PRICE ANALYSIS AND RISK MANAGEMENT IN CRYPTOCURRENCY MARKET

A Dissertation

by

ZE SHEN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	David J. Leatham
Committee Members,	Hwagyun Kim
	Ximing Wu
	Henry Bryant
Head of Department,	David J. Leatham

August 2021

Major Subject: Agribusiness and Managerial Economics

ABSTRACT

Cryptocurrency and its underlying technology, blockchain, are significantly changing the financial world. One of the critical characteristics of cryptocurrency is its substantial price fluctuations, which means investment risk. This dissertation explores the cryptocurrency market by analyzing its price formation and risk management. By investigating the inter-market relationship among multiple cryptocurrencies, it is found that cryptocurrency's price affects each other in the long run and short run, and the competition in the market matters in price formation. I forecast bitcoin's volatility and examine the cryptocurrency index (CRIX) tail behavior to address the cryptocurrency risk management issue. I adopt the Value at Risk and Expected Shortfall as risk measurements, which measure the risk exposure in the cryptocurrency market. The results of bitcoin volatility forecasting provide evidence that the RNN outperforms GARCH and EWMA in average forecasting performance. However, it is less efficient in capturing the bitcoin market's extreme events. Moreover, the RNN shows poor performance in Value at Risk forecasting, indicating that it could not work well as the econometric models in explaining extreme volatility. By investigating the cryptocurrency index tail behavior, I discover that focusing on tail events using extreme value theory and filtered historical simulation methods yields more accurate Value at Risk and Expected Shortfall estimation. This in turn significantly improves the efficiency of cryptocurrency index risk management. This dissertation is motivated by the fact that there is an increasingly important role cryptocurrency plays in the financial market but that there is relatively little existing empirical literature on cryptocurrency. This study provides reliable quantitative cryptocurrency market investigations to help fill this gap.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Leatham, and my committee members, Dr. Wu, Dr. Kim, and Dr. Bryant, for their guidance and support throughout the course of this research. Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my mother and father for their encouragement and to my husband for his patience and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor David J. Leatham, Professor Ximing Wu and Professor Henry Bryant of the Department of Agricultural Economics and Professor Hwagyun Kim of the Department of Finance.

All other work conducted for the thesis (or) dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a graduate assistantship from Texas A&M University.

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGMENTS iii

CONTRIBUTORS AND FUNDING SOURCES iv

TABLE OF CONTENTS v

LIST OF FIGURES vii

LIST OF TABLES x

1. INTRODUCTION AND LITERATURE REVIEW 1

    1.1 Blockchain technology 1
    1.2 What is Cryptocurrency? 4
    1.3 Cryptocurrency's price and risk 5

2. CRYPTOCURRENCIES PRICES AFFECT EACH OTHER 10

    2.1 Introduction 10
    2.2 Methodology 12
        2.2.1 Cointegration 12
        2.2.2 Vector error correction model (VECM) 12
    2.3 Data and empirical results 13
        2.3.1 Unit root test 14
        2.3.2 Long run equilibrium 14
        2.3.3 Short run equilibrium 15
        2.3.4 Granger causality analysis 16
        2.3.5 Impulse response function 18
    2.4 Conclusion 21

3. BITCOIN RETURN VOLATILITY FORECASTING: A COMPARATIVE STUDY BETWEEN GARCH AND RNN 23

    3.1 Introduction 23
    3.2 Materials and Methods 27
        3.2.1 Econometric methodology 28
            3.2.1.1 Conditional mean 29
            3.2.1.2 Conditional variance 30

3.2.1.3 Exponentially weighted moving average (EWMA) 30
3.2.1.4 GARCH model 30
3.2.2 Recurrent neural network (RNN) 31
3.3 Results 31
3.3.1 Forecasting 32
3.3.2 Volatility proxies 33
3.3.3 Out-of-sample performance 35
3.3.4 Value at Risk (VaR) 37
3.4 Conclusion 41

4. ESTIMATION OF TAIL RISK MEASURES IN CRYPTOCURRENCY MARKET 43

4.1 Introduction 43
4.2 Methodology 45
4.2.1 VaR and ES approach to risk measurement 45
4.2.2 Conditional mean and conditional variance 46
4.2.3 Parametric approach 47
4.2.4 Semiparametric approach 49
4.3 Data and empirical results 50
4.3.1 ARMA-GARCH process 50
4.3.2 EVT-POT 50
4.3.3 Filtered historical simulation (FHS) 57
4.3.4 Unconditional EVT 57
4.3.5 Value at Risk (VaR) and Expected Shortfall (ES) 59
4.4 Backtesting 68
4.5 Conclusion 72

5. CONCLUSION 74

REFERENCES 75

APPENDIX 79

LIST OF FIGURES

FIGURE Page

1.1 Traditional Ledger 2

1.2 How Blockchain Works 3

1.3 Broadcasting 3

2.1 Response of each variable to bitcoin price innovation (btc) 19

2.2 Response of each variable to Ethereum price innovation (eth) 19

2.3 Response of each variable to XRP price innovation (xrp) 20

2.4 Response of each variable to Litecoin price innovation (ltc) 20

3.1 Bitcoin daily return and bitcoin daily squared return. (a) Bitcoin daily return; (b) Bitcoin squared daily return. 28

3.2 Architecture of recurrent neural network (RNN) model. 32

3.3 Detailed implementation of recurrent neural network (RNN) model. 34

3.4 Out of sample standard deviation of realized volatility (1st proxy) vs standard deviation of one day ahead volatility forecasting of EWMA, GARCH and RNN model. EWMA1, GARCH1 and RNN1 denote the one day ahead forecasting of each model. 36

3.5 Out of sample standard deviation of realized volatility (2nd proxy) vs standard deviation of one day ahead volatility forecasting of EWMA, GARCH and RNN model. EWMA1, GARCH1 and RNN1 denote the one day ahead forecasting of each model. 37

3.6 Out of sample realized return vs one day ahead VaR(1%) of EWMA, GARCH and RNN model. EWMA1, GARCH1 and RNN1 denote the one day ahead forecasting of each model. 39

4.1 Plot of in sample CRIX daily loss. 51

4.2 Estimated sample mean excess. 52

4.3 Estimated excesses vs empirical excess (GARCH-EVT-Gaussian). 53

4.4 Estimated excesses vs empirical excess (GARCH-EVT-Student t). 54

4.5 Estimated excesses vs empirical excess (GARCH-EVT-Generalized Error). 54

4.6 Estimated tail probabilities vs empirical tail probabilities (GARCH-EVT-Gaussian). 55

4.7 Estimated tail probabilities vs empirical tail probabilities (GARCH-EVT-Student t). 56

4.8 Estimated tail probabilities vs empirical tail probabilities (GARCH-EVT-Generalized Error). 56

4.9 Estimated excesses vs empirical excess (Unconditional EVT). 58

4.10 Estimated tail probabilities vs empirical tail probabilities (Unconditional EVT). 59

4.11 One day ahead ES (EVT-Gaussian). 60

4.12 One day ahead ES (EVT-Student t)). 60

4.13 One day ahead ES (EVT-Generalized error). 61

4.14 One day ahead ES (EVT-u (0.05)). 62

4.15 One day ahead ES (EVT-u (0.10)). 62

4.16 One day ahead ES (EVT-u (0.15)). 63

4.17 ES-EVT vs ES-non-EVT (Gaussian, u (0.10)). 64

4.18 ES-EVT VS ES-non-EVT (Student t, u(0.10)). 64

4.19 ES-EVT VS ES-non-EVT (Generalized error, u(0.10)). 65

4.20 ES-EVT vs ES-FHS (Gaussian). 66

4.21 ES-EVT vs ES-FHS (Student t). 66

4.22 ES-EVT vs ES-FHS (Generalized error). 67

4.23 Conditional EVT vs Unconditional EVT. 68

1 Out of sample standard deviation of realized volatility (1st proxy) vs standard deviation of 5 days ahead volatility forecasting of EWMA, GARCH and RNN model . 80

2 Out of sample standard deviation of realized volatility (2nd proxy) vs standard deviation of 5 days ahead volatility forecasting of EWMA, GARCH and RNN model 80

3 Out of sample standard deviation of realized volatility (1st proxy) vs standard deviation of 10 days ahead volatility forecasting of EWMA, GARCH and RNN model 81

4 Out of sample standard deviation of realized volatility (2nd proxy) vs standard deviation of 10 days ahead volatility forecasting of EWMA, GARCH and RNN model 81

5 Out of sample realized return vs 5 days ahead VaR(1%) of EWMA, GARCH and RNN model 81

6 Out of sample realized return vs 10 days ahead VaR(1%) of EWMA, GARCH and RNN model 82

7 Out of sample realized return vs 1 day ahead VaR(2.5%) of EWMA, GARCH and RNN model 82

8 Out of sample realized return vs 5 days ahead VaR(2.5%) of EWMA, GARCH and RNN model 82

9 Out of sample realized return vs 10 days ahead VaR(2.5%) of EWMA, GARCH and RNN model 83

10 Out of sample realized return vs 1 day ahead VaR(5%) of EWMA, GARCH and RNN model 83

11 Out of sample realized return vs 5 days ahead VaR(5%) of EWMA, GARCH and RNN model 83

12 Out of sample realized return vs 10 days ahead VaR(5%) of EWMA, GARCH and RNN model 84

LIST OF TABLES

TABLE Page

2.1 Summary statistics for daily closing price from 1/3/2017 to 1/1/2020 13

2.2 Unit root test. 14

2.3 Johansen trace cointegration test. 14

2.4 The estimated vector error correction model coefficients. 17

2.5 Granger causality test. 18

3.1 Summary Statistics of Bitcoin Daily Returns in the Sample Period. 28

3.2 Unit root test. 28

3.3 Ljung-Box Q-Test for Bitcoin Daily Return. 29

3.4 ARMA (2,2)-GARCH (1,2) estimated parameters. 33

3.5 Out-of-sample Performance. 35

3.6 Out of Sample Coverage of Each Model. 40

3.7 VaR backtesting results of unconditional coverage test of Kupiec and conditional coverage test of Christoffersen -p value. 41

4.1 Parameter estimates for ARMA (2,1)-GARCH (1,2) model. 51

4.2 Estimated EVT-POT Parameters. 53

4.3 FHS quantile estimation. 57

4.4 Estimated Unconditional EVT-POT Parameters. 57

4.5 Backtesting resuts on violation ratio, unconditional coverage test of Kupiec (1995) and conditional coverage test of Christoffersen (1998)-p value. 70

4.6 Backtesting results p-value on expected shortfall test 72

1. INTRODUCTION AND LITERATURE REVIEW

The cryptocurrency market is an emerging market that has been attracting significant attention in the last decade. Its market capitalization reached $1,739 billion as of February 2021, with a total of 8,708 types of cryptocurrencies trading in the market. Cryptocurrency is an exciting phenomenon in the financial market. People from various areas are excited and curious about it, including investors, researchers, and even policymakers. They tried to understand this novel and mysterious market from different views. As an economics researcher, before we investigate the cryptocurrency market from economics and financial perspective, we need to address an important question: what is cryptocurrency? To figure out what cryptocurrency is, we first have to understand the technology behind it – blockchain.

1.1 Blockchain technology

Blockchain is the critical technique of cryptocurrency. In October 2008, a pseudonymous author or team named Satoshi Nakamoto introduced blockchain technology as a part of a bitcoin proposal to the world. Blockchain is a digitized, decentralized, public ledger of cryptocurrency transactions and is open to anyone. It records transactions between two parties in a verifiable and permanent way. Blockchain is also a decentralized database and keeps continuously updated digital records. Let us take a look at how blockchain works. Suppose there are four people in the world: Amy, Charli, Mike, and Jack. Amy wants to send $50 to Charli, Mike wants to send $3 to Amy and $7 to Charli, and Jack needs to send $20 to Mike. The traditional ledger works in the way as shown in Figure 1.1.

Amy trusts the bank and its ledger in this traditional payment way. However, as Amy starts to trust her friends and the banking system less and less, a digitized, decentralized system of verification and public ledger accessible to everyone is necessary. In Figure 1.2, Amy wants to send $50 to Charlie, and the transaction is represented online as a "block." Blockchain is a chain of blocks that contain information. The block contains information that we call data. Each block

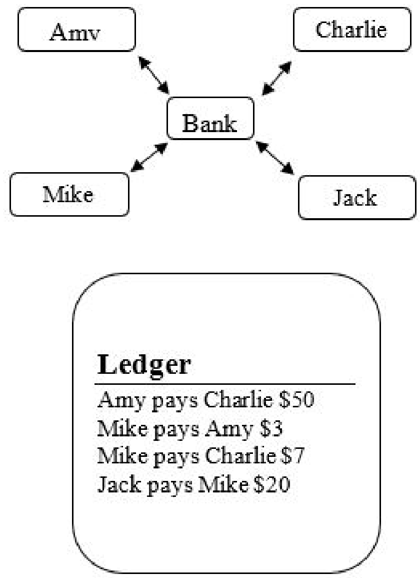


Figure 1.1: Traditional Ledger

contains data about transactions and the data of previous blocks. For example, block 61 in Figure 2 contains both data “Mike sent Amy $3” and the data in block 60. All these data inside the block can be calculated as a hash. The hash concept is that, given some input data, the computer creates an output of a fixed length. In our example, the input data of block 61 are “Mike pays Amy $3”, and together with the information of block 60. The output is a string of random letters and numbers, which is known as a digital fingerprint. Each block has a hash, and each hash identifies a block. Because each block contains the information about the previous block, if someone wants to change the information inside a block, then the hash of the next block will be changed too; otherwise, the changed block will be kicked out of the blockchain. Thus, anyone who wants to change one block has to change all the blocks after it. This type of mechanism makes it very difficult to change the information inside any block and secure a blockchain.

Once a block is created, it is broadcasted to every node in the network. In our case, as shown in Figure 1.3, Amy, Charlie, Mike, and Jack received the transaction information at the same time. Once those in the network approve that the transaction is valid, the block can be added to the chain. Meanwhile, the transaction is permanently and transparently recorded.

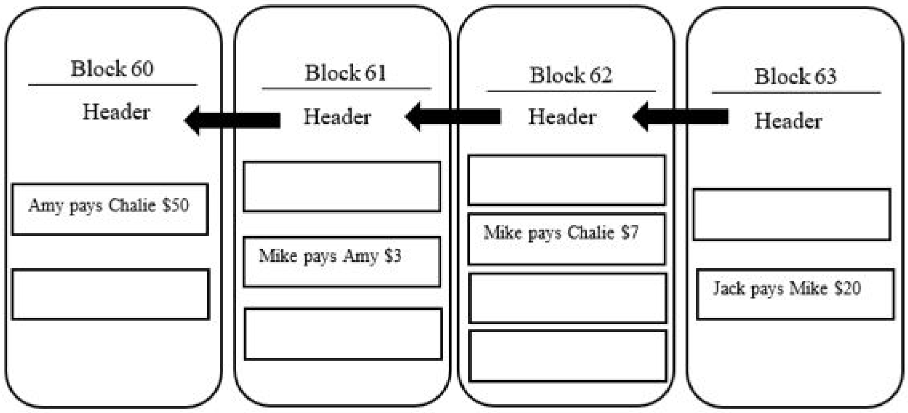


Figure 1.2: How Blockchain Works

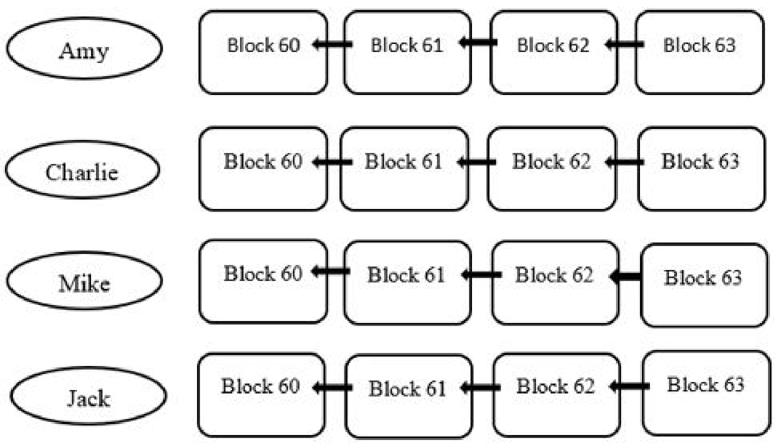


Figure 1.3: Broadcasting

The most significant advantage of blockchain is decentralization. It enables the database to be open source and share the information with the public without a central administrator. No government, no organization, or nobody can take control of the system. People do not have to depend on a third party for the transaction to take place. Let us recall the financial crisis in 2009, where some people lost trust in the banks. Blockchain technology gives us hope to solve this trust issue. Each transaction record is irreversible and transparent; thus, the information is reliable. Another benefit blockchain brings the low transaction cost. The transaction recording system and all kinds of documents required in the current financial service industry make business costly and slow. By using blockchain technology, financial services would be more efficient and faster.

1.2 What is Cryptocurrency?

Cryptocurrency is the application of blockchain technology. The first type of cryptocurrency is bitcoin, which was introduced in 2009 by Satoshi Nakamoto. As we mentioned in the blockchain section, each block is a record of transactions. People transfer the new transactions information into a block and add the new block to the old ones to become blockchain. However, who does this work? Blockchain is decentralized, which means there is no authority to "hire" employees to create blocks. Therefore, the system has to provide incentives to encourage public people to do this work. The incentive system is quite complex, but in short, the incentive is to reward anyone who can add new blocks to the efficient blockchain. This process of creating new blocks is called "mining," the people who do this work is called "miner," and the payoff is called "bitcoin." The mining process is updating the blockchain with verified bitcoin transactions, which any participant can do so. However, only the miner who first solves a difficult mathematical problem known as proof of work will win the chance to add the new block and get a mining reward. Bitcoin is a digital currency that allows people to buy and sell in the market. It has a price, and the price is essentially the exchange rate of bitcoin with another currency type, for example, USD. It is a similar story to other types of cryptocurrencies. Therefore, cryptocurrency has value, and it is priced and traded in the market.

After understanding what cryptocurrency is, the most intuitive question is: is it a real currency? So far, academic researchers have an agreement that a cryptocurrency such as bitcoin serves as a financial asset with investment and speculation functions rather than an actual currency (Buchholz et al., 2012 [1]; Kristoufec 2013 [2]; Ciaian et al. 2016 [3]; Bouoiyour et al. 2015 [4]; Yermack 2015 [5]). Yermack (2015) [5] claimed that bitcoin does not satisfy a fiat currency's criteria, which includes the functions of medium of exchange, a store of value, and a unit of account. Yermack's statement is also supported by Cheah and Fry (2015) [6], who argue that bitcoin fails to be a true unit of account or a store of value due to its high volatility. The agreement that cryptocurrencies behave as financial assets rather than currencies is consistent with our real-world experience. When we are talking about cryptocurrency, what do we think? Are we thinking of holding it like cash and using it for purchasing a meal? Or are we considering putting it in the retirement portfolio? Apparently, we regard it as an investment tool.

1.3 Cryptocurrency's price and risk

Since we have determined that cryptocurrency is a type of financial asset, it is necessary to look into its economic properties. There are two themes about cryptocurrency to explore: its price and risk. Over the decade, we have seen the miracle of bitcoin price starting from nearly zero in 2009 and rushed to $18 thousand in 2018. Numbers of altcoins poured into the market at the same time. This young market was expanding aggressively. However, not everyone welcomed the birth of blockchain technology and cryptocurrencies. In 2011, "the Silk Road" platform announced using bitcoin as the means of payment. The platform was an online black market that engaged with illegal drug trading. Cryptocurrencies also involved other illegal activities such as money laundering and terrorism. Some countries' governments noticed this and put regulations on cryptocurrency trading, and some even banned the trading. In early 2018, cryptocurrency trading was banned by the Chinese government. After 2018, the bitcoin price experienced a crash and dropped to only $4 thousand. The global cryptocurrency market capitalization also fell from $830 billion in January 2018 to $115 billion in December 2018. The crash seems to confirm people's belief in the existence of pricing bubbles in cryptocurrencies. Many studies of cryptocurrencies bubbles can be found in literature: Cheung et al. (2015) [7] detected robust bubbles in bitcoin by using econometric methodology; Corbet et al. (2018) [8] claimed the significant bubble behavior in bitcoin; Xiong et al. (2020) [9] verified bitcoin bubbles based on the production cost, and predicted the next bitcoin bubble would happen in late 2020. However, unlike tulip mania in the 17th century, bitcoin's "bubble" did not burst as people expected; on the contrary, what does not "kill" makes it stronger. The miracle of cryptocurrencies comes back: bitcoin price started to rise in early 2019 and jumped up to over $50 thousand as of February 2021; the global cryptocurrency market capitalization reached up to $1,747 billion at the same time point, which is fifteen times of that in December 2018.

A significant number of rags-to-riches stories of bitcoin holders excite people's interests and expectations. However, it also brings us to confusion: what drives cryptocurrency prices? Many researchers have investigated the cryptocurrency price formation and identified four determinants: The first is the economic driver, where the price is determined by the interaction of bitcoin supply and demand (Buchholz et al. 2012 [1], P. Ciaian et al. (2015) [10]. The second determinant is the popularity of bitcoin among investors: Kristoufek (2013) [2] studied the relationship between the bitcoin price and the number of searched terms related to bitcoin on Google Trends and Wikipedia and found a significant positive correlation between them. Then Kristoufek found out how they were correlated. Kristoufek (2015) [11] employed Wavelet Coherence analysis's approach and uncovered that the Google and Wikipedia searches on bitcoin have an asymmetric effect. Online attention pulls the price higher during the bubble formation and pushes the price lower prices to collapse. P. Ciaian et al. (2015) [10], followed Kristoufek (2013) [2], did not use Google and Wikipedia searches. They used the number of new members and new posts on bitcointalk.org as variables. Bitcointalk.com is a message board that Satoshi Nakamoto founded in 2009. People who are interested in bitcoin can talk to each other on this bitcoin forum. Their result stressed that investment attractiveness has a significant impact on bitcoin price. The third determinant is a technical driver. Kristoufek (2013) [2] observed that the difficulty of mining one bitcoin unit is positively correlated with the bitcoin price. The fourth determinant is the macroeconomic and financial indicators. Van Wijk (2013) [12] pointed out the impact of macroeconomic and financial indicators on bitcoin price. However, the study of P. Ciaian et al. (2015) [10] examined this factor, and the result shows that the macroeconomic and financial indicators do not have a significant impact on the bitcoin price in the long run.

However, these determinants are external drivers of cryptocurrency. As numerous altcoins are emerging in the market, it is necessary to consider the internal drivers when exploring price formation. Several researchers have noticed this knowledge gap that peer linkages may play an important role in cryptocurrency price formation. The first essay, "How cryptocurrency prices affect each other," investigates the inter-market relationships among selective cryptocurrency prices by employing a vector error correction model to contribute to this literature gap. We selected four cryptocurrencies, bitcoin, Ethereum, XRP, and Litecoin. These four cryptocurrencies were leading the market with large market capitalization, trading volume, and long price history. Initially, we consider the vector autoregression model (VAR) because VAR is one of the most successful multivariate time series models. VAR applies to the situation that all the time series variables are stationary without cointegrating relationships. However, we found the existence of cointegration between the four variables. Therefore, we applied the vector error correction model (VECM), which adds error correction terms to the VAR model and analyzes both long-run equilibrium and short-run adjustment of the four variables. The empirical results show that cryptocurrency's price affects each other. There exists two long-run equilibriums among these cryptocurrencies. Each price is not only influenced by its past value but other cryptocurrencies' prices as well. We also test for the causal relationship between the four variables by using the Granger causality test. Then we estimated the impulse response function (IRF) based on VECM to track how each variable responds to system shocks' impulses. It is found that a shock in each variable has either a short-term or long-term impact on other cryptocurrencies' prices. The results of VECM and causality tests provide evidence that competition in the cryptocurrency market matters. Investors need to consider it when making investment decisions. After analyzing cryptocurrency's price, we addressed the second issue, risk management. As financial assets, the most well-known feature of cryptocurrencies is their extreme fluctuations. People are excited about the tale of wealth that bitcoin makes a millionaire overnight. However, high returns always go with high risks. Let us suppose if we invested one unit of Ethereum on January 1st, 2018, held it for a month, and sold it at the end of January, then our rate of return would be positive 43.5%; but if we wait for a while and sell it at the end of June, our rate of return would drop to negative 41.2%, which is a disaster for an investor. The most important concern for a cryptocurrency investor is the level of risk exposure. In the financial world, volatility is commonly used to measure the risk of a financial asset.

