T. H. and R. A. Wood (1985). “Intraday and Overnight Returns and Day-of-the-Week Effects”. *Journal of Financial Research*, 8(2), pp. 119–126.
- McNeil, A. J. and R. Frey (2000). “Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach”. *Journal of Empirical Finance*, 7(3-4), pp. 271–300.
- Mehlitz, J. S. and B. R. Auer (2021). “Time-varying dynamics of expected shortfall in commodity futures markets”. *Journal of Futures Markets*, 41(6), pp. 895–925.
- Meng, X. and J. W. Taylor (2020). “Estimating value-at-risk and expected shortfall using the intraday low and range data”. *European Journal of Operational Research*, 280(1), pp. 191–202.
- Morgan, I. G. (1976). “Stock prices and heteroscedasticity”. *The Journal of business*, 49(4), pp. 496–508.
- Müller, U. A., M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz, and C. Morgengegg (1990). “Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis”. *Journal of Banking & Finance*, 14(6), pp. 1189–1208.
- Nagel, S. (2012). “Evaporating liquidity”. *Review of Financial Studies*, 25(7), pp. 2005–2039.
- Nanopoulos, A., R. Alcock, and Y. Manolopoulos (2001). “Feature-based classification of time-series data”. *International Journal of Computer Research*, 10(3), pp. 49–61.
- Nieto, M. R. and E. Ruiz (2016). “Frontiers in VaR forecasting and backtesting”. *International Journal of Forecasting*, 32(2), pp. 475–501.
- Nikolaev, N., P. Tino, and E. Smirnov (2013). “Time-dependent series variance learning with recurrent mixture density networks”. *Neurocomputing*, 122, pp. 501–512.

- Nonejad, N. (2017). “Modeling and forecasting aggregate stock market volatility in unstable environments using mixture innovation regressions”. *Journal of Forecasting*, 26(6), pp. 718–740.
- Osborne, M. F. (1959). “Brownian motion in the stock market”. *Operations research*, 7(2), pp. 145–173.
- Pan, Z., Y. Wang, C. Wu, and L. Yin (2017). “Oil price volatility and macroeconomic fundamentals: A regime switching GARCH-MIDAS model”. *Journal of Empirical Finance*, 43, pp. 130–142.
- Patton, A. J., J. F. Ziegel, and R. Chen (2019). “Dynamic semiparametric models for expected shortfall (and value-at-risk)”. *Journal of Econometrics*, 211(2), pp. 388–413.
- Pavlidis, E. G., I. Paya, and D. A. Peel (2012). “Forecast Evaluation of Nonlinear Models: The Case of Long-Span Real Exchange Rates”. *Journal of Forecasting*, 31(7), pp. 580–595.
- Paye, B. S. (2012). “‘Déjà vol’: Predictive regressions for aggregate stock market volatility using macroeconomic variables”. *Journal of Financial Economics*, 106(3), pp. 527–546.
- Petneházi, G. (2019). “Recurrent neural networks for time series forecasting”. *arXiv preprint arXiv:1901.00069*.
- Psaradellis, I. and G. Sermpinis (2016). “Modelling and trading the US implied volatility indices. Evidence from the VIX, VXN and VXD indices”. *International Journal of Forecasting*, 32(4), pp. 1268–1283.
- Roh, T. H. (2007). “Forecasting the volatility of stock price index”. *Expert Systems with Applications*, 33(4), pp. 916–922.
- Rousseeuw, P. J. and A. M. Leroy (1988). “A robust scale estimator based on the shortest half”. *Statistica Neerlandica*, 42(2), pp. 103–116.
- Rutledge, D. J. (1979). *Trading volume and price variability: New evidence on the price effects of speculation*.

- Schittenkopf, C., G. Dorffner, and E. J. Dockner (2000). “Forecasting time-dependent conditional densities: a semi non-parametric neural network approach”. *Journal of Forecasting*, 19(4), pp. 355–374.
- Schuster, M. and K. K. Paliwal (1997). “Bidirectional recurrent neural networks”. *IEEE transactions on Signal Processing*, 45(11), pp. 2673–2681.
- Schwert, G. W. (1989). “Why does stock market volatility change over time?” *The Journal of Finance*, 44(5), pp. 1115–1153.
- Şener, E., S. Baronyan, and L. A. Mengütürk (2012). “Ranking the predictive performances of value-at-risk estimation methods”. *International Journal of Forecasting*, 28, pp. 849–873.
- Shahzad, H., H. N. Duong, P. S. Kalev, and H. Singh (2014). “Trading volume, realized volatility and jumps in the Australian stock market”. *Journal of International Financial Markets, Institutions and Money*, 31, pp. 414–430.
- Sharkey, A. J. (2002). “Types of multinet system”. In: *International Workshop on Multiple Classifier Systems*. Springer, pp. 108–117.
- Sheela, K. G. and S. N. Deepa (2013). “Review on methods to fix number of hidden neurons in neural networks”. *Mathematical problems in engineering*, 2013.
- Sjöberg, J., Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorennec, H. Hjalmarsson, and A. Juditsky (1995). “Nonlinear black-box modeling in system identification: a unified overview”. *Automatica*, 31(12), pp. 1691–1724.
- Stark, T. (2010). “Realistic evaluation of real-time forecasts in the survey of professional forecasters”. *Technical Report Rap Special Report, Federal Reserve Bank of Philadelphia*.
- Stinchcombe, M. and H. White (1992). “Using feedforward networks to distinguish multivariate populations”. In: *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*. Vol. 1. IEEE, pp. 788–793.
- Su, Z., T. Fang, and L. Yin (2017). “The role of news-based implied volatility among US financial markets”. *Economics Letters*, 157, pp. 24–27.