Thus, the modeling and forecasting of cryptocurrency volatility are crucial for investors' decision-making and risk management. There has been a growing literature on cryptocurrency volatility analysis. Researchers implemented different types of Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models to capture the cryptocurrency return's volatility (Katsiampa, 2017 [13]; Blau, 2018 [14]; Peng et al., 2018 [15]; Cheikh et al. 2020 [16]). These earlier studies of cryptocurrency volatility were founded on economic models. However, research on cryptocurrency volatility analysis using machine learning algorithms is still scarce. Machine learning is an advanced and promising approach that may significantly impact and make future contributions to financial markets. Some people believe that machine learning methods are more advanced and more efficient than economic models. Our interest is to figure out whether this is true or not in bitcoin volatility forecasting.

The second essay, "Bitcoin Return Volatility Forecasting: A Comparative Study of GARCH Model and Machine Learning Model," uses conventional economic models - exponentially weighted moving average and GARCH model - and machine learning model-recurrent neural network models - to forecast the volatility of bitcoin return and calculate the forecasted Value at Risk of bitcoin return. We compare the out-of-sample performance of the three models by using RMSE and MAE. We also plot the figures that compare the forecasted volatility and the realized one with three different forecasting horizons: one day, five days, and ten days ahead. We found that the RNN outperforms GARCH and EWMA in average forecasting performance. However, it is less efficient in capturing the bitcoin market's extreme events. Moreover, the RNN shows poor performance in Value at Risk forecasting, indicating that it could not work well as the econometric models in explaining extreme volatility. This study proposed an alternative way of volatility analysis. It is widely believed and empirically proved by earlier studies in the financial market that the machine learning approach is more advanced in time series forecasting. However, this study shows something different. The RNN model is less efficient than traditional econometric models in bitcoin volatility forecasting and risk management. The econometric models are superior in analyzing extreme market conditions, while the machine learning approach is more suitable for less volatile market conditions.

Let us go back to the question that cryptocurrency holders are concerned about: how much risk exposure they have? In other words, how much money will they potentially lose tomorrow? In addition to volatility, another way to evaluate financial asset risk is to investigate its loss tail behavior. The third essay, "Estimation of tail risk measures in the cryptocurrency market," focuses on cryptocurrency index (CRIX) tail events. On an excellent tail estimation, extreme value theory (EVT) may be a proper parametric method. However, the iid assumption makes extreme value theory an inappropriate approach for cryptocurrency index loss data. To address this issue, we filter the asset loss by using a GARCH model. It is observed that the time series of cryptocurrency index loss is serially correlated and presents volatility clustering, which implies the existence of conditional heteroscedasticity in volatility. We take it into account by employing the GARCH model under different probability distribution assumptions. The filtered residual will satisfy the iid assumption of extreme value theory modeling. Then the task of asset conditional loss distribution estimation could be transferred to the standard innovation distribution estimation. To show a clear preference for the models that focus on tail events, we also apply a standard parametric approach that makes a specific probability distribution assumption to the full distribution and compares its performance with the extreme value theory approach. The results show that the extreme value theory method dominates the non-extreme value theory method by generating more accurate and plausible Value at Risk and expected shortfall forecasting, especially for the more extreme quantiles. We alternatively employ a semiparametric approach, the filtered historical simulation (FHS), in this study. The filtered historical simulation also focuses on tail events. FHS method yields similar results to the extreme value theory approach, which can be regarded as evidence of the results' robustness. Therefore, focusing on tail events using extreme value theory and filtered historical simulation methods could significantly improve CRIX tail risk estimation accuracy, which is more reliable in cryptocurrency market risk management.

2. CRYPTOCURRENCIES PRICES AFFECT EACH OTHER

2.1 Introduction

Since the launch of the first unit of bitcoin by Satoshi Nakamoto in 2009, the blockchain technology and cryptocurrency market have been proliferating, attracting people's attention and interests. Over 1,600 types of cryptocurrencies are trading in the market as of March 2018, with a total market capitalization of around 3,000 billion USD. People are curious about this emerging market and trying to figure out the cryptocurrency price formation. Economics researchers have identified some of the dominants of bitcoin price formation, such as bitcoin supply and demand, bitcoin popularity among investors, technical drivers which represent the difficulty of bitcoin mining, and macroeconomic and financial indicators (Buchholz et al. 2012 [1], P. Ciaian et al. 2015 [10], Li and Wang. 2017 [17], Kristoufek, 2015[11],Van Wijk, 2013 [12].)

These earlier studies on cryptocurrency price formation are focused on factors that are outside the cryptocurrency market. However, it is necessary to consider the inter-market relationship when exploring price formation. There are hundreds of competitive altcoins entering the market, and these altcoins are competing with the market leader, bitcoin. Altcoins here refer to the alternative cryptocurrencies that are not bitcoin. These altcoins have similar features to bitcoin, and some of them are even more advanced. For instance, Litecoin provides faster transaction confirmation times than bitcoin. It is reasonable to believe that these cryptocurrencies, either bitcoin or altcoin, influence each other's prices. On the one hand, the growing number of cryptocurrencies signals that the market attracts more investors and more capital, leading to an increase in their prices. On the other hand, these cryptocurrencies are substitutions and competitors to each other, which indicates the prices might move in opposite directions. Therefore, the inter-market relationship may affect cryptocurrency price formation. Understanding this relationship is quite crucial for constructing cryptocurrency portfolios and making investment decisions. Some researchers have noticed this knowledge gap in the inter-market relationship. Earlier studies found that bitcoin and altcoins have statistical similarities, and there is a strong impact of network effects on them (Gandal and Halaburda, 2017[18], Osterrieder et al., 2017 [19]). However, as many altcoins emerged in the market after the year 2017, bitcoin dominance became much weaker than before. The relationship between bitcoin and altcoins could be analyzed but it may be more meaningful to analyze the inter-market relationship among multiple cryptocurrency prices.

Therefore, we follow the literature to investigate the inter-market relationships among selective cryptocurrency prices by employing the most recent data after 2017. The representative cryptocurrencies we select are bitcoin, Ethereum, XRP, and Litecoin. These four cryptocurrencies are leading the market and have the largest market capitalization. Ethereum was launched in 2015 and became the second-largest cryptocurrency on the market as of 2018. Ethereum's market capitalization is just behind the market leader, bitcoin. XRP, the world's 3rd most prominent cryptocurrency, had a market capitalization of over 90 billion USD at the beginning of 2018, which accounted for about thirteen percent of the whole cryptocurrency market. Litecoin was released in 2011 and reached a 13 billion USD market capitalization at the beginning of 2018.

This study aims to investigate the inter-market relationships among bitcoin, Ethereum, XRP, and Litecoin prices. A vector autoregression model (VAR) is an excellent methodology because it is one of the most successful multivariate time series models. The vector autoregression model was first introduced by Chris Sims in 1980 and became one of the most applied models in empirical economics. The vector autoregression model applies to the situation that series variables are stationary all the time. However, the cryptocurrencies prices are usually non-stationary, and they are demonstrated to be cointegrated. We use the vector error correction model (VECM) instead of the vector autoregression model to address this issue. The vector error correction model adds error correction terms to the vector autoregression model, which allows us to analyze both long-run equilibrium and short-run adjustment of the variable. We also examine the influence of one cryptocurrency innovation on the other with impulse response functions. To go further, we use the Granger causality test to explore the causality among the four cryptocurrencies' prices. It is expected that the cryptocurrencies prices significantly affect each other.

2.2 Methodology

In this section, the dynamic long-run and short-run relationships between bitcoin, Ethereum, XRP, and Litecoin are examined by employing a vector error correction model. The Johansen trace test is used to assess the cointegrating relationship among these variables.

2.2.1 Cointegration

The presence of non-stationary time series data is familiar in economics studies. However, it makes the multivariate time series model complicated. If there are at least one or two non-stationary time series, it is possible to have one or more cointegrating relationships among the non-stationary variables. The time series variables are said to be cointegrated if the linear combination between them is stationary. The existence of cointegration indicates the long-run relationship among variables. In this study, it is reasonable to expect long-run relationships among bitcoin, Ethereum, XRP, and Litecoin, and the relationship between each pair is supposed to be either positive or negative. Because as the technology behind these cryptocurrencies – blockchain - attracts more attention and capital, the prices are supposed to move together. However, these cryptocurrencies are also substitutions for each other. For example, bitcoin's market share (the percentage of the cryptocurrency market capitalization attributed to bitcoin) was 94.29% in April 2013 but dropped to 33.44% in January 2018. The price drop provides evidence that the co-movement of bitcoin and altcoins are supposed to be negative. Johansen trace cointegration test is used to detect the cointegrating relationships among these four prices.

2.2.2 Vector error correction model (VECM)

After specifying the long-run equilibrium, the vector error correction model analyzes the dynamic relationships among these four cryptocurrencies' prices. Equation (4.1) presents the vector error correction model in this study:

$$\Delta x_t = \mu + \sum_{i=1}^{p-1} \tau_i \Delta x_{t-i} + \alpha\beta^{'} x_{t-1} + u_t \tag{2.1}$$

Where μ is the deterministic shift vector, x_t is the vector of the natural logarithm of cryptocurrency prices $x_t = [btc_t, eth_t, xrp_t, lte_t]'^1$. p is the lag length and τ_i denotes the (4×4) parameter matrices of the lagged stationary difference. β is a (4×r) matrix of r cointegration vectors, and α is a (4×r) matrix that indicates how each series responds to perturbations in the long-run equilibrium given $\beta' x_{t-1}.\Pi = \alpha\beta'$ represents the long-run relationships between the four variables in x_t. The vector error correction model (VECM) adds error correction terms to the vector autoregression (VAR) model and analyses both the long-run effect and short-run adjustment process in this model.

2.3 Data and empirical results

We employ the daily closing price of bitcoin, Ethereum, XRP, and Litecoin that range from January 03, 2017, to January 01, 2020, with a total of 1,094 observations. The cryptocurrency price is defined as its exchange rate against the USD. We use the natural logarithm of each variable, where $btc_t, eth_t, xrp_t, ltc_t$ is the nature logarithm of bitcoin, Ethereum, XRP, and Litecoin price. The price data are available on the website: coinmarketcap.com. Table 2.1 summarizes the descriptive statistics for each time series variable: btc_t, eth_t, xrp_t,and ltc_t.

Description	btc	eth	xrp	ltc
Mean	8.55	5.32	0.31	3.97
Median	8.78	5.39	0.27	4.01
Max	9.88	7.24	1.48	5.88
Min	6.66	2.27	0.01	1.31
S.D.	0.72	1.00	0.21	0.99
Skewness	-0.93	-1.15	1.82	-1.20
Obs	1094	1094	1094	1094

Note: $btc_t, eth_t, xrp_t, ltc_t$ denote the natural logarithm of bitcoin, Ethereum, XRP and Litecoin.

Table 2.1: Summary statistics for daily closing price from 1/3/2017 to 1/1/2020 .

[1] $btc_t, eth_t, xrp_t, ltc_t$ denotes the nature logarithm of bitcoin, Ethereum, XRP, and Litecoin price, respectively.

2.3.1 Unit root test

We first examine the stationary of each variable by using the augmented Dickey-Fuller unit root test. Table 2.2 shows that all four cryptocurrencies are non-stationary, and their first difference is stationary. Therefore, these four variables are I(1) time series.

	Level prices		First difference of prices	
Cryptocurrency	ADF	p-value	ADF	p-value
btc	1.17	0.94	-33	0.01***
eth	0.73	0.86	-33	0.01***
xrp	-1.04	0.31	-31.3	0.01***
ltc	0.43	0.77	-32.5	0.01***

Note: $btc_t, eth_t, xrp_t, ltc_t$ denote the natural logarithm of bitcoin, Ethereum, XRP and Litecoin; *** significant at 1% level.

Table 2.2: Unit root test.

2.3.2 Long run equilibrium

Johansen trace cointegration test is used to determine whether these four variables have long-run relationships. The test results in table 2.3 indicate two long run equilibriums (r=2) at a significant level 5%.

H_0	T	C(5%)	C(1%)
$r \leq 3$	3.07	8.18	11.65
$r \leq 2$	17.93	17.95	23.52
$r \leq 1$	35.37**	31.52	37.22
$r \leq 0$	74.89***	48.28	55.43

Note: ***significant at 1% level, ** significant at 5% level. T is trace test value; C (5%) is the critical value at significant 5% level; C(1%) is the critical value at significant 1% level.

Table 2.3: Johansen trace cointegration test.

The lag length is selected based on Akaike Information Criterion (AIC), and the optimum lag order was 10. The normalized equation (4.2) and equation (4.3) represent the significantly long-run equilibrium among bitcoin, Ethereum, XRP, and Litecoin price:

$$btc_t = 7.49xrp_t - 1.63ltc_t \tag{2.2}$$

$$eth_t = -36.86xrp_t + 10.16ltc_t \tag{2.3}$$

The $btc_t, eth_t, xrp_t, ltc_t$ denotes the nature logarithm of the respective cryptocurrency price. The normalized cointegrating equation (4.2) and (4.3) indicate that: the bitcoin's long-run relationship with XRP is positive, while its relationship with Litecoin is negative; the Ethereum's long-run relationship with XRP is negative, while its relationship with Litecoin is positive; there does not exist a long-run relationship between bitcoin and Ethereum. The estimated coefficients are interpreted as the cryptocurrency's price elasticities. A one percent increase in XRP price would contribute to a 7.49 percent rise in bitcoin price and a 36.86 percent decline in Ethereum price in the long-run equilibrium. A one percent increase in Litecoin price would be associated with a 1.63 percent decrease in bitcoin price but a 10.16 percent increase in Ethereum price.

2.3.3 Short run equilibrium

The cointegration analysis implies that the cryptocurrencies price would return to the equilibrium in the long run, however, it does not tell how they move toward the equilibrium. To analyze the short run dynamic of each variable, vector correction model is used to show how each variable responds to the disequilibrium.

Table 2.4 reports the short-run adjustment effects. First, the estimated error correction coefficients are significantly negative for each variable, implying that each cryptocurrency price contributes to reducing the deviation by moving toward the long-run equilibrium when the system deviates from the equilibrium. In other words, if the cryptocurrency price at the time $t-1$ were above the equilibrium, it would fall in the next day; and if its price were below the equilibrium, it would rise in the next day. Second, the estimated error correction coefficient's value illustrates the speed of adjustment to the long-run equilibrium. A relatively small value implies a slow rate of adjustment to equilibrium, while a relatively large value implies a fast rate. For instance, bitcoin's estimated error correction coefficient is -0.0136, indicating that the bitcoin price corrects for 1.36 percent of the discrepancy every trading day. We notice that the adjustment rate of XRP price and Litecoin price to the second equilibrium (equation(4.3)) is slower than the first equilibrium (equation(4.2)), implying that they respond more quickly to the deviations from the equilibrium with bitcoin price than that of Ethereum price. Third, each cryptocurrency price is influenced by the lagged price of itself and others.

2.3.4 Granger causality analysis

To further investigate the relationship among bitcoin, Ethereum, XRP, and Litecoin prices, we test each variable's causality relationship. Granger causality analysis is a statistical method that is the most widely applied in economics when people investigate information flow between time series. It was proposed by Granger in 1969, and this approach is used to determine the causality between two time series data. To explore the intuition behind Granger causality, suppose there are two time series data sets, X and Y. If the predictions of variable Y based on its past value and the past value of X are better than the predictions based on its past value alone, then we say variable X Granger causes variable Y. Now suppose X_t is a d-dimensional multivariate time series with $t = 1, 2, , T$.The mathematical formulation for multivariate Granger causality (Granger 2969 [20]) is as follows :

$$X_t = \sum_{\tau=1}^{p} A_\tau X_{t-\tau} + \epsilon_t \tag{2.4}$$

Where p is the maximum number of lagged observations; matrix A is the coefficients matrix and ϵ_t is the error term.

The Granger causality test specifies the direction of the causal relationship between two variables. Table 2.5 presents the Granger causality relationship results. The estimated results indicate there is bi-directional causality between bitcoin and XRP prices, and between Ethereum and XRP prices; there is a unidirectional causality relationship from bitcoin to Ethereum, from XRP to Lite-

	Δbtc	Δeth	Δxrp	Δltc
ecm1	-0.0136***	-0.0135***	-0.0046***	-0.0125***
ecm2	-0.0036***	-0.0027**	-0.0016***	-0.0029**
interpret	0.0897***	0.1094***	0.0209*	0.0926***
Δ btc(-1)	0.0679	-0.0129	-0.0566**	0.0158
Δ btc(-2)	-0.0287	0.0292	-0.0357	0.0357
Δ btc(-3)	0.0227	-0.0266	0.0181	0.0207
Δ btc(-4)	-0.1104***	-0.0208	-0.065***	-0.1046*
Δ btc(-5)	0.0365	0.1161**	-0.0468**	0.2046***
Δ btc(-6)	-0.0094	-0.0385	-0.012	-0.0968
Δ btc(-7)	-0.0001	0.0401	0.0632***	0.0018
Δ btc(-8)	0.0703	0.0978*	0.0252	0.094
Δ btc(-9)	-0.012	-0.02	-0.0505**	-0.0247
Δ btc(-10)	0.1263***	0.1114*	-0.0228	0.0825
Δeth(-1)	-0.0548*	-0.0005	0.0148	-0.0651
Δeth(-2)	0.0219	0.0128	0.0054	-0.0243
Δeth(-3)	0.0035	0.0703	-0.0261	-0.0444
Δeth(-4)	0.0149	0.0283	0.0284	-0.0131
Δeth(-5)	-0.003	0.0174	0.006	-0.0173
Δeth(-6)	-0.0039	0.0023	-0.0299*	0.058
Δeth(-7)	0.0048	-0.0448	-0.0462***	0.06
Δeth(-8)	-0.023	-0.0089	-0.0114	0.001
Δeth(-9)	-0.053*	-0.0042	0.0276	-0.0408
Δeth(-10)	0.0543*	0.0403	0.0645***	0.0057
Δxrp(-1)	-0.1042*	-0.2024**	0.0254	-0.1285
Δxrp(-2)	0.2149***	0.1324	0.1128***	0.0354
Δxrp(-3)	0.0224	0.075	0.0169	0.1355
Δxrp(-4)	0.01	0.0884	0.0736**	0.0198
Δxrp(-5)	0.0666	0.08	0.0485	-0.0264
Δxrp(-6)	0.1838***	0.3051***	0.1284***	0.1911**
Δxrp(-7)	0.0589	0.0925	0.1191***	-0.0003
Δxrp(-8)	-0.0478	-0.1174	0.1333***	0.0446
Δxrp(-9)	0.0852	0.2408***	0.0103	-0.037
Δxrp(-10)	-0.1528**	-0.212**	-0.2228***	-0.0735
Δltc(-1)	0.0011	0.0566	0.0104	0.0652
Δltc(-2)	-0.0236	-0.0259	0.0007	-0.0117
Δltc(-3)	-0.0076	-0.0403	0.0372**	0.0091
Δltc(-4)	0.0516*	-0.0425	-0.0096	0.0958**
Δltc(-5)	-0.0237	-0.058	0.0144	-0.0787*
Δltc(-6)	0.0096	0.0216	0.009	0.0731*
Δltc(-7)	-0.0335	-0.0275	-0.0469***	-0.0734*
Δltc(-8)	-0.0376	-0.0387	-0.0202	-0.0962**
Δltc(-9)	-0.0083	-0.0679*	0.0098	0.0047
Δltc(-10)	-0.0774***	-0.0363	-0.0002	-0.0243

Note: Δ stands for the first difference of each variable; the number in the parentheses stands for the number of lags; ecm1 denotes the first cointegrating relationship ((4.2))), and ecm2 denotes the second cointegrating relationship ((4.3)). *** is significant at 1% level, ** is significant at 5% level, and * is significant at 10% level.

Table 2.4: The estimated vector error correction model coefficients.

coin, and from Litecoin to bitcoin. In terms of technique, Litecoin was proposed to be the leading rival for bitcoin, which might explain its unidirectional Granger causal effect on bitcoin.

Null hypothesis	F-statistics	p-value	Decision
eth does not Granger cause btc	0.62	0.80	eth does not Granger cause btc
xrp does not Granger cause btc	2.27	0.01***	xrp Granger causes btc
ltc does not Granger cause btc	1.89	0.04**	ltc Granger causes btc
btc does not Granger cause eth	1.79	0.06*	btc Granger causes eth
xrp does not Granger cause eth	4.07	0.00***	xrp Granger causes eth
ltc does not Granger cause eth	1.03	0.42	ltc does not Granger cause eth
btc does not Granger cause xrp	3.65	0.00***	btc Granger causes xrp
eth does not Granger cause xrp	3.35	0.00***	eth Granger causes xrp
ltc does not Granger cause xrp	0.99	0.45	ltc does not Granger cause xrp
btc does not Granger cause ltc	1.41	0.17	btc does not Granger cause ltc
eth does not Granger cause ltc	0.69	0.73	eth does not Granger cause ltc
xrp does not Granger cause ltc	3.34	0.00***	xrp Granger causes ltc

Note: *** is significant at 1% level, ** is significant at 5% level, and * is significant at 10% level. $btc_t, eth_t, xrp_t, ltc_t$ denote the natural logarithm of bitcoin, Ethereum, XRP and Litecoin.

Table 2.5: Granger causality test.

2.3.5 Impulse response function

The impulse response function (IRF) is used to track how the system's variables respond to the system's shocks' impulses. It represents the mechanisms of how the shock spreads over time. We estimate the impulse response function base on the vector error correction model. Figure 2.1, Figure 2.2, Figure 2.3 and Figure 2.4 plot the impulse response of each variable to the innovation of bitcoin, Ethereum, XRP, and Litecoin price, respectively, with 95% bootstrap confidence interval (redlines). The vertical axis is the impulse response function, and the horizontal axis represents 180 days aftershock.

Figure 2.1 shows that innovation to the bitcoin price induces a significant positive effect on Ethereum, XRP, Litecoin, and itself. The effect declines over time but is still significantly positive

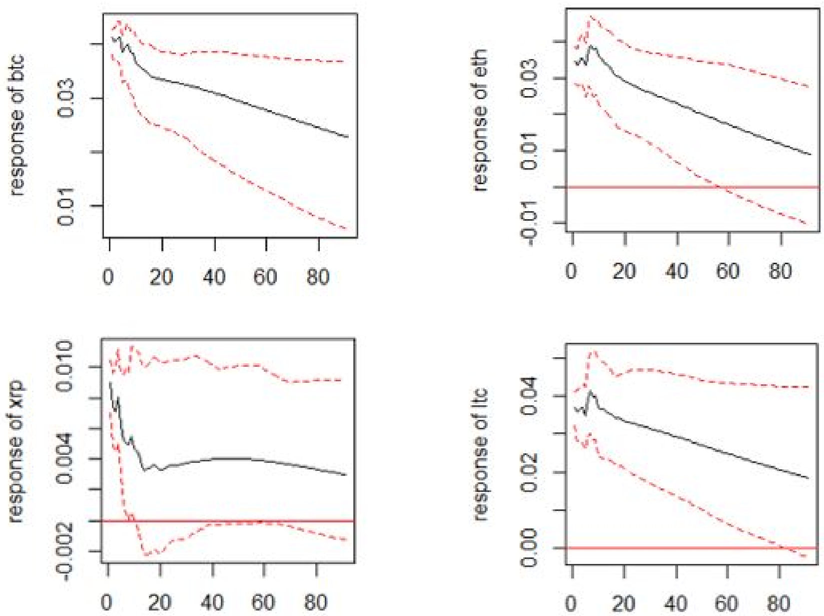


Figure 2.1: Response of each variable to bitcoin price innovation (btc)

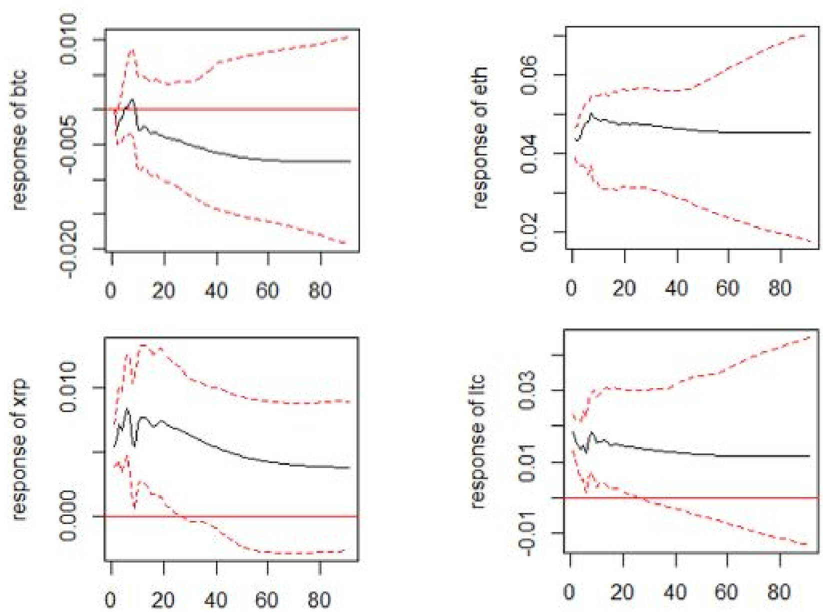


Figure 2.2: Response of each variable to Ethereum price innovation (eth)

in the long run. Figure 2.2 shows that innovation to the Ethereum price directly affects itself, XRP, and Litecoin while negatively affecting bitcoin. These impacts decline over time but last for a

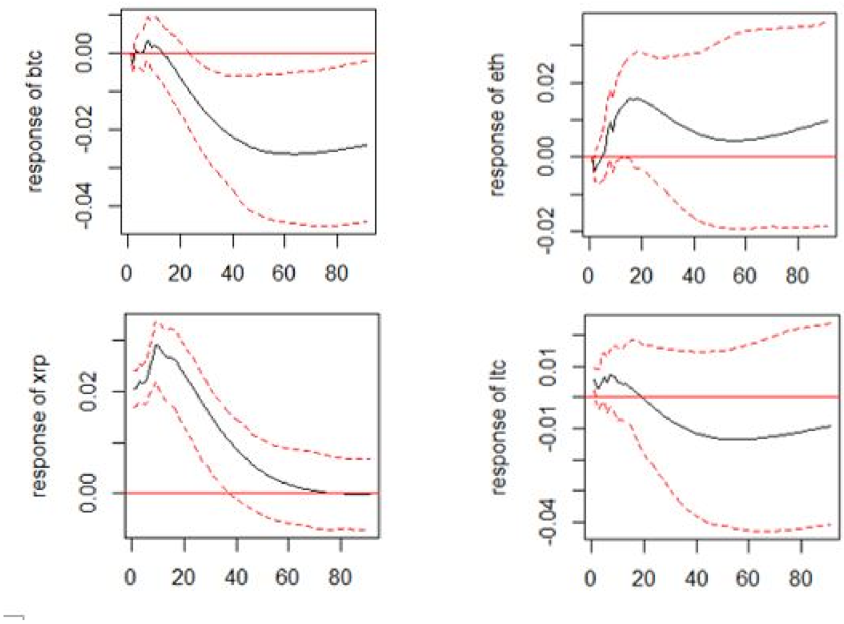


Figure 2.3: Response of each variable to XRP price innovation (xrp)

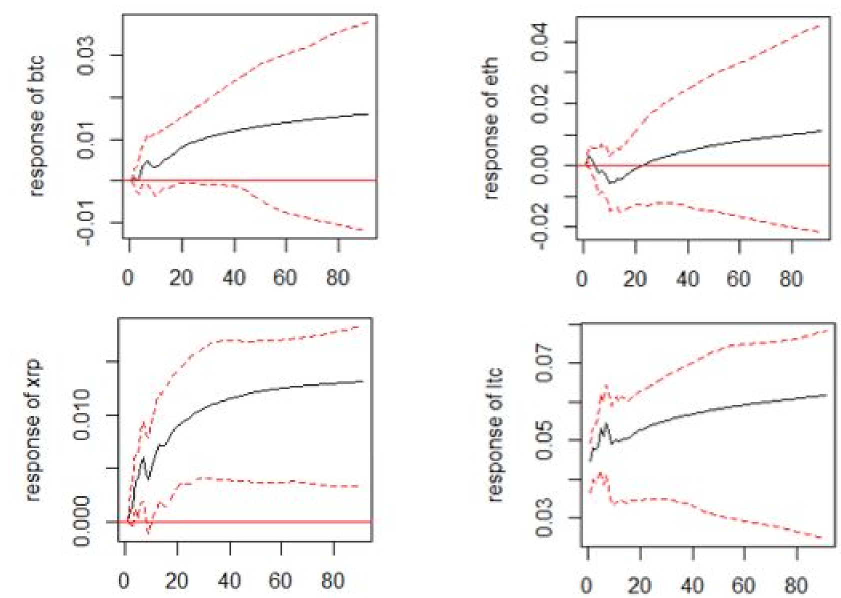


Figure 2.4: Response of each variable to Litecoin price innovation (ltc)

long time. Figure 2.3 shows that the bitcoin and Litecoin to XRP innovation are initially slightly positive but finally decline to negative. Ethereum and XRP to XRP innovation are initially positive and increase over time, reaching a peak in the second to the third week but dies out around 60 days. Figure 2.4 shows that the response of bitcoin, XRP, and Litecoin to Litecoin innovation is significantly positive, and the positive effect lasts for a long time. However, the innovation hurts Ethereum initially, but the impact turns positive after 20 days and remains positive in the long run.

2.4 Conclusion

We investigate the inter-market relationships among four different cryptocurrencies. By studying the long-run equilibrium and short-run dynamics of bitcoin and selective altcoins under the vector error correction model's framework, we found cointegrating relationships among bitcoin, Ethereum, XRP, and Litecoin prices. The cointegrating relationships reveal two long-run equilibriums. One is between bitcoin and altcoins, and the other is among altcoins themselves. The long-run relationships influence the cryptocurrencies' short-run dynamics. In terms of the first long-run equilibrium among bitcoin and altcoins, the prices respond to the deviation by correcting their short-run dynamics and return to the equilibrium quickly. However, the response and adjustment to the second equilibrium are slower than that of the first one. One explanation is that although altcoins weaken bitcoin's leadership, it still dominates the market. The short-run dynamics imply that each cryptocurrency's price is affected by past value and other cryptocurrencies. Therefore, the cointegrating analysis and vector error correction model provide evidence of inter-market relationships.

However, the cointegrating relationships are unable to provide information on the direction of the causal relationship between variables. Therefore, we use Granger causality analysis to investigate the causal relationship between each two cryptocurrency prices. We found a bi-directional Granger causal relationship between bitcoin and XRP price and between XRP and Ethereum. An interesting phenomenon is that during the sample period from January 2017 to January 2020, we detect that bitcoin price and Ethereum price have neither a long-run equilibrium nor Granger causal relationship. The linkage between these two cryptocurrencies, with the largest market capitalization, seems to be relatively weak. It is worth discussing the reason for this. Adem et al. (2018) [21] demonstrated that neither bitcoin nor Ethereum dominants the other regarding their underlying technique. Therefore, investors interested in blockchain technology should not have a strong preference for either one. The high total cryptocurrency market capitalization after the year 2017 suggests that investors are also interested in speculation. One unit of bitcoin's price was much higher than that of Ethereum: the price of one unit of bitcoin was over one thousand dollars at the beginning of 2017 and reached 17 thousand at the end of 2017. However, one unit of Ethereum was only around 8 dollars at the beginning of 2017 and reached 1 thousand dollars at the end of 2017. The relatively more risk-averse speculators might feel that bitcoin was too hot to invest, and they turned to the less-priced alternative, Ethereum. Therefore, we suppose even if Ethereum did not exist in the market, some of the speculators would not invest in bitcoin. Some of the bitcoin investors and Ethereum investors belong to two groups of people whose investment decisions do not affect each other.

The impulse response function analysis shows that each cryptocurrency price significantly responds to innovations of other cryptocurrency prices. Once the investor observes a shock in a cryptocurrency price, it is a signal that the others will respond to it, and the effect will last a long time.

Therefore, cryptocurrency prices affect each other, and investors need to consider their inter-market relationships when constructing portfolios and making investment decisions.

3. BITCOIN RETURN VOLATILITY FORECASTING: A COMPARATIVE STUDY BETWEEN GARCH AND RNN

3.1 Introduction

Since Satoshi Nakamoto proposed the first cryptocurrency in 2009, the cryptocurrency market has received much attention. Bitcoin is the most successful and popular globally, accounting for over fifty percent of the whole cryptocurrency market capitalization in April 2019. Bitcoin's enthusiasm is due to its innovational features of decentralization and anonymity. Some public companies start to hold bitcoin as an asset, and some financial institutions consider bitcoin into investment strategy by allocating it in their portfolios. Many industries get interested in the blockchain technology behind bitcoin and even start to launch their cryptocurrencies. The agricultural industry is an excellent example of applying blockchain in agricultural insurance, product transactions, supply chain, and smart agriculture (Xiong et al., 2020 [9]). The Covantis, a company co-owned by a global agribusiness group, has launched a blockchain platform for global commodities trading. Some cryptocurrencies are connected to agricultural industry trading in the market, for instance, Carboncoin, Blocery, and Herbalist Token. Although the application of blockchain technology and the cryptocurrencies related to agribusiness are in the infancy stage, it is still necessary for agribusiness researchers to understand this market. Therefore, looking insight the bitcoin's behavior is a good starting point for understanding the Agri-crypto market. Researchers' analysis of bitcoin has received growing interest. David Yermack (2015) [5] studied bitcoin's features and functions and concluded that bitcoin appears to be more like a speculative investment than a real currency due to its high volatility. So far, academic researchers have an agreement that bitcoin serves as a financial asset rather than a real currency. Then it is necessary to look into its economic properties. There are two themes about it to explore: its price and risk.

People are excited about the tale of wealth that bitcoin makes a millionaire overnight. However, high return always goes with high risk. If people look at the bitcoin price history from 2009 till now, its violent fluctuations will be discovered. Investors have to be aware of these vast fluctuations and consider them when making investment decisions. As a financial asset, bitcoin is famous for its extreme volatility. Focusing on asset return volatility is a key to portfolio constructing, asset pricing, risk measuring, and managing. In particular, the asset return's volatility is a simple but widely used risk measurement. Researchers notice the fat tail of bitcoin return, indicating the high probability of significant losses. For better risk control, a sufficient amount of capital covering the potential losses of the asset trading at a given confidence level in the given period is required. And this required value is called Value at Risk (VaR). A more accurate Value at Risk implies a higher risk management efficiency. And the asset return's volatility is a key factor in Value at Risk calculation. Therefore, the modeling and forecasting of bitcoin volatility are crucial for bitcoin investors' decision-making analysis and risk management.

Economic researchers have been making efforts to improve the bitcoin return's volatility forecasting accuracy, and the econometric methods are usually applied. Earlier studies mainly explored bitcoin volatility by using GARCH family models. Bouoiyour and Selmi (2015, 2016) [22] [23]compared different GARCH type models on sub-period bitcoin volatility, and Katsiampa (2017) [13] compared the GARCH family models over the whole period. Balcilar et al. (2017) [24] found the bitcoin trading volume fails to predict bitcoin volatility by studying their causal relationship. Troster et al. (2019) [25] considered the heavy tail character of bitcoin return and compared bitcoin's return and volatility forecasting performance of GARCH and GAS model. They found that the heavy tail models outperform normally distributed models, and the heavy tailed GAS model provides better Value at Risk forecasting. Therefore, in terms of research techniques, the econometrics models are usually and maturely applied in bitcoin volatility forecasting.

However, research on bitcoin volatility forecasting using machine learning algorithms is still void. S. Athey (2018) [26] pointed out that machine learning would dramatically impact the field of economics shortly. Unlike the economic models, where the researcher picks a specific model based on economic principles and estimates the parameters, a machine learning algorithm is usually data-driven modeling focused on the selection process. Thus, a machine learning model is not fixed or predetermined but will be refined during a training process. Applying machine learning methods to solve economic issues can potentially make a difference in the economic and financial fields. Some researchers have noticed this literature gap and started to apply machine learning approaches in cryptocurrency trading. A survey by Fang et al. (2020) [27] indicates that up to 2019, among the research on cryptocurrency that involves technical methods, 13.8% applied machine learning methods. However, most of these researches use machine learning methods to predict cryptocurrency prices instead of volatility. Therefore, this study contributes to the literature gap by applying a machine learning method to bitcoin volatility forecasting. In this study, both conventional econometric models and a machine learning model are used to forecast bitcoin return volatility, and their forecasting performance is evaluated. This study aims to compare their performance and discover if machine learning can improve econometrics time series forecasting. The successful development of machine learning techniques in time series forecasting encourages people to apply them in the financial market. Moreover, machine learning's success in stock market prediction leads us to believe that it may also work well for cryptocurrency price forecasting. Also, the empirical studies show that the machine learning method is more efficient than the ARIMA model in bitcoin price prediction. McNally, Roche, and Caton (2018) [28] compared the forecasting performance of the recurrent neural network (RNN), long short term memory (LSTM) network, and ARIMA on bitcoin price and reported that the machine learning models outperformed ARIMA. Laura A. et al. (2018) [29] examined the forecasting performance on cryptocurrency portfolios and reported that machine learning methods overwhelm the standard benchmark simple moving average. It makes sense for the machine learning method to be superior to the traditional economic model (simple moving average and ARIMA). The machine learning model is proposed in a more general scope that considers both linear and nonlinear features. It also preserves more temporal information of a time series during training.

As discussed above, machine learning methods are more advanced than traditional economic models in time series forecasting theoretically and empirically. However, this assertion needs to be cautious. First of all, economic models involve economic intuition, while machine learning mainly deals with data. In the economic world, economic intuition is the key to economic analysis. In contrast, machine learning captures information only from data. However, the information contained in the data is limited in analyzing economic issues. Secondly, the performance of machine learning depends on the amount of data. Its performance is dramatically improved as the data amount getting larger. However, in this study, the bitcoin market history is relatively short. Finally, machine learning is sensitive to fluctuations. Compared to other approaches, machine learning is more efficient in identifying time-series trends and patterns. However, this leads to the problem that a shock or abnormal perturbation will be treated more seriously. Nevertheless, in the real world, many factors affecting the market reaction to the shock or abnormal perturbation, the fluctuation sensitivity might cause overreaction problems in the forecasting, especially for the volatility analysis.

This study compares the forecasting performance between traditional econometric models and machine learning methods in forecasting accuracy and risk management efficiency. Investors are interested in the bitcoin volatility forecasting accuracy performance because they need information on how volatile the market would be in the future. Meanwhile, since the bitcoin market investors face significant risks every day, they are also concerned about risk management. This study contributes to the bitcoin volatility analysis literature in four ways. First, this is the first study that uses the RNN approach with GRU cell for bitcoin volatility forecasting. Second, the GARCH and the RNN are the most popular methods in their field. However, no research work gives detailed descriptions of how they perform differently in the bitcoin market. This study straightforwardly compares them and gives a clear preference for applying either one under different situations. Third, in addition to volatility forecasting accuracy, this study also examines the Value at Risk efficiency, providing more applicable guidance for investors to implement in practice. Four, this study is not the first one to investigate whether the machine learning method is more advanced in financial time series forecasting. However, it contributes to the existing literature by providing more evidence on the limitations of applying machine learning approaches to solve economic issues.

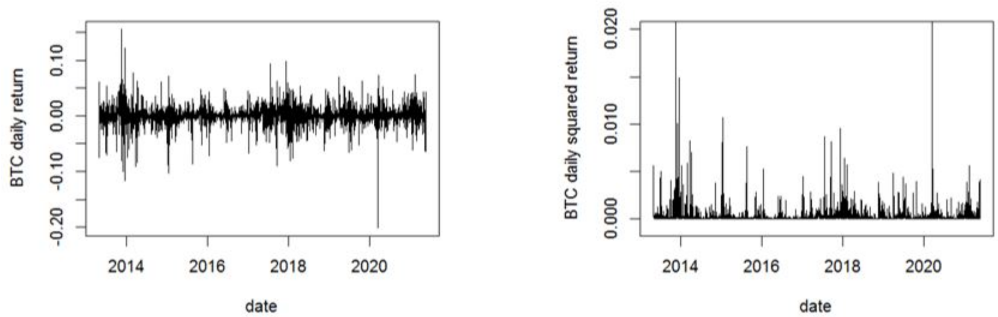
This chapter is structured as follows. First, the econometric models are presented. It starts with the naive model, an exponentially weighted moving average (EWMA) as a benchmark, and then moves to a more complex but conventionally applied model, generalized autoregressive conditional heteroscedasticity (GARCH) model, to forecast bitcoin return volatility. Then a machine learning model based on Recurrent Neural Network (RNN) is proposed. The next step is to evaluate the out-of-sample performance of the three models. The root mean squared error (RMSE) and mean absolute error (MAE) are used to evaluate their forecasting accuracy performances, and the Value at Risk (VaR) is used to compare their risk management efficiency. Because bitcoin return's true conditional volatility is unobservable, the bitcoin daily squared return and Garman-Klass volatility (Garman and Klass, 1980 [30]) are used as proxies for the realized volatility.

3.2 Materials and Methods

In this section, the econometric methodology is discussed first, and then the recurrent neural network model, which is a machine learning methodology, will be presented. Engle, in 1982 proposed the autoregressive conditional heteroscedasticity model (ARCH), which assumes that the volatility of asset returns is time-varying instead of a constant. Bollerslev (1986) [31] generalized the ARCH model and developed a more commonly used GARCH model. In this study, the GARCH model is applied as the econometric method.

The bitcoin return time series is used rather than the raw bitcoin price data. The bitcoin daily return is defined as the difference in the daily bitcoin closing price's natural logarithm. Bitcoin daily opening, high, low, and closing prices are used to estimate the realized bitcoin volatility. All the data are available on the website: CoinMarketCap.com. The data ranges from April 30, 2013, to May 21, 2021, with 2944 observations. Figure 3.1 illustrates the bitcoin daily return and bitcoin daily squared return respectively, and table 3.1 shows the descriptive statistics of the bitcoin daily return.

Before going further to the econometric modeling, the stationary of the time series must be checked. The augmented Dickey-Fuller- test (ADF) and Phillips-Perron (PP) unit root test are used to check for the stationary of bitcoin daily return series, and table 3.2 indicates that the financial time series is stationary.



(a) (b)

Figure 3.1: Bitcoin daily return and bitcoin daily squared return. (a) Bitcoin daily return; (b) Bitcoin squared daily return.

Description	BTC
Sample size	2944
Mean	0.000819
Variance	0.000343
Std. Dev.	0.018524
Skewness	-0.55305
Kurtosis	11.38964

Note: This table displays the summary statistics of bitcoin daily return from April 30, 2013 to May 21, 2021; BTC denotes bitcoin daily return..

Table 3.1: Summary Statistics of Bitcoin Daily Returns in the Sample Period.

	Without Trend		With Trend	
	ADF	PP	ADF	PP
BTC daily return	-55.1	-55.2	-55.2	-55.1
Critical values (1%)	-3.43	-3.43	-3.96	-3.96

Note: ADF and PP test statistics are much smaller than the 1% critical value, indicating the bitcoin daily return from April 30, 2013 to May 21, 2021 is stationary at 1% significance level.

Table 3.2: Unit root test.

3.2.1 Econometric methodology

Figure 3.1 shows notable fluctuations in bitcoin daily return. It is also found that large changes follow large turbulence and small changes follow calm periods. This phenomenon in time series asset return is known as "volatility clustering." The bitcoin daily squared return plot in Figure 3.1 provides more evidence that changes tend to be clustered together.

Table 3.3 shows the results of the Ljung-Box Q-test for the bitcoin daily squared return. Table 3 indicates that the bitcoin daily squared return is serially correlated, suggesting the existence of conditional heteroscedasticity in bitcoin price volatility. Thus, the econometric model needs to capture the feature of heteroscedasticity.

The basic structure of the econometric model is as follows:

$$r_t = \mu_t + Z_t \tag{3.1}$$

$$\mu_{t-1} = \mathbb{E}[r_t \mid \mathcal{F}_{t-1}] \tag{3.2}$$

$$h_t{}^2 = \mathrm{Var}[r_t \mid \mathcal{F}_{t-1}] = \mathbb{E}[(r_t - \mu_t)^2 \mid \mathcal{F}_{t-1}] = \mathbb{E}[Z_t{}^2 \mid \mathcal{F}_{t-1}] \tag{3.3}$$

Where $r_t = \log(\frac{p_t}{p_{t-1}})$ is the conditional mean and $h_t{}^2$ is the conditional variance, $\mathcal{F}_{t-1}$ denotes the past information.

No. of lags	Lag 10	Lag 15	Lag 20
P-value	0.000496***	0.000693***	0.000006***

Note: Triple asterisk (***) denotes variable significant at 1% level.

Table 3.3: Ljung-Box Q-Test for Bitcoin Daily Return.

3.2.1.1 Conditional mean

ARMA(p,q) process is applied to model the conditional mean:

$$\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{j=1}^{q} \theta_j Z_{t-j} \tag{3.4}$$

With the autoregressive order p and moving average order q. After applying the ARMA(p,q) process, the estimated parameters and the residuals are obtained. As discussed above, the bitcoin daily return exhibits volatility clustering, which indicates the conditional heteroscedasticity volatility. The ARCH effects of the residuals are tested. If there is an ARCH effect in the residuals, the conditional variance models will be specified in the next section.

3.2.1.2 Conditional variance

Given the conditional mean model and using equation (3), the residuals $Z_t = r_t - \mu_t$ are obtained. Then the conditional variance models can be built. Two different models are presented in the following section. It starts with EWMA model, then moves to the GARCH model to forecast bitcoin return volatility.

3.2.1.3 Exponentially weighted moving average (EWMA)

The exponentially weighted moving average is one of the simplest models for volatility forecasting. It models the time-varying variance and captures past information and historical variance. Although the exponentially weighted moving average model incorporates neither conditional mean nor conditional variance in the sense of GARCH, it is presented here as a benchmark to evaluate the performance of the other models.

The exponentially weighted moving average model is presented as:

$$\sigma_{t+1}{}^2 = \lambda \sigma_t{}^2 + (1 - \lambda) r_t{}^2 \tag{3.5}$$

Where λ is set to be 0.94 in RiskMetrics model, and $r_t{}^2$ is the bitcoin daily squared return.