- Taskaya-Temizel, T. and M. C. Casey (2005). “A comparative study of autoregressive neural network hybrids”. *Neural Networks*, 18(5-6), pp. 781–789.
- Tauchen, G. E. and M. Pitts (1983). “The price variability-volume relationship on speculative markets”. *Econometrica: Journal of the Econometric Society*, pp. 485–505.
- Taylor, J. W. (2008). “Estimating value at risk and expected shortfall using expectiles”. *Journal of Financial Econometrics*, 6(2), pp. 231–252.
- (2019). “Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution”. *Journal of Business & Economic Statistics*, 37(1), pp. 121–133.
- Tino, P., C. Schittenkopf, and G. Dorffner (2001). “Financial volatility trading using recurrent neural networks”. *IEEE transactions on neural networks*, 12(4), pp. 865–874.
- Tseng, C.-H., S.-T. Cheng, Y.-H. Wang, and J.-T. Peng (2008). “Artificial neural network model of the hybrid EGARCH volatility of the Taiwan stock index option prices”. *Physica A: Statistical Mechanics and its Applications*, 387(13), pp. 3192–3200.
- Vrontos, S. D., J. Galakis, and I. D. Vrontos (2021). “Implied volatility directional forecasting: a machine learning approach”. *Quantitative Finance*, 21(10), pp. 1687–1706.
- Wang, T. and Z. Huang (2012). “The relationship between volatility and trading volume in the Chinese stock market: a volatility decomposition perspective”. *Annals of economics and finance*, 13(1), pp. 211–236.
- Wheelock, D. C., M. E. Wohar, et al. (2009). “Can the term spread predict output growth and recessions? A survey of the literature”. *Federal Reserve Bank of St. Louis Review*, 91(5 Part 1), pp. 419–440.
- Xiong, R., E. P. Nichols, and Y. Shen (2015). “Deep learning stock volatility with google domestic trends”. *arXiv preprint arXiv:1512.04916*.

- Xu, Y., X. Wang, and H. Liu (2021). “Quantile-based GARCH-MIDAS: Estimating value-at-risk using mixed-frequency information”. *Finance Research Letters*, 43, p. 101965.
- Yamai, Y. and T. Yoshihara (2005). “Value-at-risk versus expected shortfall: A practical perspective”. *Journal of Banking and Finance*, 29(4), pp. 997–1015.
- Ying, C. C. (1966). “Stock market prices and volumes of sales”. *Econometrica: Journal of the Econometric Society*, pp. 676–685.
- Zakoian, J.-M. (1994). “Threshold heteroskedastic models”. *Journal of Economic Dynamics and Control*, 18(5), pp. 931–955.
- Zhang, G. P. and M. Qi (2005). “Neural network forecasting for seasonal and trend time series”. *European journal of operational research*, 160(2), pp. 501–514.
- Zhang, G. P. and V. L. Berardi (2001). “Time series forecasting with neural network ensembles: an application for exchange rate prediction”. *Journal of the operational research society*, 52(6), pp. 652–664.
- Zheng, Y., Q. Liu, E. Chen, Y. Ge, and J. L. Zhao (2016). “Exploiting multi-channels deep convolutional neural networks for multivariate time series classification”. *Frontiers of Computer Science*, 10(1), pp. 96–112.
- Zhou, Y.-L., R.-J. Han, Q. Xu, Q.-J. Jiang, and W.-K. Zhang (2019). “Long short-term memory networks for CSI300 volatility prediction with Baidu search volume”. *Concurrency and Computation: Practice and Experience*, 31(10), e4721.
- Zhu, D. and J. W. Galbraith (2011). “Modeling and forecasting expected shortfall with the generalized asymmetric student-t and asymmetric exponential power distributions”. *Journal of Empirical Finance*, 18(4), pp. 765–778.
- Zou, H. (2006). “The adaptive lasso and its oracle properties”. *Journal of the American Statistical Association*, 101(476), pp. 1418–1429.