3.2.1.4 GARCH model

Bollerslev developed the generalized autoregressive conditional heteroscedasticity model (GARCH) in 1986. Both the ARCH process and GARCH process model the variations of a financial assets' volatility, and the GARCH process allows the conditional variance to be an ARMA process. The GARCH process is as follows:

$$Z_t = h_t \epsilon_t, \epsilon_t \sim IID(0, 1) \tag{3.6}$$

$$h_t{}^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i Z_{t-i}{}^2 + \sum_{j=1}^{n} \beta_j h_{t-1}{}^2 \tag{3.7}$$

$\{Z_t\}$ is the residual series of the best-fitting ARMA(p,q) model. Thus, the conditional variance of the residual series essentially acts like an ARMA process. It is expected that the standardized squared residuals obtained from the best fitted ARMA-GARCH model should not be autocorrelated, and there should not remain any ARCH effects. The ARCH LM test is used to check whether this is true or not in this study.

3.2.2 Recurrent neural network (RNN)

The sequencing model for predicting bitcoin return volatility is built on the concept of Recurrent Neural Networks (RNN). RNN deals well with sequence problems and has a remarkable architecture that considers the order of data. Each RNN has a type of memory unit concatenated into multi-stages and each of which will turn previous states and current input to activations and pass necessary information forward to the next stage. In this study, a GRU (Gated Recurrent Units) cell is employed to serve as the memory unit. The cost function is redesigned based on a tangent function. This model does not build any embedding or probability layer inside usual configurations in some engineering tasks. Besides, by considering some uncertainty of the volatility, the range is equally cut into 250 intervals to convert a real volatility value to a vector with a dimension of 250. This conversion serves as an encoder for an RNN cell's input. The whole architecture of the RNN model is listed in Figure 3.2. In general, the encoding process will turn a fixed length of sequential data into the same length of vectors for RNN, fed into multiple layers of perceptron (MLP). The MLP will decode states from RNN into sequential vectors and transfer them to a predictor for output.

3.3 Results

In this section, the forecasting results of the EWMA model, GARCH model, and RNN model will be presented. Then, their out-of-sample forecasting accuracy performance will be evaluated and compared. Before the evaluation, appropriate proxies for the realized volatility have to be

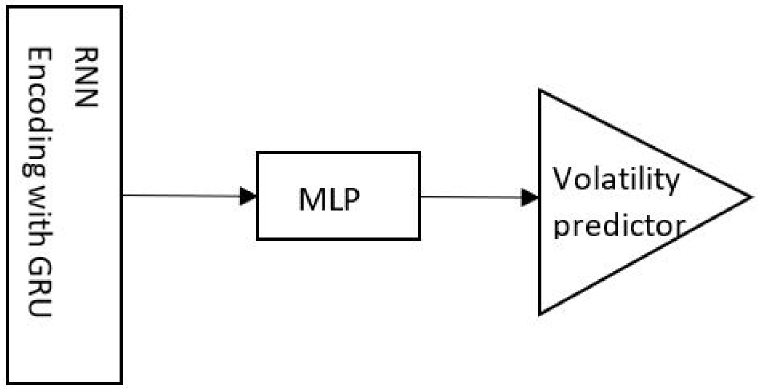


Figure 3.2: Architecture of recurrent neural network (RNN) model.

found.

3.3.1 Forecasting

The sample data is divided into two parts, the in-sample period from April 30, 2010, to August 02, 2020 (2652 observations) and the out-of-sample period from August 03, 2020, to May 21, 2021 (292 observations).

In the econometric GARCH model, the ARMA order is selected by AIC and BIC, and the best-fitted conditional mean model was found to be ARMA (2,2). Then, the residuals' ARCH effects are tested, and the result indicates there remains an ARCH effect in the residual series. Finally, the best-fitted ARMA -GARCH model is obtained. The ARMA-GARCH parameters are estimated in a rolling window. Table 3.4 presents the estimated parameters of the ARMA (2,2)-GARCH (1,2) model of the in-sample data.

The autocorrelation in the standardized residuals of the fitted ARMA-GARCH model is checked, and the result indicates that there is no remaining ARCH effect in the residuals.

For the RNN model, 30 days samples of the volatility are used to predict the next one day, five days, and ten days with an out-of-sample method. For example, the first 30 days of volatility values were used to predict the 31st. The sequential data generated by this process is called tuple 1; then, the 2nd to 31st volatility values are used to predict the 32nd, and it is called tuple 2. The

Parameters	Estimated value	t-value	p-value
ϕ_0	0.001	3.280	0.001***
ϕ_1	1.424	128.130	0.000***
ϕ_2	-0.434	-39.606	0.000***
θ_1	-1.458	-300530	0.000***
θ_2	0.474	1452	0.000***
α_0	0.000	1.851	0.064***
α_1	0.191	8.254	0.000***
β_1	0.413	2.965	0.003***
β_2	0.394	3.120	0.002***

Note: *, ** and *** denote variables significant at 10%, 5% and 1% levels respectively.

Table 3.4: ARMA (2,2)-GARCH (1,2) estimated parameters.

total data length was 2031. By rolling this process, 1994 tuples were generated. In the out-of-sample method, the first 1794 (90%) observations were appointed to training, and the remaining 200 observations were used as a test volume.

A more detailed implementation is illustrated in Figure 3.3. Two layers of RNN with GRU cells are built as a core. The first layer has 512 units, while the second shrink to 256 units. Sequential data were fed in cells on the bottom from left to right. The predicated data were collected on the top from left to right.

Both training and testing were taken on the GTX 1070 GPU. An SGD (Stochastic Gradient Descend) algorithm that shuffles the whole dataset is used in each iteration; the RMSProp gradient update algorithm was chosen as an optimizer; the learning rate and batch size were set to 0.0001 and 20, respectively. As stated before, the model 1000 epochs are trained on the 1794 tuples, and the 200 tuples are tested every five epochs.

3.3.2 Volatility proxies

One difficulty evaluating the forecasting performance is that the true conditional volatility of the bitcoin return is unobservable. Thus, a proxy for the realized bitcoin return volatility has to be found. The most commonly used proxy for volatility is the bitcoin daily squared return. Thus, the first volatility proxy in this study is the daily squared return. However, it may lead

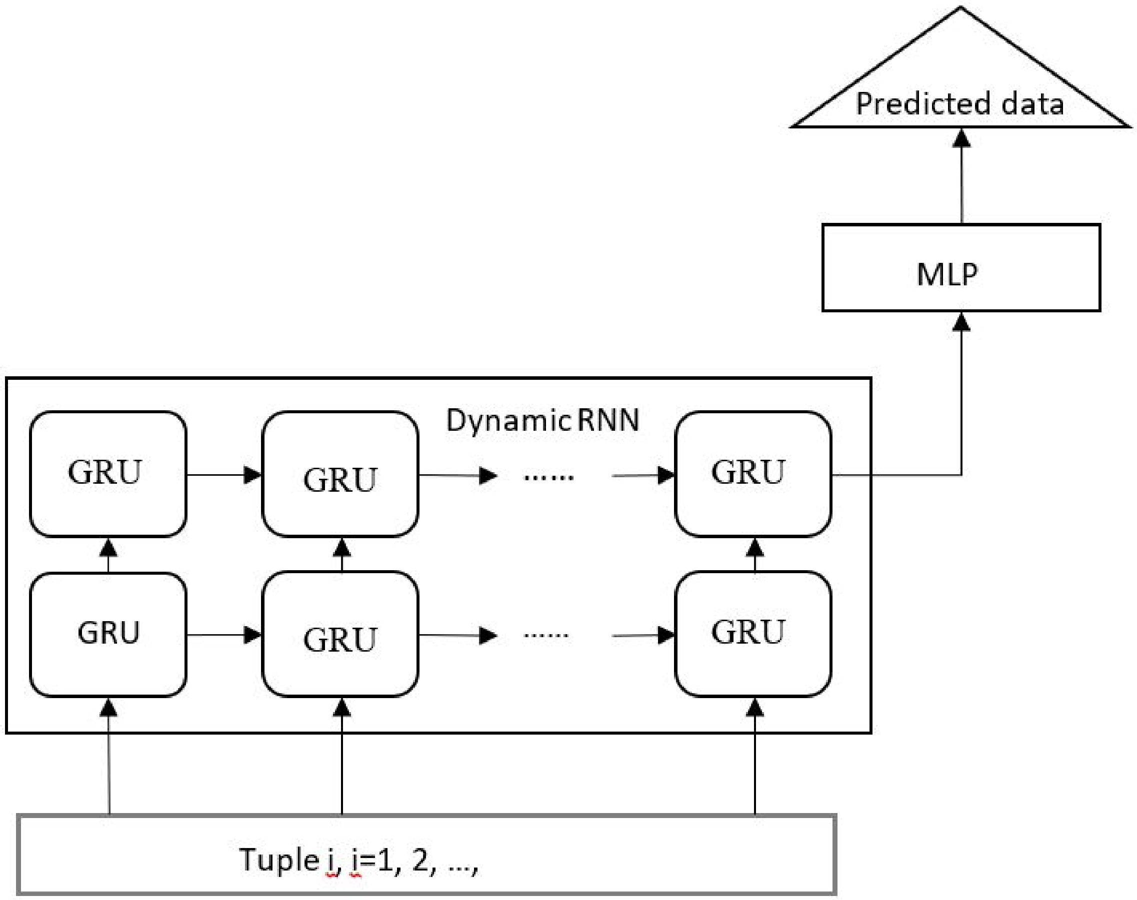


Figure 3.3: Detailed implementation of recurrent neural network (RNN) model.

to poor out-of-sample performance (Anderson and Bollerslev, 1998) [32]. To get a more robust forecasting performance comparison result, a second volatility proxy is necessary. The cumulative squared intra-day returns are a more efficient proxy for volatility (Chou et al., 2010) [33], but it requires high-frequency bitcoin prices in one day, which is not available in this case. Garman-Klass volatility (Garman and Klass, 1980) [30] is used as the second proxy for bitcoin return volatility. This proxy includes the information of daily high, low, opening, and closing prices. Garman and Klass (1980)'s estimator in practical is presented as:

$$\hat{\sigma}^2_{GK} = 0.5[\ln(\frac{BTC_{Ht}}{BTC_{Lt}})]^2 - [2\ln 2 - 1][\ln(\frac{BTC_{Ct}}{BTC_{Ot}})]^2 \tag{3.8}$$

Where BTC_{Ht} and BTC_{Lt} is the highest bitcoin price and lowest bitcoin price at the trading day; BTC_{Ct} and BTC_{Ot} is the closing price and opening price respectively.

3.3.3 Out-of-sample performance

To compare the three models' out-of-sample performance, the root mean squared error (RMSE) and mean absolute error (MAE) are used to evaluate and rank them. RMSE and MAE are the most commonly used metric for model evaluation. MAE is a good indicator of average model performance (Willmott & Matsuura, 2005) [34], while RMSE deals well with outliers by penalizing large errors more (Chai & Draxler, 2014) [35]. Table 3.5 exhibits the WEMA (benchmark) model, GARCH model, and RNN model's out-of-sample performance. One day ahead, five days ahead, and ten days ahead out-of-sample performances are presented in table 3.5.

	RMSE		MAE	
	1st Proxy	2nd Proxy	1st Proxy	2nd Proxy
1 day ahead				
EWMA	2.4797E-06	4.5816E-04	3.1640E-04	3.8195E-03
GARCH	**2.4795E-06**	**4.5814E-04**	3.3930E-04	**3.7762E-03**
RNN	2.4985E-06	4.5818E-04	**2.8091E-04**	4.0179E-03
5 days ahead				
EWMA	2.4969E-06	4.6213E-04	3.1742E-04	3.8721E-03
GARCH	**2.4760E-06**	**4.5814E-04**	3.4605E-04	**3.7607E-03**
RNN	2.5063E-06	4.5897E-04	**2.8533E-04**	4.0535E-03
10 day ahead				
EWMA	2.4779E-06	4.5817E-04	3.2214E-04	3.8277E-03
GARCH	**2.4741E-06**	**4.5813E-04**	3.5656E-04	**3.7426E-03**
RNN	2.5065E-06	4.5897E-04	**2.8426E-04**	4.048E-03

Note:The lowest RMSE and MAE of each model is in bold..

Table 3.5: Out-of-sample Performance.

Comparing the RMSE, the GARCH model performs best with the lowest RMSE, and the RNN model performs worst. When using the first proxy, the 1 (5, 10) day ahead RMSE of the RNN model are 0.76% (1.21%, 1.29%) larger than the GARCH model; when using the second proxy, the

1 (5, 10) day ahead RMSE of RNN model is 0.009% (0.18%, 0.18%) larger than GARCH model. Comparing the MAE, the RNN model outperforms GARCH and EWMA with the lowest MAE in the first proxy but is outperformed in the second proxy. When using the first proxy, the 1 (5, 10) day ahead MAE of the GARCH model is 17.21% (17.55%, 20.28%) larger than the RNN model; when using the second proxy, the MAE of RNN is 6.02% (7.22%, 7.54%) larger than GARCH model. It can be seen from Table 5, when using the first proxy, RNN performs better in MAE but performs poorly in RMSE. One explanation for this is that since the RMSE punishes more on the outliers than MAE, implying the RNN model generates more outliers than the econometrics models. Figure 3.4 presents the standard deviation of the first proxy (daily squared return) and the standard deviation of one day ahead volatility forecasting of each model. Figure 3.5 shows the standard deviation of the second proxy (Garman-Klass volatility) and the standard deviation of one day ahead volatility forecasting of each model. EWMA1/5/10, GARCH1/5/10, and RNN1/5/10 denote the 1/5/10 day(s) ahead of the volatility forecasted by each model.

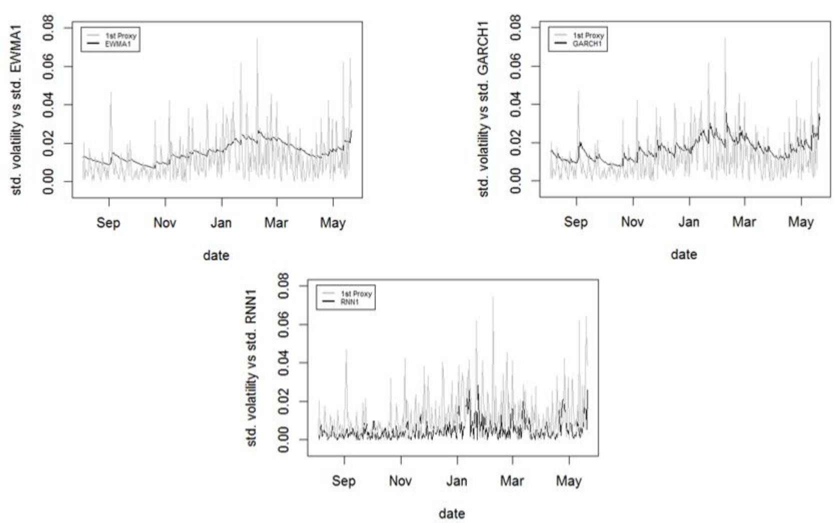


Figure 3.4: Out of sample standard deviation of realized volatility (1st proxy) vs standard deviation of one day ahead volatility forecasting of EWMA, GARCH and RNN model. EWMA1, GARCH1 and RNN1 denote the one day ahead forecasting of each model.

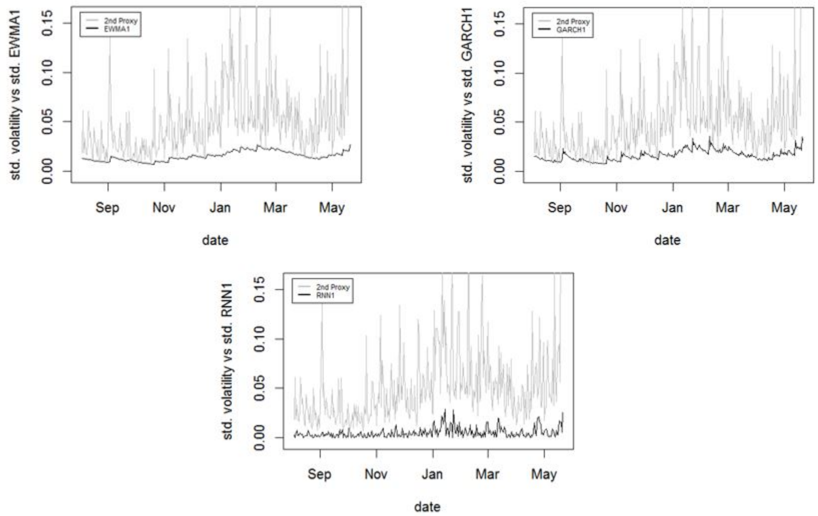


Figure 3.5: Out of sample standard deviation of realized volatility (2nd proxy) vs standard deviation of one day ahead volatility forecasting of EWMA, GARCH and RNN model. EWMA1, GARCH1 and RNN1 denote the one day ahead forecasting of each model.

It can be seen from Figure 3.4and Figure 3.5 that the RNN model does better in capturing the volatility trends and clustering than the econometric models[1] ; however, it underestimates the volatility. The second proxy is less volatile than the first one. Thus, the Garman-Klass volatility proxy does not perform as well as the bitcoin daily squared return proxy. The RNN model is not as efficient as we expected. It does better in corresponding to the volatility dynamics, but it underestimates the volatility and hurts the forecasting accuracy.

3.3.4 Value at Risk (VaR)

After analyzing the bitcoin volatility forecasting accuracy, it comes to the issue of risk management. Investors are also concerned about which method performs better in risk management. To provide more helpful information about the bitcoin market in terms of risk management, the Value at Risk (VaR) is a suitable measurement. The bitcoin Value at Risk is defined as the maximum amount of money that the bitcoin investors could lose at a given confidence level over a defined period. J.P. Morgan proposed the Value at risk in 1994. There are two categories of methodology

[1] The standard deviation of the first and second proxy and the standard deviation of the 5 days ahead and 10 days ahead volatility forecasting of each model are presented in appendix.

for VaR calculation, parametric models and nonparametric models. The most used nonparametric VaR model is the Monto Carlo method, which is complicated, while the parametric VaR is simpler for investors to apply. The parametric models are mainly considered and discussed in this article. Mathematically, the forecasted VaR of bitcoin is defined by:

$$P(r_t \leq VaR_t(\alpha) \mid \mathcal{F}_{t-1}) = \alpha \tag{3.9}$$

Where r_t is the bitcoin daily return, and $\mathcal{F}_{t-1}$ is the past information. The α is the given confidence level, 1%, 2.5% and 5%. Thus, the VaR estimation involves the assumption of bitcoin daily return distribution. The traditional VaR calculation assumes the portfolio return is normally distributed. However, it is empirically documented that financial asset return distributions always exhibit heavy tails. Table 3.1 lists the summary statistics for bitcoin daily returns, which shows there is excess kurtosis in the sample data. Therefore, the Student's t-distribution assumption is applied in this study instead of the normal one. The forecasting of daily bitcoin Value at Risk under Student's t-distribution is estimated as:

$$VaR_t(\alpha) = \mathbb{E}[r_t \mid \mathcal{F}_{t-1}] + {t_\alpha}^v \sigma_t \sqrt{\frac{v}{v-2}} \tag{3.10}$$

Where ${t_\alpha}^v$ is the Student's t-distribution critical value at confident level α (1%, 2.5% and 5%); $\mathbb{E}[r_t \mid \mathcal{F}_{t-1}]$ is the conditional mean generated by ARMA (2,2); σ_t is the standard deviation of volatility forecasted by the three models at time t. v is the estimated degree of freedom. Following Heikkinen and Kanto (2002) [36] and Andreev and Kanto (2004) [37], the degree of freedom is allowed to be non-integer. Applying method of moments, the consistent estimator of the degree of freedom is estimated by:

$$\hat{v} = 4 + \frac{6}{\hat{k}}, \forall v > 4 \tag{3.11}$$

Where $\hat{k}$ is the sample excess kurtosis.

Since the EWMA model and RNN model do not involve conditional mean, then the $\mathbb{E}[r_t \mid \mathcal{F}_{t-1}]$ of the two models are supposed to be zero. Therefore, the Value at Risk of EWMA and RNN models is estimated as:

$$VaR_t(\alpha) = t_{\alpha}{}^{v}\sigma_t\sqrt{\frac{v}{v-2}} \tag{3.12}$$

The sample excess kurtosis $\hat{k}$ is 8.39, then the estimated degree of freedom $\hat{v}$ is 4.72; the critical value $t_{\alpha}{}^{v}$ at confident level 1%, 2.5%, and 5% is -3.45, -2.62, -2.04, respectively.

Then the one day ahead, five days ahead and ten days ahead VaR at 1%, 2.5%, and 5% confidence level is calculated by the conditional expected return and forecasted volatility, which are generated from the EWMA model, ARMA (2,2)-GARCH (1,2) model and RNN model, respectively. Figure 3.6 presents the realized returns and one day ahead VaR (1%) [2] forecasts for each of the three models [3].

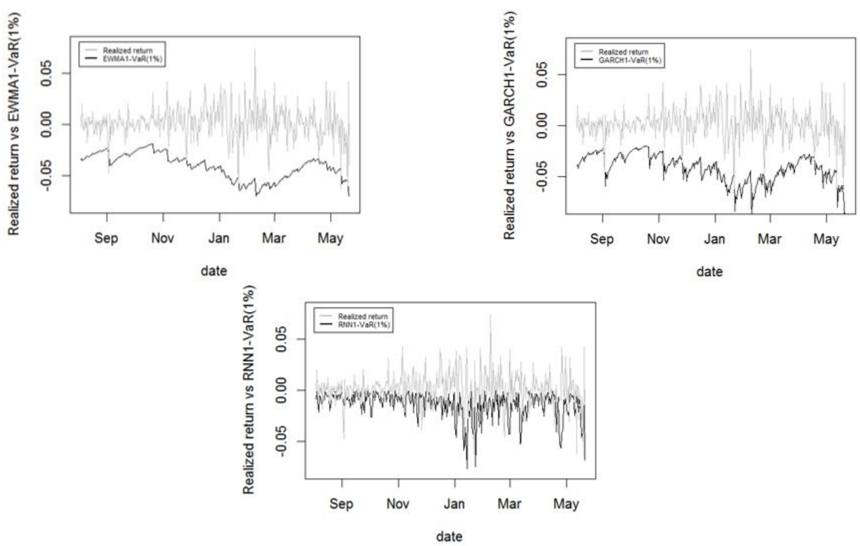


Figure 3.6: Out of sample realized return vs one day ahead VaR(1%) of EWMA, GARCH and RNN model. EWMA1, GARCH1 and RNN1 denote the one day ahead forecasting of each model.

Table 3.6 gives the out-of-sample coverage $\hat{\alpha}$ for each model to compare the Value at Risk forecasting performance. The $\hat{\alpha}$ is calculated by the number of realized losses that exceed the

[2]The figures of realized return and 5/10 days ahead VaR (1%) of each model are presented in appendix.

[3]The figures of realized return and 1/5/10 day(s) ahead VaR (2.5% and 5%) of each model are presented in appendix.

forecasted Value at Risk on that day divided by the number of totals out of sample observations:

$$\hat{\alpha} = \frac{No.(loss > VaR_t(\alpha))}{No.(\textit{out of sample obs.})} \tag{3.13}$$

	EWMA	GARCH	RNN
1 day ahead			
α=1%	**2.06%**	**2.06%**	16.15%
α=2.5%	**2.75%**	**2.75%**	18.9%
α=5%	4.47%	**5.15%**	23.37%
5 days ahead			
α=1%	**1.72%**	2.06%	21.31%
α=2.5%	3.44%	**2.75%**	25.43%
α=5%	6.19%	**4.47%**	26.80%
10 days ahead			
α=1%	**1.72%**	**1.72%**	18.90%
α=2.5%	3.78%	**2.41%**	23.71%
α=5%	**5.16%**	4.12%	28.18%

Note:The smallest $|\alpha - \hat{\alpha}|$ in each model is in bold.

Table 3.6: Out of Sample Coverage of Each Model.

The closer $\hat{\alpha}$ to α, the more accurate VaR would be, making it easier for investors to manage the bitcoin market risk. Thus, the model with the smallest $|\alpha - \hat{\alpha}|$ provides the best risk coverage.

Table 3.6 indicates that the sample coverage $\hat{\alpha}$ of the GARCH model is closest to the given confidence level α, while the sample coverage of the RNN model appears to be the most volatile, with the largest distance between $\hat{\alpha}$ and α. Thus, the RNN model performs poorly in Value at Risk forecasting, even worse than the benchmark EWMA model.

After getting the forecasted Value at Risk of each model, two approaches to Value at Risk backtesting are conducted: unconditional coverage test and conditional coverage test.

The unconditional coverage test was proposed by Kupiec in 1995. It tests whether the violation rate of the Value at Risk model is equal to the theoretical rate. Christoffersen introduced the conditional coverage test in 1998, which examines whether the VaR violation process is independent. Table 3.7 reports the VaR backtesting results of each model.

	1 day ahead		5 days ahead		10 days ahead	
0.99 Quantile						
EWMA	0.11	0.25	0.26	0.49	0.26	0.49
GARCH	0.11	0.25	0.11	0.25	0.26	0.49
RNN	0.00	0.00	0.00	0.00	0.00	0.00
0.975 Quantile						
EWMA	0.79	0.43	0.33	0.40	0.20	0.29
GARCH	0.79	0.43	0.79	0.43	0.92	0.84
RNN	0.00	0.00	0.00	0.00	0.00	0.00
0.95 Quantile						
EWMA	0.67	0.80	0.37	0.67	0.90	0.94
GARCH	0.90	0.96	0.67	0.80	0.48	0.63
RNN	0.00	0.00	0.00	0.00	0.00	0.00

Note:LR_uc denotes the p value of unconditional coverage Kupiec test and LR_cc denotes the conditional coverage Christoffersen test.

Table 3.7: VaR backtesting results of unconditional coverage test of Kupiec and conditional coverage test of Christoffersen -p value.

The p-value of the EWMA model and GARCH model show that they fail to reject the null hypothesis, indicating the VaR is correctly estimated. However, the RNN model fails in the two tests.

3.4 Conclusion

Bitcoin is the most successful and popular cryptocurrency in the market, with around 130 billion daily trading volumes as of April 2019. Bitcoin has historically had more significant fluctuations in price than most other financial assets. Therefore, the analysis of bitcoin return volatility is crucial for investors' decision-making and risk management. Both economic models and the machine learning method are used to forecast the bitcoin return volatility.

The machine learning method in time series forecasting is expected to be superior to the traditional econometrics models. The earlier empirical studies in stock price forecasting and cryptocurrency prices forecasting provided evidence of this statement. By comparing the out-of-sample performance of each volatility forecasting model, the result indicates that the RNN model is more sensitive and corresponds more quickly to the volatility change than the traditional econometrics models.. The RNN outperforms GARCH and EWMA in MAE evaluation criteria in forecasting accuracy but is overwhelmed in RMSE criteria. Since MAE does well in average model performance, while RMSE provides more information on outliers, the two opposite performances could be regarded as evidence that the RNN model is less efficient in capturing the bitcoin market extreme events.

In addition to the bitcoin volatility forecasting, the RNN model is outperformed by the GARCH and EWMA model in risk management efficiency in the framework of Value at Risk. The Value at Risk essentially focuses on the tail events of bitcoin return. Therefore, the RNN's poor performance in VaR provides another evidence of the robust results that it could not work well as econometric models in explaining extreme volatility. It underestimates the fluctuations in the more volatile price period. In other words, it underestimates the risk. This result is consistent with the previous study that Seq2Seq RNNs improve the bitcoin price forecasting accuracy over the ARIMA model during less volatile periods but shows poor performance in extreme cases (Rebane et al., 2018) [38].

This study proposed an alternative way of volatility analysis. It is widely believed and empirically proved by earlier studies in the financial market that the machine learning approach is more advanced in time series forecasting. However, this study shows something different. The RNN model is less efficient than traditional econometric models in bitcoin volatility forecasting and risk management. The econometric models are superior in analyzing extreme market conditions, while the machine learning approach is more suitable for less volatile market conditions.

Many investors have considered including bitcoin in their investment portfolio. This study provides implications for the investors on trading strategy and risk management. For the financial institutions required to hold sufficient risk capital to cover potential losses on the portfolio, the econometric models are recommended for a good Value at Risk estimation.

4. ESTIMATION OF TAIL RISK MEASURES IN CRYPTOCURRENCY MARKET

4.1 Introduction

The cryptocurrency market has been proliferating since bitcoin was invented in 2009 (Nakmoto 2009 [39]). As of February 2020, there exit approximately 5,360 types of cryptocurrencies that are being traded with over 300 billion USD market capitalization. This new and growing market has attracted significant attention and involves plenty of participants, including individuals, institutions, and governments. In this regard, analyzing the cryptocurrency market is crucial for investors and government regulations. In this study, we focus on the risk management analysis of the cryptocurrency market.

The main target of risk management in the cryptocurrency market is investigating its loss tail behavior. Cryptocurrency prices show more significant fluctuations than traditional financial assets and exhibit extreme tail events. Another issue is risk measurement. The most modern and practical approach to measuring financial asset risk is based on asset loss distribution. It makes perfect sense in the cryptocurrency market because, for cryptocurrency holders, risk management's primary interest is their losses. Value at Risk (VaR) and conditional Value at Risk is the most commonly used risk measurements based on loss distributions (Alexander et al., 2012 [40]). Conditional Value at Risk is also called Expected Shortfall (ES). VaR and ES are essentially the quantiles of asset loss distribution over a given period.

Researchers have frequently discussed cryptocurrency tail-risk characteristics. Prices for Cryptocurrencies are shown to be non-normal and exhibit heavy tail (Osterrieder 2017 [19]; Phillip et al., 2018 [41]). Osterrieder and Lorenz (2017) [42] analyze the bitcoin return's tail behavior using extreme value theory and comparing it with G10 currencies. They find bitcoin is more volatile and is risker than traditional fiat currencies. Gkillas and Katsiampa (2018) [43] also use extreme value theory to study five major cryptocurrencies' tail behavior instead of only bitcoin. They rank the five cryptocurrencies from highest risky to least risky according to Value at Risk and Expected

Shortfall estimation and find that bitcoin cash is the riskiest while bitcoin and Litecoin are the least risky cryptocurrencies.

This study analyzes the cryptocurrency market risk by calculating and forecasting the cryptocurrency index's VaR and ES. Accurate forecasting of VaR and ES are crucial for cryptocurrency market risk management since it helps investors and other participators better understand their risk exposures and make more informed decisions.

The commonly used method is to apply the standard parametric approach to the full loss distribution. However, not only from the full distribution, but we also intend to grab information from the extreme tail events. Therefore, this study focuses on tail events. An ARMA process was considered for the conditional mean loss. VaR and ES calculation's main challenge is to estimate the tail of cryptocurrency index conditional loss distribution. On an excellent tail estimation, extreme value theory (EVT) may be a proper parametric method. However, the iid assumption makes EVT an inappropriate approach for cryptocurrency loss data. To address this issue, we filter the asset loss by using a GARCH model. It is observed that the time series of cryptocurrency loss is serially correlated and presents volatility clustering, which implies the existence of conditional heteroscedasticity in volatility. We take it into account by employing the GARCH model under different probability distribution assumptions. The filtered residual will satisfy the iid assumption of EVT modeling. Then the task of asset conditional loss distribution estimation could be transferred to the standard innovation distribution estimation. The typical approach of EVT estimation is Peak Over Threshold (POT), which fits a general Pareto distribution (GPD) to the standard innovations that filtered from the GARCH process (Alexander et al., 2012 [40]). This GARCH-EVT method is essentially a conditional EVT approach developed by McNeil and Frey (2000) [44]. Few attempts were made on cryptocurrency markets using this method. Some recent studies apply EVT in bitcoin and other cryptocurrencies without considering conditional volatility background.

However, there is an issue with the POT approach because the researcher has to choose an appropriate threshold u. The choice of threshold u may affect the tail parameters estimation. Therefore, we alternatively employ a semiparametric approach, the filtered historical simulation (FHS), in this study. FHS combines the conditional variance model GARCH with a historical simulation method for standard innovations (Christoffersen, 2011 [45]). It releases assumptions about the innovation tail distribution. By calculating one day ahead value at risk and expected shortfall, we compare the performance of parametric GARCH-EVT with semiparametric FHS.

To show a clear preference for the models that focus on tail events, we also apply a standard parametric approach that makes a specific probability distribution assumption to the full distribution and compares its performance with GARCH-EVT, and FHS approaches. Moreover, we evaluate EVT models' performance under both conditional and unconditional background to show a significant improvement in considering conditional background. The unconditional EVT method fits GPD with raw loss data instead of the data filtered from the GARCH process. Besides, we also examine each conditional model's sensitivity to the choice of probability distribution with Gaussian, Student t, and generalized error assumption.

This study contributes to cryptocurrency market risk analysis in three ways: 1. It uses recent cryptocurrency index (CRIX) data rather than just bitcoin or any specific type of cryptocurrency, which is more generalized and valuable for cryptocurrency investors; 2. This study is not the first one to examine the tail behavior of cryptocurrency prices. However, it is the first study that allows one to discriminate among different approaches; 3. We evaluate the improvement of involving conditional background to the unconditional one.

4.2 Methodology

In this section, we introduce VaR and ES measurements and then describe the three steps in VaR and ES calculation.

4.2.1 VaR and ES approach to risk measurement

Value at Risk is defined as the maximum amount of money that the asset holder could potentially lose at a given confidence level over a defined period. Mathematically, given the asset loss l_t, confidence level α and past information $\mathcal{F}_{t-1}$, $P(l_t \leq VaR_t(\alpha) \mid \mathcal{F}_{t-1}) = \alpha$. Here we define the loss as negative return, $l_t = -r_t$, where $r_t = \ln P_t - \ln P_{t-1}$ and P_t is the cryptocurrency index at time t. Assuming that we are forecasting one day ahead VaR, at confidence level α, and $l_{t+1} = \mu_{t+1} + \sigma_{t+1} z_{t+1}$, where μ_{t+1} is the forecasted conditional mean at time t+1 that is obtained from ARMA process, ${\sigma_{t+1}}^2$ is the conditional variance forecasted by GARCH process, and z_{t+1} is the standard innovations of GARCH process which is white noise time series. Now we can write the one day ahead VaR at level α as $P(\mu_{t+1} + \sigma_{t+1} z_{t+1} \leq VaR_{t+1}(\alpha) \mid \mathcal{F}_t) = \alpha$, or $P(z_{t+1} \leq (\frac{VaR_{t+1} - \mu_{t+1}}{\sigma_{t+1}}) \mid \mathcal{F}_t) = \alpha$. By assuming the loss distribution follows a given probability distribution F that $F(\frac{VaR_{t+1} - \mu_{t+1}}{\sigma_{t+1}} \mid \mathcal{F}_t) = \alpha$, then we can rewrite

$$VaR_{t+1}(\alpha) = \mu_{t+1} + \sigma_{t+1} F^{-1}(\alpha) \tag{4.1}$$

An alternative risk measurement related to VaR is conditional VaR. Conditional VaR is also known as Expected Shortfall, and we denote it as ES in this study. Compared to VaR, ES looks further into the tail of loss distribution by measuring the expected loss, which exceeds VaR. Mathematically, $ES_{t+1}(\alpha) = \mathbb{E}_{t+1}[l_{t+1} \mid l_{t+1} \geq VaR_{t+1}(\alpha)]$ and can be rewrite as

$$ES_{t+1}(\alpha) = \mu_{t+1} + \sigma_{t+1} \mathbb{E}_{t+1}[z_{t+1} \mid z_{t+1} \geq F^{-1}(\alpha)] \tag{4.2}$$

As we can see from equations (4.1) and (4.2), the following three steps to VaR and ES forecasting are forecasted conditional mean, forecasted conditional variance, and the estimated probability distribution of standard innovations.

4.2.2 Conditional mean and conditional variance

For the given loss series $l_1, , l_T$, we estimate the conditional mean and conditional variance

$$l_t = \phi_0 + \sum_{i=1}^{m} \phi_i l_{t-i} + \sum_{j=1}^{n} \theta_j \varepsilon_{t-j} + \varepsilon_t, \varepsilon_t = z_t \sigma_t \tag{4.3}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{4.4}$$

With z_t is iid with the distribution of Gaussian, student – t and generalized error.

4.2.3 Parametric approach

McNeil and Frey (2000) [44] first combine the GARCH process with the EVT method and implement it to S&P 500 return to show its advantage in modeling volatility dynamics and extreme tail behavior of financial time series. In our study, as we have obtained the filtered time series data standard innovation z_t, under the iid assumption, with a given threshold u, we consider the excess innovations over threshold u, $Y = Z - u$, follow a generalized Pareto distribution $GPD(\xi, \beta)$, where ξ is the tail shape parameter and β is the tail scale parameter with distribution function:

$$GPD_{\xi,\beta}(y) = f(x) = \begin{cases} 1 - (1 + \frac{\xi y}{\beta})^{-\frac{1}{\xi}} & \text{if} \quad \xi \neq 0 \\ 1 - \exp(-\frac{y}{\beta}) & \text{if} \quad \xi = 0 \end{cases} \tag{4.5}$$

Where the support is $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\frac{\beta}{\xi}$ when $\xi \ < 0$, and $\beta > 0$.

Now suppose the general distribution function of Z is F, then the distribution of excess innovations over threshold u is

$$F_u(y) = P(Y \leq y \mid Z > u) = P(Z \leq y + u \mid Z > u) = \frac{F(y+u) - F(u)}{1 - F(u)} \tag{4.6}$$

After fitting the filtered z_t to GPD, the estimated $\hat{\xi}$ and $\hat{\beta}$ could be obtained by using MLE. Notably, we would get $\hat{\xi} > 0$, which corresponds to the heavy tail distribution features. Therefore, the estimated excess innovations distribution is

$$F_u(y) = 1 - (1 + \frac{\hat{\xi} y}{\hat{\beta}})^{-\frac{1}{\hat{\xi}}} \tag{4.7}$$

From equation 4.6, the standard innovation distribution could be rearranged by putting F to the left-hand side:

$$F(z) = F(u) + (1 - F(u))F_u(z - u) \tag{4.8}$$

If we estimate $1 - F(u)$ by the proportion of the tail filtered data $\frac{N_u}{N}$, where N_u is the number of excess innovations over threshold u and N is the number of innovations, and we estimate $F_u(z-u)$ with the GPD, we will have the estimated standard innovation distribution:

$$\hat{F(z)} = 1 - \frac{N_u}{N}(1 + \frac{\hat{\xi}(z-u)}{\hat{\beta}})^{-\frac{1}{\hat{\xi}}} \tag{4.9}$$

Therefore, we can forecast one day ahead VaR and ES at level α by:

$$V\hat{a}R_{t+1}(\alpha) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{F}^{-1}(\alpha) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}(u + \frac{\hat{\beta}}{\hat{\xi}}[(\frac{1-\alpha}{\frac{N_u}{N}})^{-\hat{\xi}} - 1]) \tag{4.10}$$

$$E\hat{S}_{t+1}(\alpha) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}(\frac{1}{1-\alpha}\int_{\alpha}^{1}\hat{F}^{-1}(S)ds) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}[\frac{\hat{F}^{-1}(\alpha)}{1-\hat{\xi}} + (\frac{\hat{\beta}-\hat{\xi}u}{1-\hat{\xi}})] \tag{4.11}$$

One of the critical issues of POT is to find the appropriate threshold u. The selection of threshold u involves a tradeoff of bias and variance. EVT might not hold if u is too low, in which case GPD may not sufficiently define the excess data; however, if u is too high, there are only a few observations of excess data left, which increases the variance of the estimated tail shape parameter. Here we use the linearity of the mean excess function to choose the appropriate u. It is found that if the excess distribution follows a GPD for a given threshold u_0, then the excess distribution remains GPD with the same tail shape parameter ξ for the threshold $u > u_0$, but the scale parameter β grows linearly with the threshold u. As ξ is less than 1, we could get the mean excess function

$$e(u) = \mathbb{E}(z-u \mid z > u) = \frac{\beta + \xi(u-u_0)}{1-\xi} \tag{4.12}$$

Hence, we use the linearity of the mean excess function to select threshold u. We plot the sample mean excess function. If the filtered data admits GPD over the threshold and $\xi > 0$, then the plot should be an upward trend line. We should not expect a perfect line in practice, but we might choose u towards the beginning part of the linear section if the plot is visually linear. The sample mean excess is estimated by

$$e_{N_u}(u) = \frac{\sum_{i=1}^{N_u}(z_i - u)}{N_u} \tag{4.13}$$

Where $z_i > u$ and N_u denotes the number of observations that exceed selected threshold.

To show the superiority of EVT in tail estimation, we compare it with non EVT models, which apply the standard parametric approach to the full distribution. We assume Gaussian, student t, and generalized error to the loss distribution and calculate one day ahead VaR and ES by:

$$V\hat{a}R_{t+1}(\alpha) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\Phi^{-1}(\alpha) \tag{4.14}$$

$$V\hat{a}R_{t+1}(\alpha) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}t_{\hat{v}}^{-1}(\alpha) \tag{4.15}$$

$$V\hat{a}R_{t+1}(\alpha) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}g_{\hat{\beta}}^{-1}(\alpha) \tag{4.16}$$

$$\hat{ES}_{t+1}(\alpha) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\frac{1}{1-\alpha}\int_{\alpha}^{1} q_u(F_L)du \tag{4.17}$$

Where Φ denotes the Gaussian df; $t_{\hat{v}}$ denotes the student t df with estimated degree of freedom v; $g_{\hat{\beta}}$ denotes the generalized error df with estimated shape parameter β; $q_u(F_L)$ denotes the quantile function of the underlying distribution function.

4.2.4 Semiparametric approach

The attractive feature of the filtered historical simulation (FHS) approach is that it captures volatility dynamics while does not need to make a specific distribution assumption about the standard innovations. In our univariate case, as the estimated standard innovation set $\{\hat{Z}_t\}$ is obtained, the one day ahead VaR can be calculated by empirical quantile estimation:

$$V\hat{a}R_{t+1}(\alpha) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}Percentile\{\{\hat{z}_t\}_{t=1}^{N}, 100\alpha\} \tag{4.18}$$

$$\hat{ES}_{t+1}(\alpha) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}(\frac{1}{N(1-\alpha)})\sum_{i=1}^{N}\hat{z}_{l,t+1}\mathbb{I}\{\hat{z}_{l,t+1} > \frac{V\hat{a}R_{t+1}(\alpha) - \hat{\mu}_{t+1}}{\hat{\sigma}_{t+1}}\} \tag{4.19}$$

4.3 Data and empirical results

We use the daily Cryptocurrency Index (CRIX) in this study. Simon and Wolfgang (2018) proposed a method to construct a dynamic index in the cryptocurrency market, named CRIX [1] , and claimed CRIX is a good index as an investable and accurate benchmark for cryptocurrencies. This study uses CRIX ranges from 8/1/2014 to 3/31/2020, with 2071 observations. The CRIX data is available from the website: https://thecrix.de/. The data source is from Humboldt-Innovation GmbH, CRIX- Network. We use the CRIX index from 8/1/2014 to 9/12/2019 as in-sample data and 9/13/2019 to 3/31/2020 as out-of-sample data. The ratio is approximately 9:1.

4.3.1 ARMA-GARCH process

It is seen from figure 4.1 that the CRIX daily loss is extraordinarily volatile and that there exists volatility clustering. We did the ADF test and Ljung Box test and found that the in-sample time series is stationary but serially correlated. An ARMA (2,1)-GARCH (1,2) model is implemented for the conditional mean and conditional variance. The estimated parameters with Gaussian, Student t, and generalized error innovation assumptions are shown in table 4.1.

After obtaining the standard innovations from ARMA-GARCH, we test for the ARCH effect. Ljung Box test shows that the standard innovations are not serially correlated, which indicates that no ARCH effect remains. The standard innovation is filtered CRIX daily loss that satisfies the iid assumption.

4.3.2 EVT-POT

Before fitting GPD to the filtered time series, we need to select an appropriate threshold u by using the linearity property of the mean excess function in equation 4.13

[1] "CRIX an Index for Cryptocurrencies" by Simon T. and Wolfgang K. H. detailly explains how CRIX is constructed and its application in the German stock market and Mexican stock market.

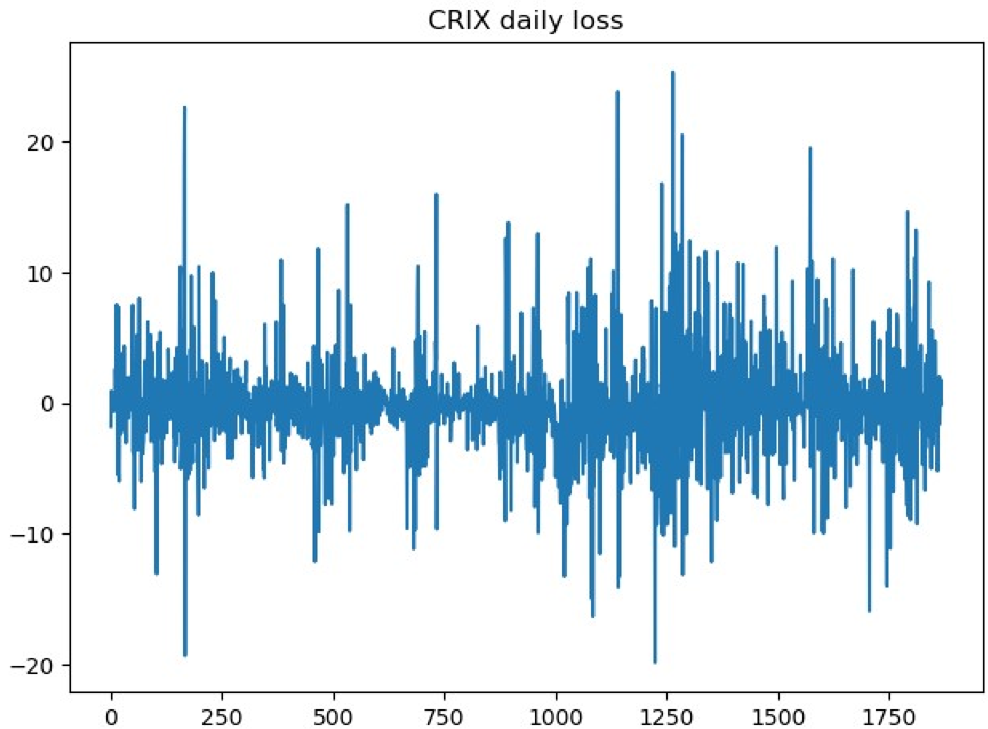


Figure 4.1: Plot of in sample CRIX daily loss.

	Gaussian			Student t			Generalized Error		
	Estimates	S.E.	p-value	Estimates	S.E.	p-value	Estimates	S.E.	p-value
ϕ_0	-0.170	0.138	0.218	-0.401	0.139	0.004	-0.269	0.020	0.000
ϕ_1	0.993	0.007	0.000	0.932	0.009	0.000	0.913	0.002	0.000
ϕ_2	0.001	0.007	0.886	0.060	0.009	0.000	0.076	0.001	0.000
θ_1	-0.988	0.000	0.000	-0.976	0.000	0.000	-0.976	0.001	0.000
ω	0.733	0.146	0.000	0.366	0.140	0.009	0.360	0.129	0.005
α_1	0.198	0.026	0.000	0.212	0.032	0.000	0.215	0.037	0.000
β_1	0.420	0.095	0.000	0.454	0.141	0.001	0.419	0.132	0.002
β_2	0.354	0.083	0.000	0.333	0.120	0.006	0.366	0.116	0.002

Table 4.1: Parameter estimates for ARMA (2,1)-GARCH (1,2) model.

Figure 4.2 plots the sample mean excess function of filtered CRIX daily loss with threshold u from 0 to 6. As expected, the sample mean excess plots are visually linear with upward trends at the beginning part, which indicates the GPD model is appropriate, and its shape parameter is

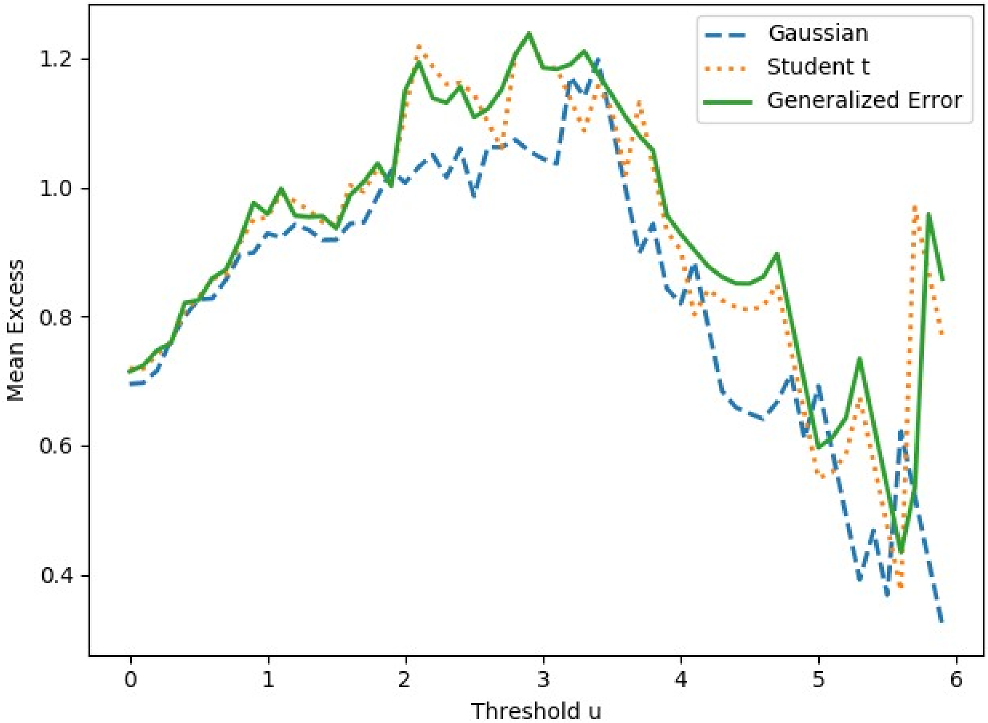


Figure 4.2: Estimated sample mean excess.

positive in this study. The appropriate threshold u should be selected along the upward line and towards the beginning section. As mentioned, the selection of threshold u involves a tradeoff of bias and variance. As u moves from low to high, the estimation variance increases, and bias is reduced. To address this issue, we exam the sensitivity of VaR and ES forecasting to threshold u choice. For each innovation distribution assumption, we select different u that corresponds to a quantile of 0.15, 0.1, and 0.05, with 280, 187, 93 exceedances, respectively. These selected thresholds are in the appropriate range. After u is chosen, we fit filtered excess loss data to GPD and estimate the parameters by using MLE. Table 4.2 gives the estimated tail shape parameter $\hat{\xi}$ and tail scale parameter $\hat{\beta}$ with a different u.

The estimated shape parameters $\hat{\xi}$ are positive, implying a heavy tail of filtered data under each assumption.

Recall the estimated excess innovation distribution function in equation 4.7, where y denotes

	Gaussian	Student t	Generalized Error
u (0.15)	0.715	0.770	0.780
$\hat{\xi}$	0.096	0.138	0.124
$\hat{\beta}$	0.786	0.766	0.791
u (0.10)	1.010	1.060	1.050
$\hat{\xi}$	0.031	0.037	0.064
$\hat{\beta}$	0.910	0.947	0.902
u (0.05)	1.640	1.710	1.700
$\hat{\xi}$	0.039	0.087	0.095
$\hat{\beta}$	0.926	0.919	0.915

Table 4.2: Estimated EVT-POT Parameters.

the exceedance over threshold u that $Y = Z - u$.We plot the estimated excess innovation distribution with the empirical excess distribution under each assumption in Figures 3, 4, and 5.

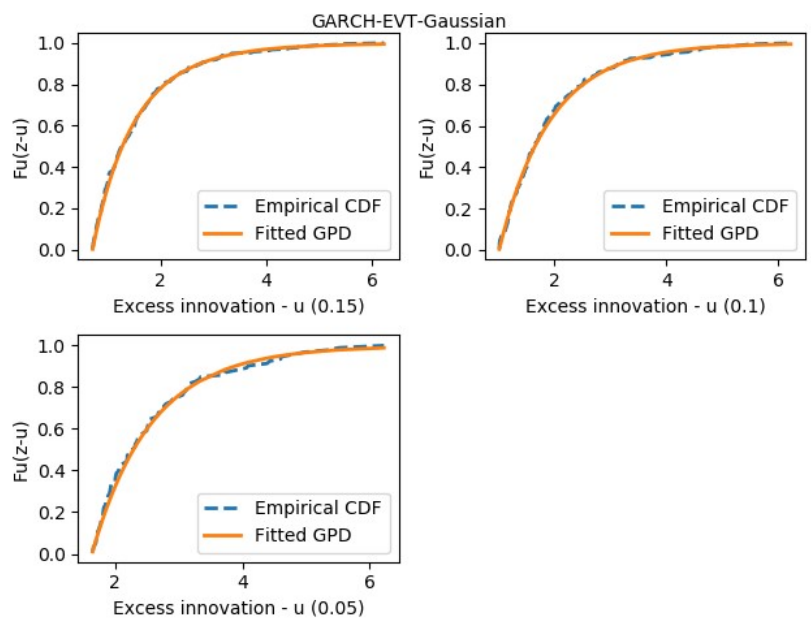


Figure 4.3: Estimated excesses vs empirical excess (GARCH-EVT-Gaussian).

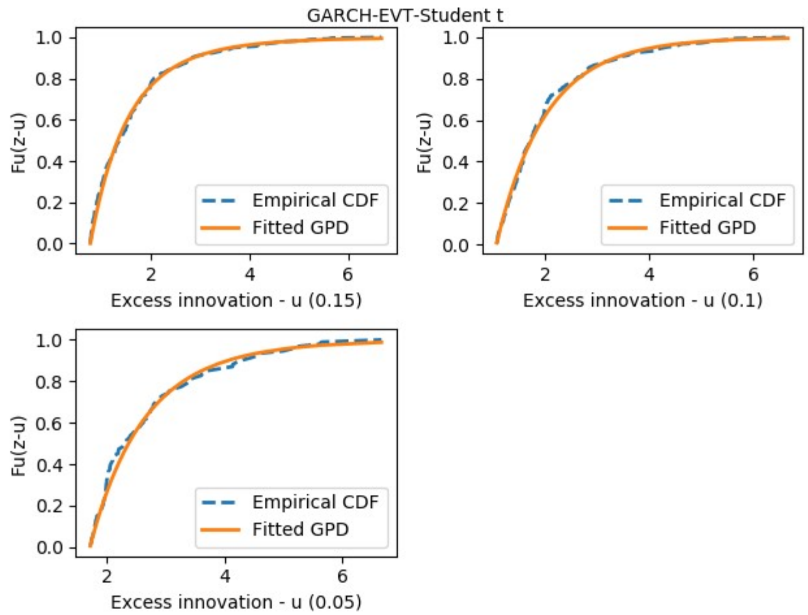


Figure 4.4: Estimated excesses vs empirical excess (GARCH-EVT-Student t).

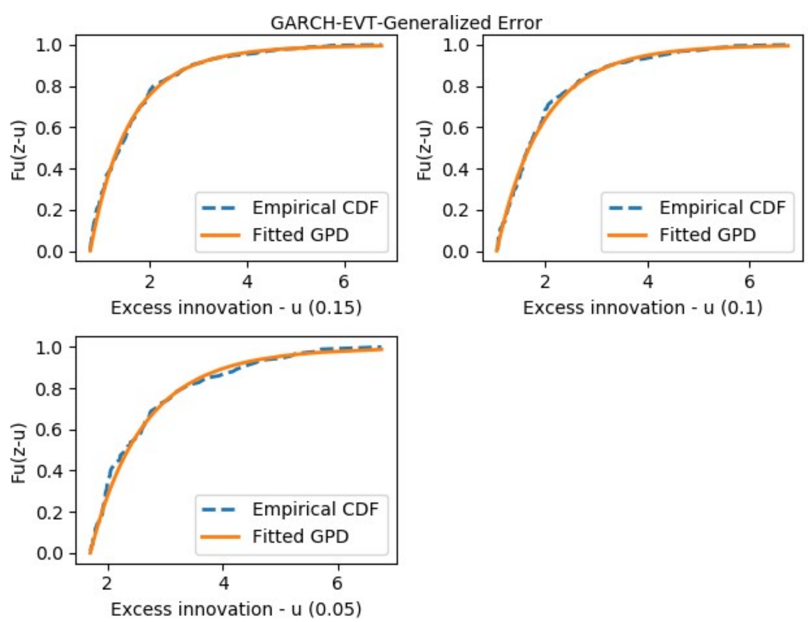


Figure 4.5: Estimated excesses vs empirical excess (GARCH-EVT-Generalized Error).

We can see from Figures 4.3, 4.4, and 4.5 that the fitted GPD curve fits the empirical excess innovations well, especially when the chosen threshold u corresponds to 0.15 quantile.

As we have mentioned before, the filtered standard innovations tail F(z) can be estimated by $F_u(z-u)$, after rearrangement, see equation 4.9

Figure 4.6, 4.7, 4.8 show the estimated tail probabilities with the empirical tail probabilities. The correspondence between empirical tail probabilities and the estimated tail probabilities is close when the chosen threshold u corresponds to 0.15 and 0.10 quantile. Also, we could not see a significant difference in probability distribution assumptions from these figures.

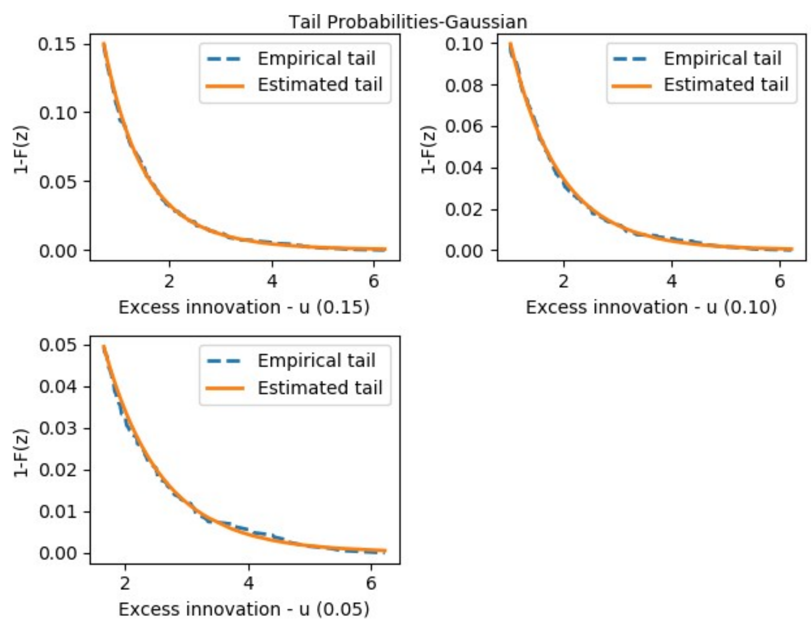


Figure 4.6: Estimated tail probabilities vs empirical tail probabilities (GARCH-EVT-Gaussian).

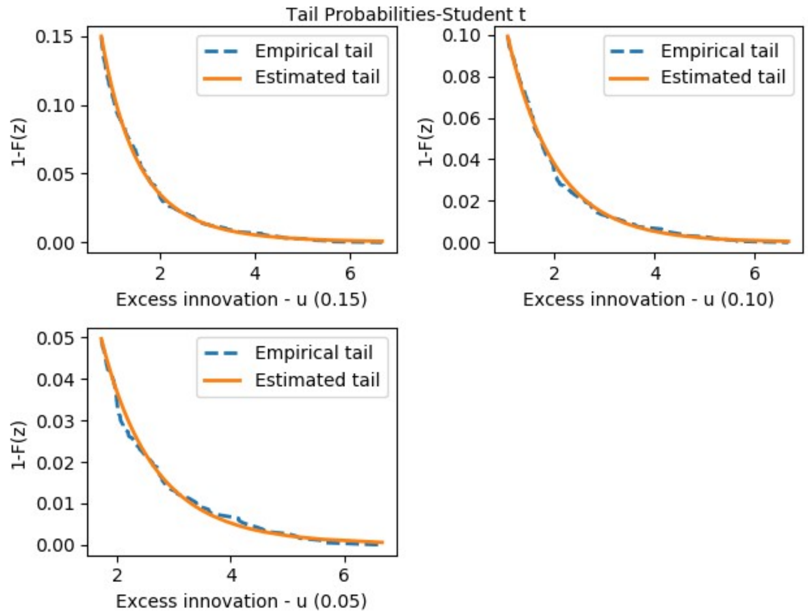


Figure 4.7: Estimated tail probabilities vs empirical tail probabilities (GARCH-EVT-Student t).

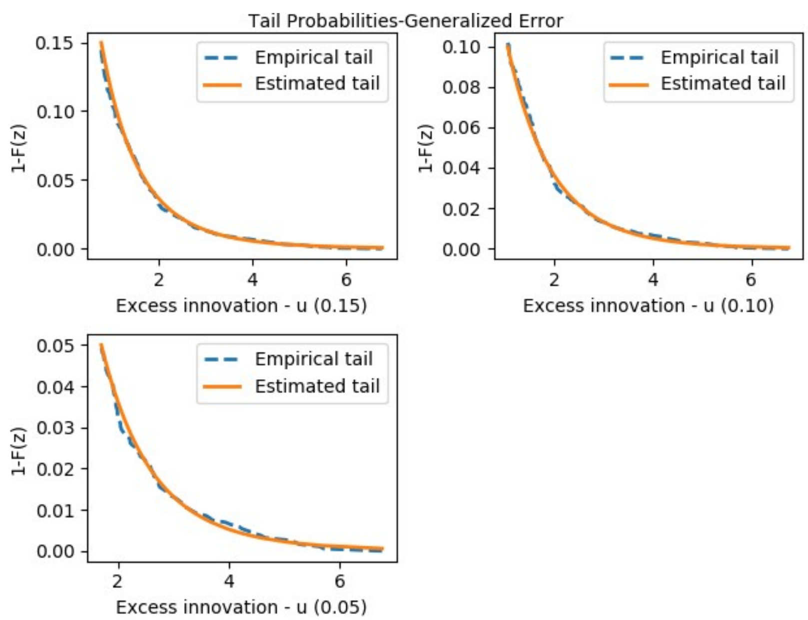


Figure 4.8: Estimated tail probabilities vs empirical tail probabilities (GARCH-EVT-Generalized Error).

4.3.3 Filtered historical simulation (FHS)

The filtered historical simulation (FHS) assumes the obtained standard innovations that its past distribution could estimate tomorrow's filtered loss distribution. Therefore, we could calculate one day ahead of VaR and ES by using empirical quantile estimation (see equation 4.18, 4.19). Table 4.3 gives the filtered historical loss quantile estimation.

Quantile	0.99	0.975	0.95
Gaussian	3.137	2.259	1.630
Student t	3.394	2.302	1.708
Generalized error	3.303	2.317	1.697

Table 4.3: FHS quantile estimation.

4.3.4 Unconditional EVT

To evaluate the improvement of the involving conditional background to the unconditional one, we apply EVT to unconditional data. In other words, we fit the raw daily loss data to GPD instead of filtered data. Table 4.4 gives the estimated parameters under different thresholds u.

	u (0.15)	u (0.10)	u (0.05)
u	2.45	3.83	6.25
$\hat{\xi}$	0.028	0.030	-0.003
$\hat{\beta}$	3.530	3.580	3.896

Table 4.4: Estimated Unconditional EVT-POT Parameters.

We plot the estimated excess innovation distribution with the empirical excess distribution in figure 4.19 and estimated tail probabilities with the empirical tail probabilities in figure 4.10. Figures 4.9 and 4.10 find that the estimated distribution fits empirical distribution well when u is small but does not fit well when threshold u gets more extensive. In other words, when the number of threshold exceedances is small, the EVT could not work well for unfiltered non-iid time series.

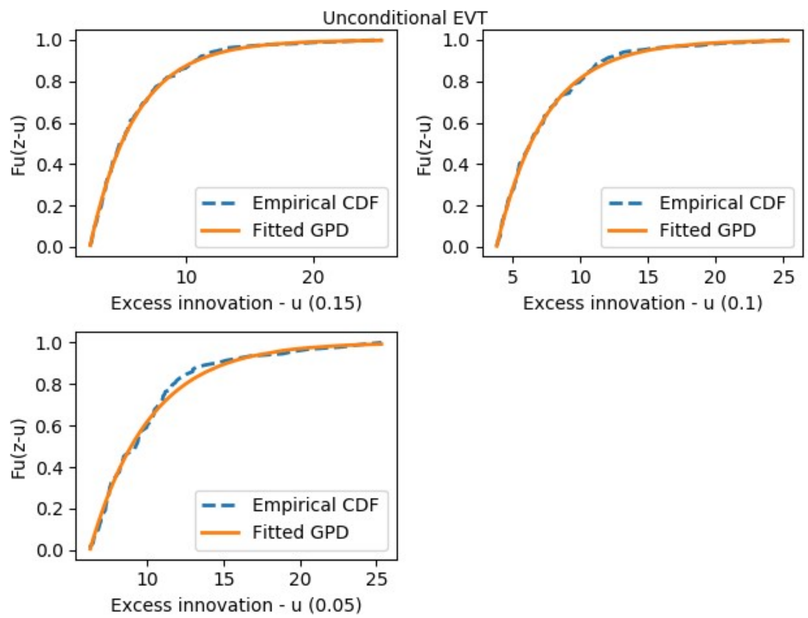


Figure 4.9: Estimated excesses vs empirical excess (Unconditional EVT).

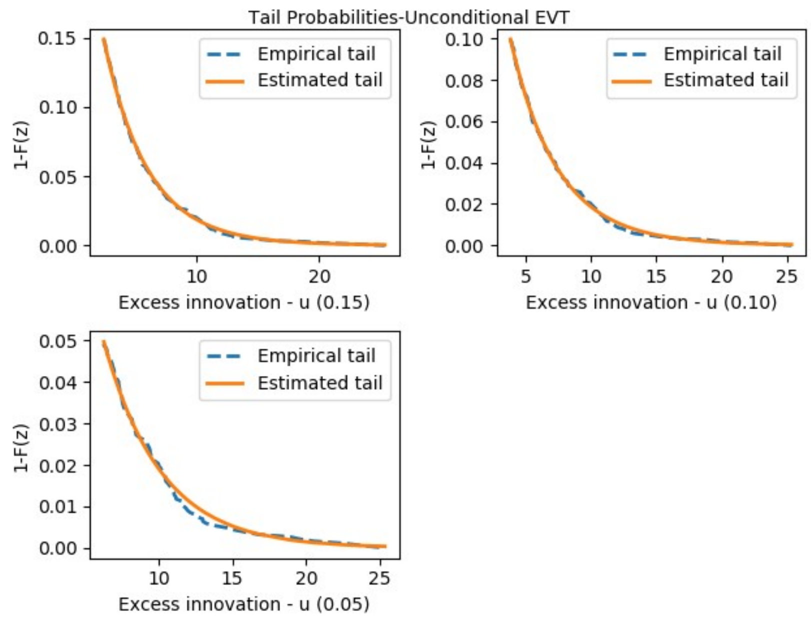


Figure 4.10: Estimated tail probabilities vs empirical tail probabilities (Unconditional EVT).

4.3.5 Value at Risk (VaR) and Expected Shortfall (ES)

We use a rolling window to estimate one day ahead Value at Risk and expected shortfall and plot each model's performance. We find similar results from VaR and ES forecasting plots. Therefore, we mainly discuss the ES forecasting performance here. The following figures in this section are ES plots.

Figure 4.11, 4.12 and 4.13 exam whether the conditional EVT models are sensitive to the choice of threshold u. We could see that given the probability distribution Gaussian, Student t, or generalized error, and given confidence level 99%, 97.5%, or 95%, the difference between ES generated by the underlying model is too small to be visually captured. It seems that the performance of each model is not sensitive to different thresholds u. This study makes sense since all the thresholds u we selected here are in the appropriate range. Therefore, within an appropriate range, the choice of threshold u is not an essential factor in one day ahead ES forecasting.

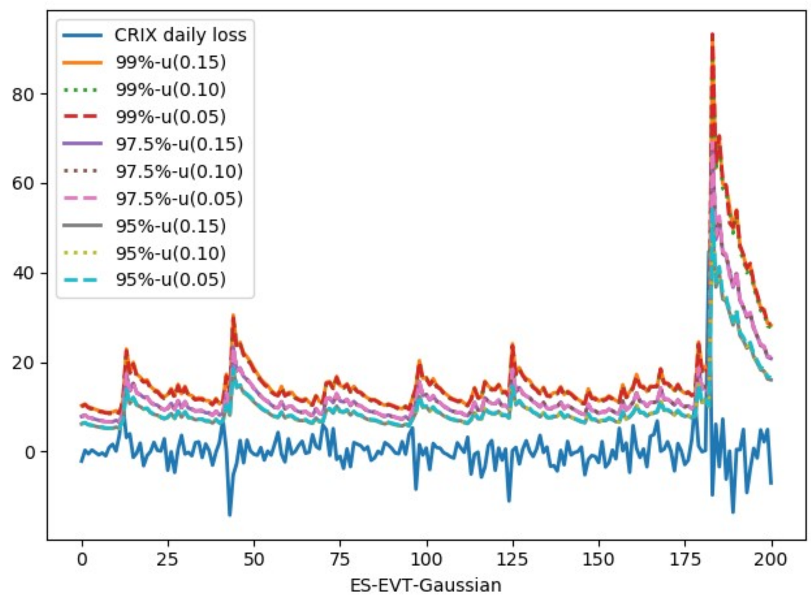


Figure 4.11: One day ahead ES (EVT-Gaussian).

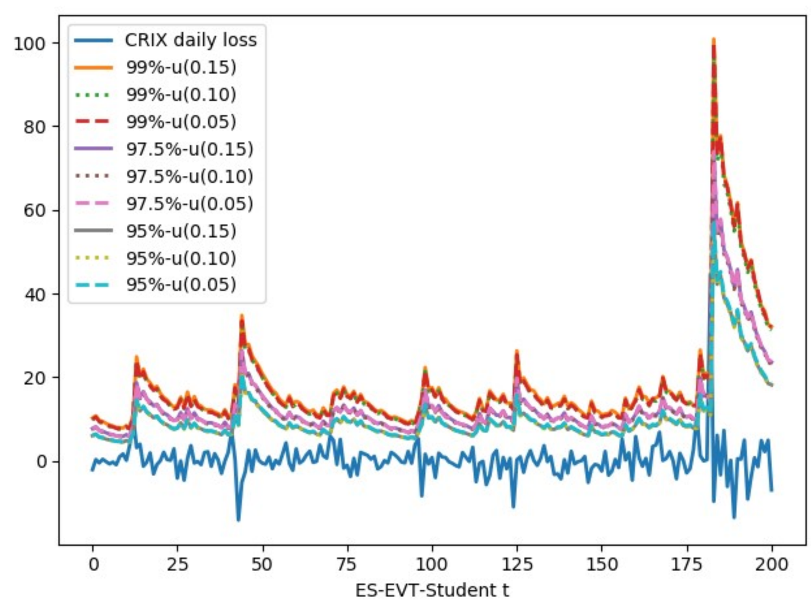


Figure 4.12: One day ahead ES (EVT-Student t)).

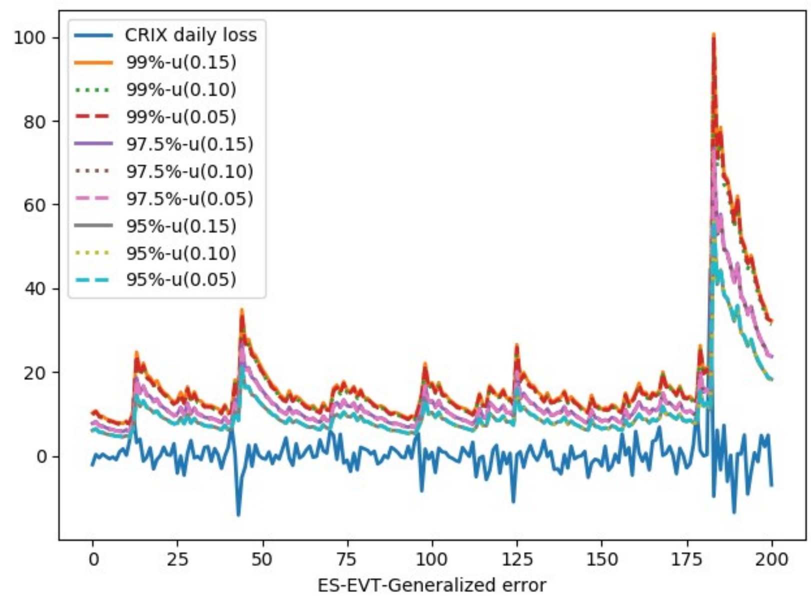


Figure 4.13: One day ahead ES (EVT-Generalized error).

Figure 4.14, 4.15 and 4.16 plot the one day ahead ES calculated by conditional EVT models under probability distribution assumption Gaussian, Student t, and generalized error. At each confidence level, the ES generated by each underlying model is similar. The ES under the Gaussian assumption is slightly below the ES under student t and generalized assumption, but the differences are not significant. The estimated shape parameter of standard innovations from the GARCH process under student t and generalized error assumption is 3.12 and 0.86, respectively, which indicates a heavy tail. This may explain the ES under student t, and generalized error assumption are always above Gaussian assumption. Nevertheless, overall, from the figures, the conditional ES model is not significantly sensitive to probability distribution choice.

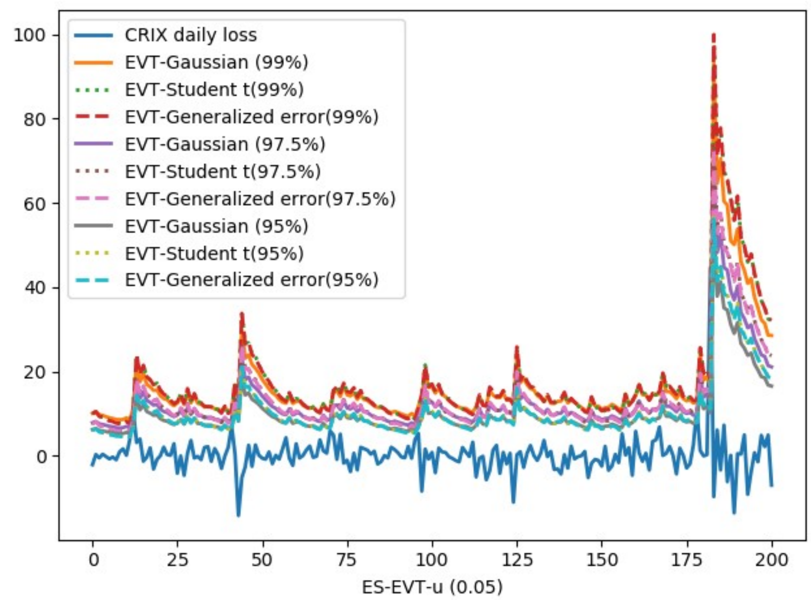


Figure 4.14: One day ahead ES (EVT-u (0.05)).

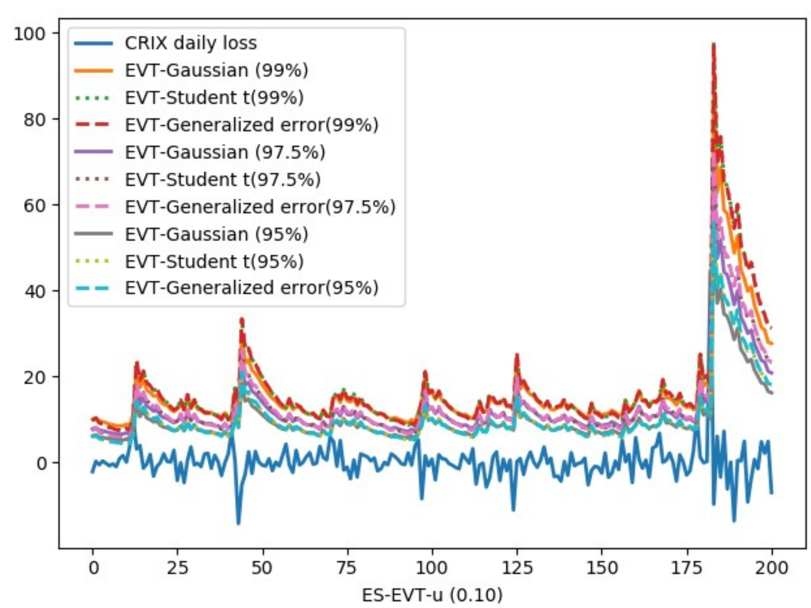


Figure 4.15: One day ahead ES (EVT-u (0.10)).

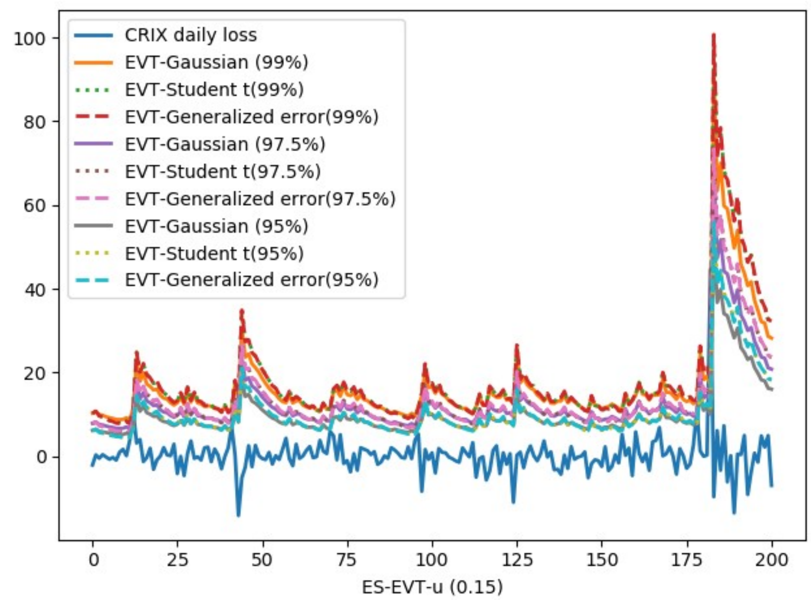


Figure 4.16: One day ahead ES (EVT-u (0.15)).

Figure 4.17, 4.18 and 4.18 show the performance of one day head ES calculated by conditional non EVT models under each distribution assumption at confidence level 99%, 97.5%, and 95%, compared with conditional EVT models. We can see from figure 4.17, the difference in ES between the two models is relatively small at 95% confidence level, but it becomes more critical at 97.5% confidence level and even more important at 99% confidence level. Also, the ES generated by ES models is always above the one generated by non-EVT models. A question arises: is the EVT model overestimating risk, or is the non-EVT model underestimating risk? We will discuss this question in the next section.

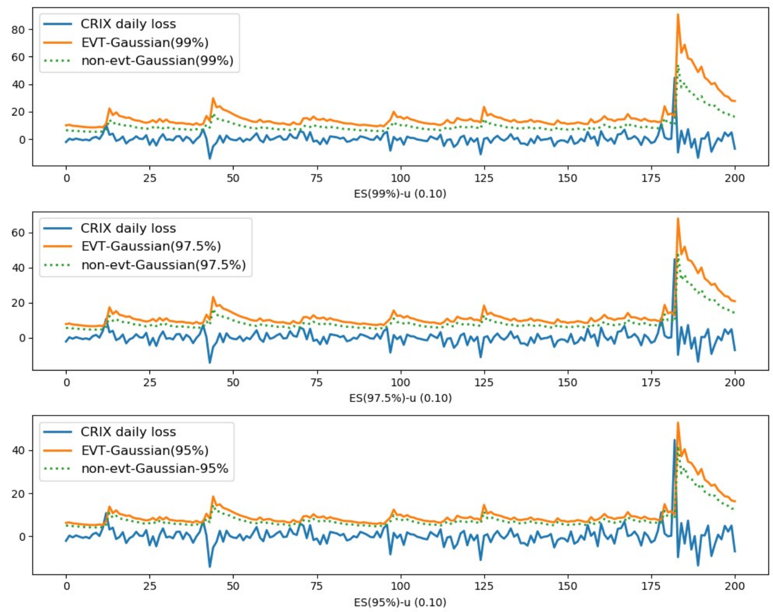


Figure 4.17: ES-EVT vs ES-non-EVT (Gaussian, u (0.10)).

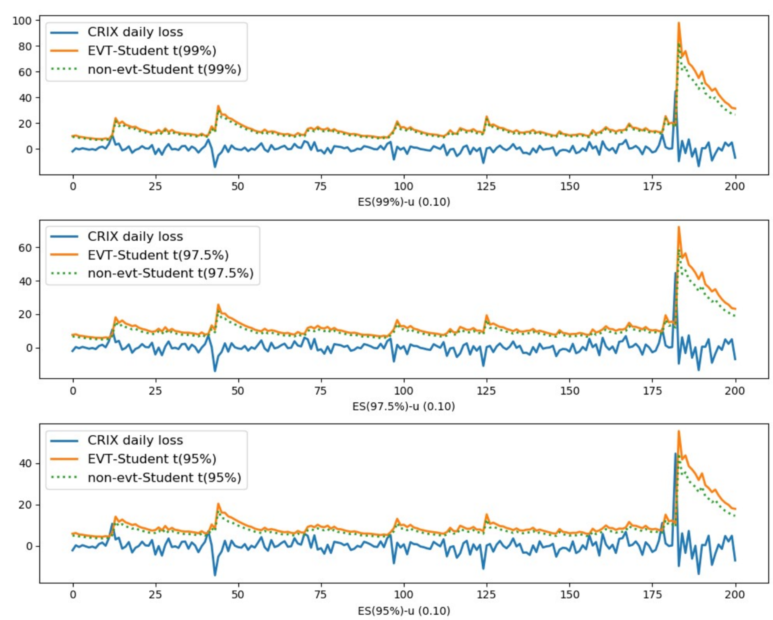


Figure 4.18: ES-EVT VS ES-non-EVT (Student t, u(0.10)).

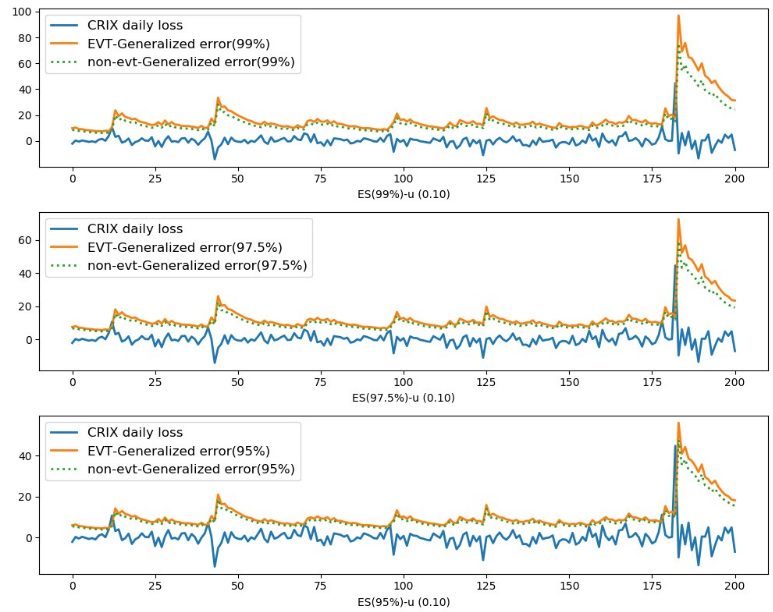


Figure 4.19: ES-EVT VS ES-non-EVT (Generalized error, u(0.10)).

Figure 4.20, 4.21 and 4.22plot the performance of filtered historical simulation (FHS) methods. The performance of the FHS models is similar to the EVT models. We are not able to see the significant difference between them from these figures.

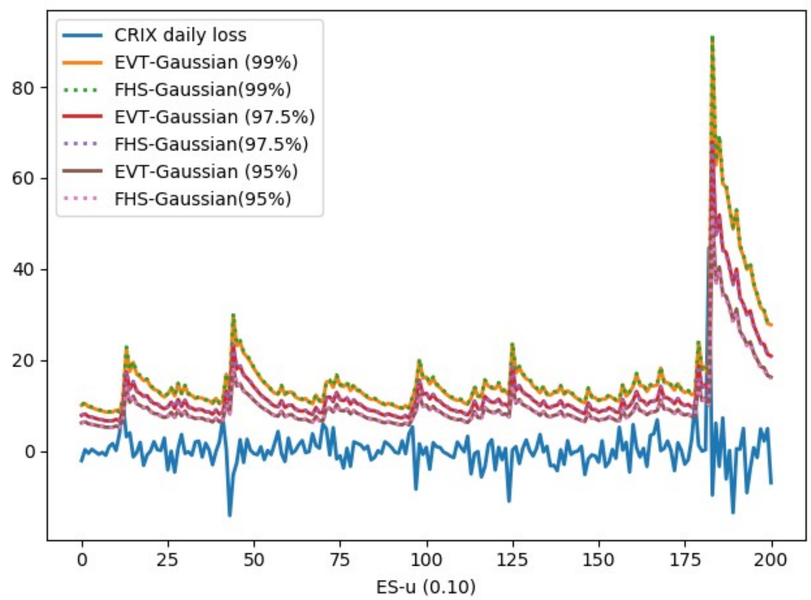


Figure 4.20: ES-EVT vs ES-FHS (Gaussian).

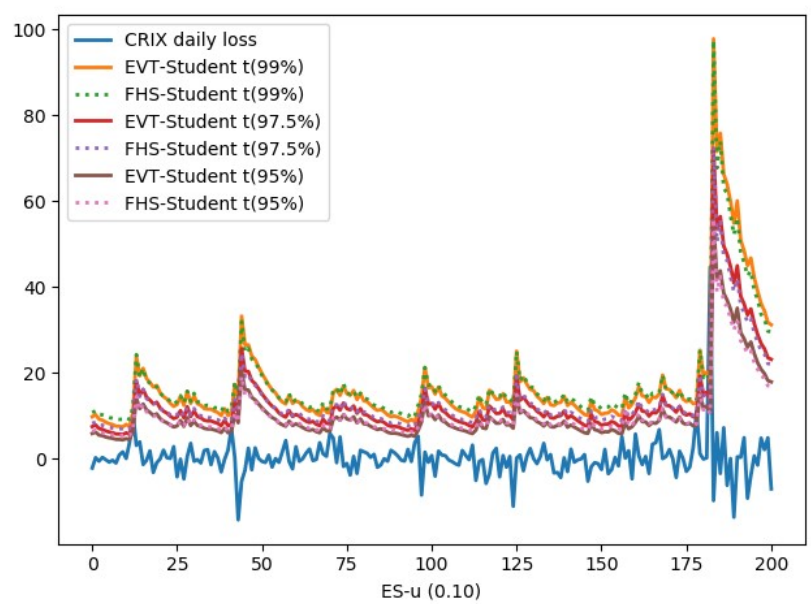


Figure 4.21: ES-EVT vs ES-FHS (Student t).

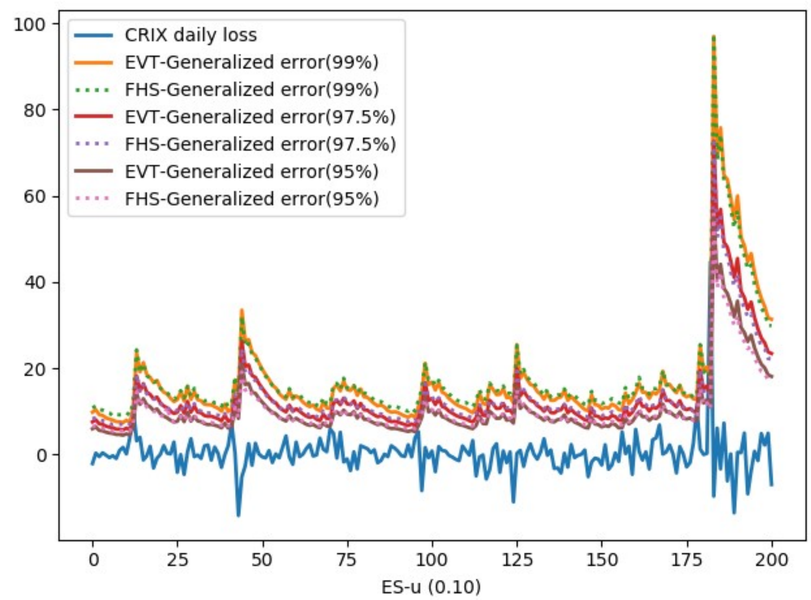


Figure 4.22: ES-EVT vs ES-FHS (Generalized error).

Figure 4.23 plots the performance of unconditional EVT compared to conditional EVT models under each probability distribution assumption. The ES generated by the unconditional EVT model is close to a straight line, which could not respond to the changing volatility as the conditional methods.

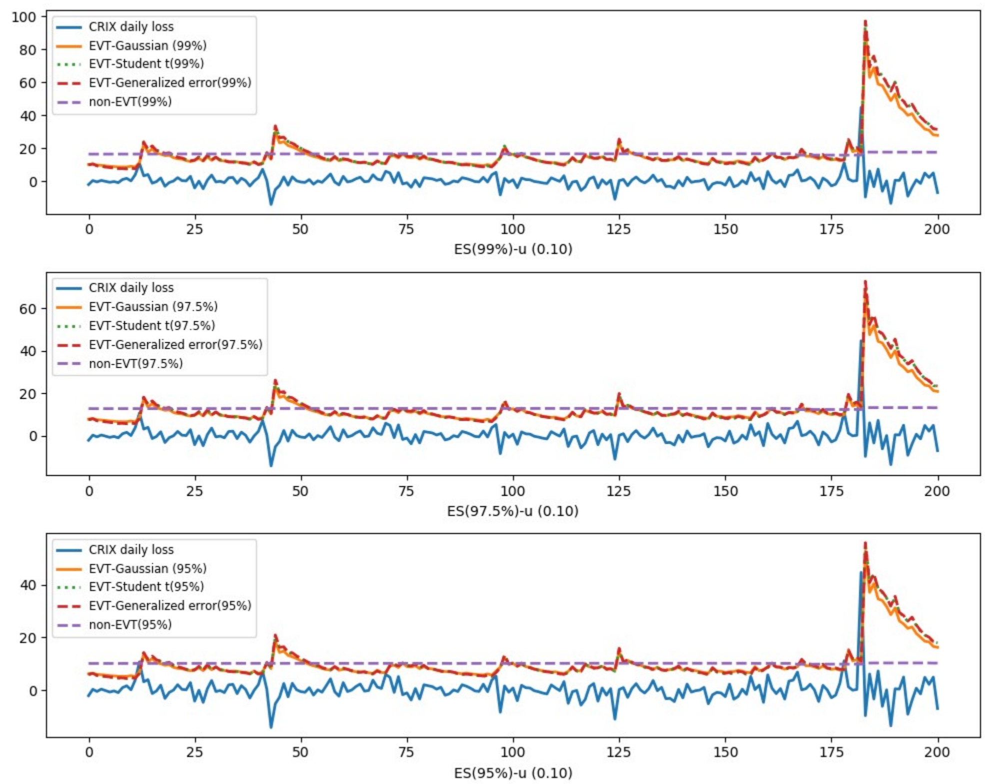


Figure 4.23: Conditional EVT vs Unconditional EVT.

4.4 Backtesting

So far, we have compared the performance of each model from the figures. In this section, we evaluate the performances theoretically. We did the VaR and ES backtesting.

We conduct two VaR tests: unconditional coverage test and conditional coverage test. The main idea of VaR evaluation is to compare the estimated VaR with the realized loss. Kupiec (1995) [46] proposed the unconditional coverage test, which tests for the number of violations. The violation refers to the realized loss is uncovered by (greater than) the estimated VaR. Let $N_1 = \sum_{i=1}^{N} I_{t+1}$ be the total number of days that the realized loss is uncovered by the forecasted VaR, $N_0 = N - N_1$ be the total number of days that are covered:

$$l_{t+1} = \begin{cases} 1 & \text{if} \quad l_{t+1} > V\hat{a}R(\alpha) \\ 0 & \text{if} \quad l_{t+1} < V\hat{a}R(\alpha) \end{cases} \tag{4.20}$$

If the VaR model is correctly specified, $\frac{N_1}{N}$ should be equal to the given failure rate $1 - \alpha$. Therefore, the null hypothesis of the test is $H_0 : \frac{N_1}{N} = 1 - \alpha$,

$$LR_{uc} = 2\ln[\frac{(\frac{N_1}{N})^{N_1}(1 - \frac{N_1}{N})^{N_0}}{(1-\alpha)^{N_1}(\alpha)^{N_0}}] \xrightarrow{N\to\infty} \chi_1^2 \tag{4.21}$$

The conditional coverage test was introduced by Christoffersen (1998) [47]. To test whether the failure rate is equal to the expected one, the test also checks the VaR violation process's independence. Let N_{ij} be the total number of days of state i followed by state j, $i, j = 0, 1$. For example, N_{11} is the number of days that a violation is followed by a violation; N_{10} denotes the number of days that a violation is followed by a non-violation. Now the null hypothesis is that $\frac{N_1}{N}$ should be equal to the given failure rate $1 - \alpha$ and the violation process is independent,

$$LR_{cc} = 2\ln[\frac{(1 - \frac{N_{01}}{N_0})^{N_{00}}(\frac{N_{01}}{N_0})^{N_{01}}(1 - \frac{N_{11}}{N_1})^{N_{10}}(\frac{N_{11}}{N_1})^{N_{11}}}{(1-\alpha)^{N_1}(\alpha)^{N_0}}] \xrightarrow{N\to\infty} \chi_2^2 \tag{4.22}$$

Table 4.5 shows the backtesting results of these approaches by listing violation ratio, unconditional coverage test (Kupiec, 1995) [46], and conditional coverage test (Christoffersen, 1998) [47] p-value.

The bold italic number indicates the p-value is less than the significance level. We desire the violation rate to be close to the theoretical rate. At 0.99 quantile, the theoretical violation rate is 0.01. The violation rate of the unconditional EVT model is equal to the expected one, while conditional EVT models are slightly higher than 0.01 and non-EVT models. The p values of conditional EVT models with different thresholds u are the same, consistent with the results we get from the figures. We can also see a big difference in p values across different models. The p-value of EVT models and FHS models are significantly higher than non-EVT models. It implies that at 0.99 quantiles, the EVT and FHS models, which focus on tail events, are more plausible

	Gaussian			Student t			Generalized Error		
	Viol	LR_{uc}	LR_{cc}	Viol	LR_{uc}	LR_{cc}	Viol	LR_{uc}	LR_{cc}
0.99 Quantile		u (0.15)			u (0.10)			u (0.05)	
GARCH-EVT-Gaussian	0.015	0.513	0.771	0.015	0.513	0.771	0.015	0.513	0.771
GARCH-EVT-Student t	0.015	0.513	0.771	0.015	0.513	0.771	0.015	0.513	0.771
GARCH-EVT-GE	0.015	0.513	0.771	0.015	0.513	0.771	0.015	0.513	0.771
FHS-Gaussian	0.015	0.513	0.771	–	–	–	–	–	–
FHS-Student t	0.015	0.513	0.771	–	–	–	–	–	–
FHS-GE	0.015	0.513	0.771	–	–	–	–	–	–
Unconditional EVT	0.010	0.428	0.727	0.005	0.428	0.727	0.015	0.428	0.727
Non-EVT-Gaussian	0.020	0.214	0.426	–	–	–	–	–	–
Non-EVT-Student t	0.020	0.214	0.426	–	–	–	–	–	–
Non-EVT-GE	0.020	0.214	0.426	–	–	–	–	–	–
0.975 Quantile		u (0.15)			u (0.10)			u (0.05)	
GARCH-EVT-Gaussian	0.025	0.991	0.880	0.025	0.991	0.880	0.020	0.631	0.821
GARCH-EVT-Student t	0.030	0.669	0.758	0.035	0.669	0.758	0.030	0.669	0.758
GARCH-EVT-GE	0.035	0.399	0.334	0.025	0.991	0.880	0.025	0.991	0.880
FHS-Gaussian	0.025	0.991	0.880	–	–	–	–	–	–
FHS-Student t	0.020	0.631	0.821	–	–	–	–	–	–
FHS-GE	0.020	0.631	0.821	–	–	–	–	–	–
Unconditional EVT	0.015	0.323	0.586	0.015	0.586	0.323	0.030	0.323	0.586
Non-EVT-Gaussian	0.035	0.399	0.544	–	–	–	–	–	–
Non-EVT-Student t	0.055	***0.019***	***0.020***	–	–	–	–	–	–
Non-EVT-GE	0.040	0.215	0.275	–	–	–	–	–	–
0.95 Quantile		u (0.15)			u (0.10)			u (0.05)	
GARCH-EVT-Gaussian	0.055	0.762	0.288	0.055	0.762	0.288	0.055	0.762	0.288
GARCH-EVT-Student t	0.060	0.540	0.331	0.060	0.762	0.288	0.055	0.762	0.288
GARCH-EVT-GE	0.060	0.540	0.331	0.092	0.762	0.288	0.055	0.762	0.288
FHS-Gaussian	0.055	0.762	0.288	–	–	–	–	–	–
FHS-Student t	0.055	0.762	0.288	–	–	–	–	–	–
FHS-GE	0.050	0.987	0.802	–	–	–	–	–	–
Unconditional EVT	0.030	0.158	0.306	0.030	0.158	0.306	0.090	0.158	0.306
Non-EVT-Gaussian	0.055	0.762	0.288	–	–	–	–	–	–
Non-EVT-Student t	0.080	0.075	0.072	–	–	–	–	–	–
Non-EVT-GE	0.060	0.540	0.331	–	–	–	–	–	–

Table 4.5: Backtesting resuts on violation ratio, unconditional coverage test of Kupiec (1995) and conditional coverage test of Christoffersen (1998)-p value.

than non-EVT models that interest in full distribution. At 0.975 quantile, the violation rate of non-EVT models is higher than the EVT and FHS models; the p-value of non-EVT models is much lower than EVT and FHS models; even more, the p-value of the non-EVT model under student t distribution is lower than the significant level. At 0.95 quantile, the difference among these models is not as significant as at more extreme quantiles, except for the non-EVT model under the student t distribution assumption. From table 5, we find that the performance of EVT models and FHS models are similar, and they dominant non-EVT models, especially for extreme tail events. This might answer the question we raised previously, whether the EVT models are overestimating risk or non-EVT models are underestimating risk. The non-EVT models are underestimating risk. Unexpectedly, we find unconditional EVT model performs well according to violation rate and VaR tests. However, it performs poorly in responding to loss change movements.

To test for the ES forecasting performance, we follow McNeil and Frey's (2000) [44] ES testing. If bad things happen, where the VaR violation occurs, we define the exceedance residual:

$$ER_{t+1}(\alpha) = \frac{l_{t+1} - \hat{ES}_{t+1}(\alpha)}{\hat{sigma}_{t+1}}, where l_{t+1} > \hat{VaR}_{t+1}(\alpha) \tag{4.23}$$

If the ES is correctly estimated, the exceedance residuals are expected to be an iid sample with zero mean. Therefore, we conduct the test to check for zero mean of exceedance residuals. Table 4.6 gives each model's p-value for the ES test with the null hypothesis that the exceedance residuals have zero mean. All of these models fail to reject the null hypothesis. The p-values of EVT and FHS models are higher than these of non-EVT models, indicating EVT and FHS models are more plausible and reliable in expected shortfall forecasting

	0.99 Quantile			0.975 Quantile			0.95 Quantile		
u	0.15	0.10	0.05	0.15	0.10	0.05	0.15	0.10	0.05
GARCH-EVT-Gaussian	0.192	0.186	0.189	0.168	0.167	0.142	0.198	0.202	0.211
GARCH-EVT-Student t	0.198	0.184	0.189	0.188	0.184	0.187	0.197	0.167	0.190
GARCH-EVT-GE	0.195	0.181	0.186	0.208	0.156	0.158	0.199	0.177	0.183
FHS-Gaussian		0.191			0.169			0.201	
FHS-Student t		0.211			0.157			0.228	
FHS-GE		0.212			0.157			0.213	
Unconditional EVT	/	/	/	0.195	0.224	0.197	0.229	0.223	0.230
Non-EVT-Gaussian		0.100			0.107			0.100	
Non-EVT-ST		0.187			0.224			0.164	
Non-EVT-GE		0.155			0.156			0.130	

Table 4.6: Backtesting results p-value on expected shortfall test .

4.5 Conclusion

This study investigates the tail risk of CRIX daily loss measured by Value at Risk (VaR) and Expected Shortfall (ES). To calculate one day ahead VaR and ES, we estimate loss tail distribution by applying parametric approach extreme value theory, which fits the standard innovations obtained from ARMA-GARCH process to the generalized Pareto distribution (GPD); and by applying semiparametric method filtered historical simulation (FHS), which estimates the quantile of filtered standard innovations. In this study, we find that:

1. Non-EVT models are standard parametric models that capture the full risk of CRIX loss innovations, while EVT models estimate tail risk by focusing on tail events. EVT models dominate non EVT models by generating more accurate and plausible VaR and ES forecasting, especially for the more extreme quantiles.

2. Discriminating among probability distributions does not play an important role when we are

focusing on extreme observations.

3. For the in-sample data, the choice of threshold u slightly affects the GPD fitting performance, while for the out of sample data, EVT models are not sensitive to the selection of threshold u.

4. FHS also performs well in VaR and ES forecasting and yields similar results to EVT models. This can be regarded as evidence of the robustness of the results.

5. The unconditional EVT models were supposed to reject VaR and ES tests' null hypotheses because the EVT method requires iid sample data. Unexpectedly, they fail to reject. However, it performs poorly in responding to volatility change and cannot capture volatility clustering as conditional models.

6. Finally, we recommend conditional EVT methods and FHS methods in VaR and ES forecasting for cryptocurrency market investors who are exploring their potential risk exposures.

5. CONCLUSION

In summary, We explore the inter-market relationships among bitcoin, Ethereum, XRP, and Litecoin prices and found their prices affect each other.

We forecast bitcoin return's volatility and Value at Risk. The results indicate recurrent neural network method outperforms GARCH and EWMA in average forecasting performance. However, it is less efficient in capturing the bitcoin market's extreme events. Moreover, the RNN shows poor performance in Value at Risk forecasting, indicating that it could not work well as the econometric models in explaining extreme volatility.

We investigate the tail behavior of cryptocurrency index return and found that focusing on tail events using extreme value theory and filtered historical simulation methods could significantly improve CRIX tail risk accuracy, which is more reliable in cryptocurrency market risk management.

REFERENCES

[1] M. Buchholz, J. Delaney, J. Warren, and J. Parker, "Bits and bets, information, price volatility, and demand for bitcoin," *Economics*, vol. 312, pp. 2–48, 2012.

[2] L. Kristoufek, "Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era," *Scientific reports*, vol. 3, no. 1, pp. 1–7, 2013.

[3] P. Ciaian, M. Rajcaniova, *et al.*, "The digital agenda of virtual currencies: Can bitcoin become a global currency?," *Information Systems and e-Business Management*, vol. 14, no. 4, pp. 883–919, 2016.

[4] J. Bouoiyour and R. Selmi, "What does bitcoin look like?," *Annals of Economics & Finance*, vol. 16, no. 2, 2015.

[5] D. Yermack, "Is bitcoin a real currency? an economic appraisal," in *Handbook of digital currency*, pp. 31–43, Elsevier, 2015.

[6] E.-T. Cheah and J. Fry, "Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin," *Economics letters*, vol. 130, pp. 32–36, 2015.

[7] A. Cheung, E. Roca, and J.-J. Su, "Crypto-currency bubbles: an application of the phillips–shi–yu (2013) methodology on mt. gox bitcoin prices," *Applied Economics*, vol. 47, no. 23, pp. 2348–2358, 2015.

[8] S. Corbet, B. Lucey, and L. Yarovaya, "Datestamping the bitcoin and ethereum bubbles," *Finance Research Letters*, vol. 26, pp. 81–88, 2018.

[9] J. Xiong, Q. Liu, and L. Zhao, "A new method to verify bitcoin bubbles: Based on the production cost," *The North American Journal of Economics and Finance*, vol. 51, p. 101095, 2020.

[10] P. Ciaian, M. Rajcaniova, and d. Kancs, "The economics of bitcoin price formation," *Applied Economics*, vol. 48, no. 19, pp. 1799–1815, 2016.

[11] L. Kristoufek, “What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis,” *PloS one*, vol. 10, no. 4, p. e0123923, 2015.

[12] D. Van Wijk, “What can be expected from the bitcoin,” *Erasmus Universiteit Rotterdam*, vol. 18, 2013.

[13] P. Katsiampa, “Volatility estimation for bitcoin: A comparison of garch models,” *Economics Letters*, vol. 158, pp. 3–6, 2017.

[14] B. M. Blau, “Price dynamics and speculative trading in bitcoin,” *Research in International Business and Finance*, vol. 43, pp. 15–21, 2018.

[15] Y. Peng, P. H. M. Albuquerque, J. M. C. de Sá, A. J. A. Padula, and M. R. Montenegro, “The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression,” *Expert Systems with Applications*, vol. 97, pp. 177–192, 2018.

[16] N. B. Cheikh, Y. B. Zaied, and J. Chevallier, “Asymmetric volatility in cryptocurrency markets: New evidence from smooth transition garch models,” *Finance Research Letters*, vol. 35, p. 101293, 2020.

[17] X. Li and C. A. Wang, “The technology and economic determinants of cryptocurrency exchange rates: The case of bitcoin,” *Decision Support Systems*, vol. 95, pp. 49–60, 2017.

[18] N. Gandal and H. Halaburda, “Can we predict the winner in a market with network effects? competition in cryptocurrency market,” *Games*, vol. 7, no. 3, p. 16, 2016.

[19] J. Osterrieder, “The statistics of bitcoin and cryptocurrencies,” in *2017 International Conference on Economics, Finance and Statistics (ICEFS 2017)*, pp. 285–289, Atlantis Press, 2017.

[20] C. W. Granger, “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

[21] A. E. Gencer, S. Basu, I. Eyal, R. Van Renesse, and E. G. Sirer, "Decentralization in bitcoin and ethereum networks," in *International Conference on Financial Cryptography and Data Security*, pp. 439–457, Springer, 2018.

[22] J. Bouoiyour and R. Selmi, "Bitcoin price: Is it really that new round of volatility can be on way?," 2015.

[23] J. Bouoiyour, R. Selmi, *et al.*, "Bitcoin: A beginning of a new phase," *Economics Bulletin*, vol. 36, no. 3, pp. 1430–1440, 2016.

[24] M. Balcilar, E. Bouri, R. Gupta, and D. Roubaud, "Can volume predict bitcoin returns and volatility? a quantiles-based approach," *Economic Modelling*, vol. 64, pp. 74–81, 2017.

[25] V. Troster, A. K. Tiwari, M. Shahbaz, and D. N. Macedo, "Bitcoin returns and risk: A general garch and gas analysis," *Finance Research Letters*, vol. 30, pp. 187–193, 2019.

[26] S. Athey, "21. the impact of machine learning on economics," in *The economics of artificial intelligence*, pp. 507–552, University of Chicago Press, 2019.

[27] F. Fang, C. Ventre, M. Basios, H. Kong, L. Kanthan, L. Li, D. Martinez-Regoband, and F. Wu, "Cryptocurrency trading: a comprehensive survey," *arXiv preprint arXiv:2003.11352*, 2020.

[28] S. McNally, J. Roche, and S. Caton, "Predicting the price of bitcoin using machine learning," in *2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP)*, pp. 339–343, IEEE, 2018.

[29] L. Alessandretti, A. ElBahrawy, L. M. Aiello, and A. Baronchelli, "Anticipating cryptocurrency prices using machine learning," *Complexity*, vol. 2018, 2018.

[30] M. B. Garman and M. J. Klass, "On the estimation of security price volatilities from historical data," *Journal of business*, pp. 67–78, 1980.

[31] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[32] T. G. Andersen and T. Bollerslev, "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International economic review*, pp. 885–905, 1998.

[33] R. Y. Chou, H. Chou, and N. Liu, "Range volatility models and their applications in finance," in *Handbook of quantitative finance and risk management*, pp. 1273–1281, Springer, 2010.

[34] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," *Climate research*, vol. 30, no. 1, pp. 79–82, 2005.

[35] T. Chai and R. R. Draxler, "Root mean square error (rmse) or mean absolute error (mae)?– arguments against avoiding rmse in the literature," *Geoscientific model development*, vol. 7, no. 3, pp. 1247–1250, 2014.

[36] V.-P. Heikkinen and A. Kanto, "Value-at-risk estimation using non-integer degrees of freedom of student's distribution," *Journal of Risk*, vol. 4, pp. 77–84, 2002.

[37] A. Andreev and A. Kanto, "Conditional value-at-risk estimation using non-integer values of degrees of freedom in student's t-distribution," *The Journal of Risk*, vol. 7, no. 2, p. 1, 2004.

[38] J. Rebane, I. Karlsson, P. Papapetrou, and S. Denic, "Seq2seq rnns and arima models for cryptocurrency prediction: A comparative study," in *SIGKDD Fintech'18, London, UK, August 19-23, 2018*, 2018.

[39] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," tech. rep., Manubot, 2019.

[40] C. Alexander and J. M. Sarabia, "Quantile uncertainty and value-at-risk model risk," *Risk Analysis: An International Journal*, vol. 32, no. 8, pp. 1293–1308, 2012.

[41] A. Phillip, J. S. Chan, and S. Peiris, "A new look at cryptocurrencies," *Economics Letters*, vol. 163, pp. 6–9, 2018.

[42] J. Osterrieder and J. Lorenz, "A statistical risk assessment of bitcoin and its extreme tail behavior," *Annals of Financial Economics*, vol. 12, no. 01, p. 1750003, 2017.

[43] K. Gkillas and P. Katsiampa, "An application of extreme value theory to cryptocurrencies," *Economics Letters*, vol. 164, pp. 109–111, 2018.

[44] A. J. McNeil and R. Frey, "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach," *Journal of empirical finance*, vol. 7, no. 3-4, pp. 271–300, 2000.

[45] P. Christoffersen, *Elements of financial risk management*. Academic Press, 2011.

[46] P. Kupiec, "Techniques for verifying the accuracy of risk measurement models," *The J. of Derivatives*, vol. 3, no. 2, 1995.

[47] P. F. Christoffersen, "Evaluating interval forecasts," *International economic review*, pp. 841–862, 1998.

APPENDIX

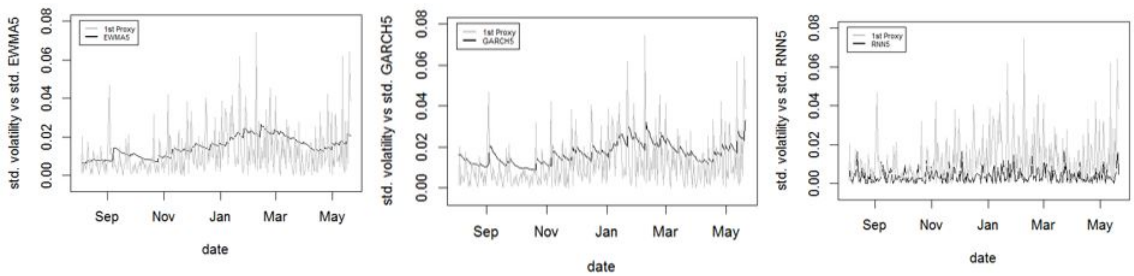


Figure 1: Out of sample standard deviation of realized volatility (1st proxy) vs standard deviation of 5 days ahead volatility forecasting of EWMA, GARCH and RNN model

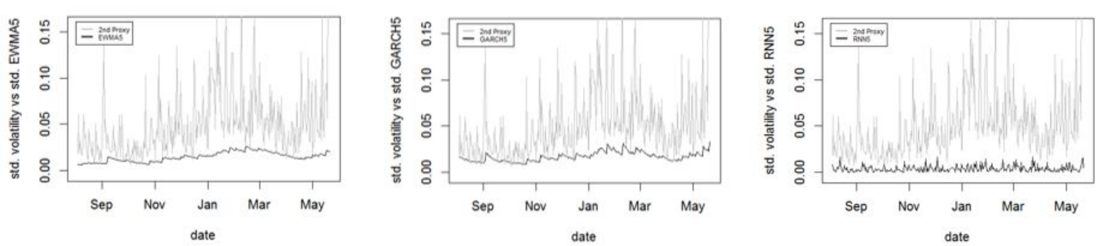


Figure 2: Out of sample standard deviation of realized volatility (2nd proxy) vs standard deviation of 5 days ahead volatility forecasting of EWMA, GARCH and RNN model

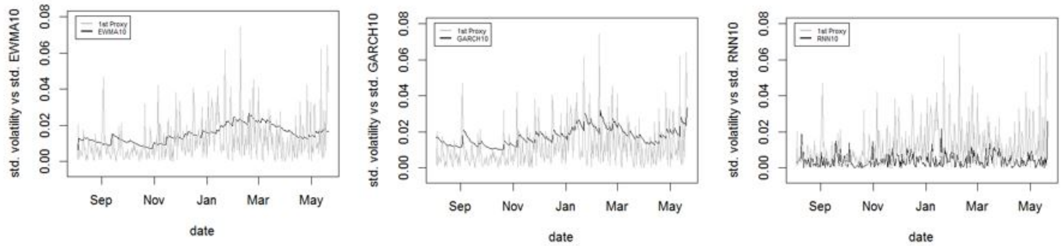


Figure 3: Out of sample standard deviation of realized volatility (1st proxy) vs standard deviation of 10 days ahead volatility forecasting of EWMA, GARCH and RNN model

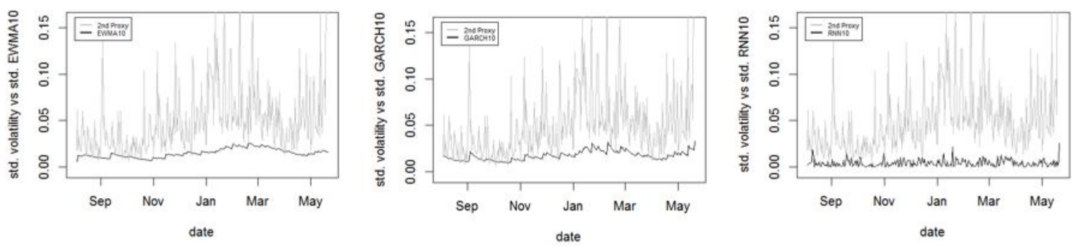


Figure 4: Out of sample standard deviation of realized volatility (2nd proxy) vs standard deviation of 10 days ahead volatility forecasting of EWMA, GARCH and RNN model

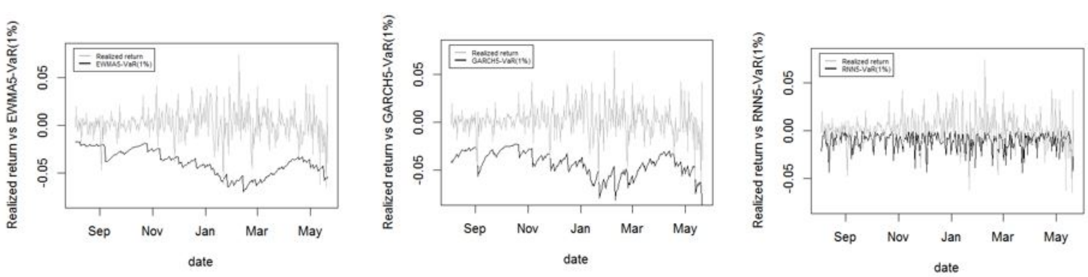


Figure 5: Out of sample realized return vs 5 days ahead VaR(1%) of EWMA, GARCH and RNN model

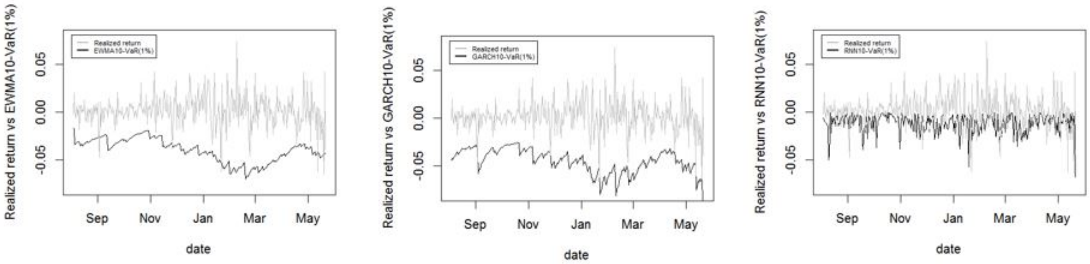


Figure 6: Out of sample realized return vs 10 days ahead VaR(1%) of EWMA, GARCH and RNN model

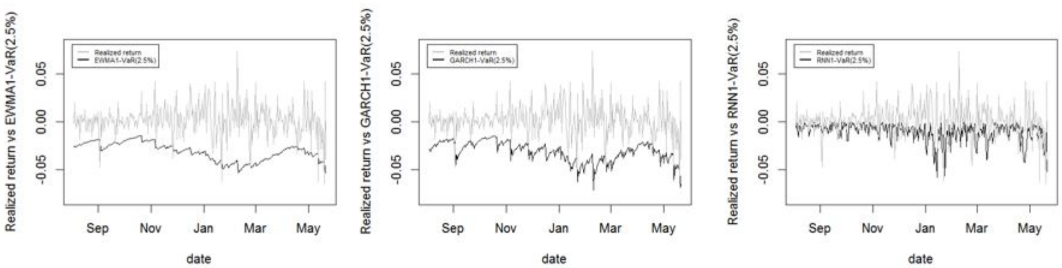


Figure 7: Out of sample realized return vs 1 day ahead VaR(2.5%) of EWMA, GARCH and RNN model

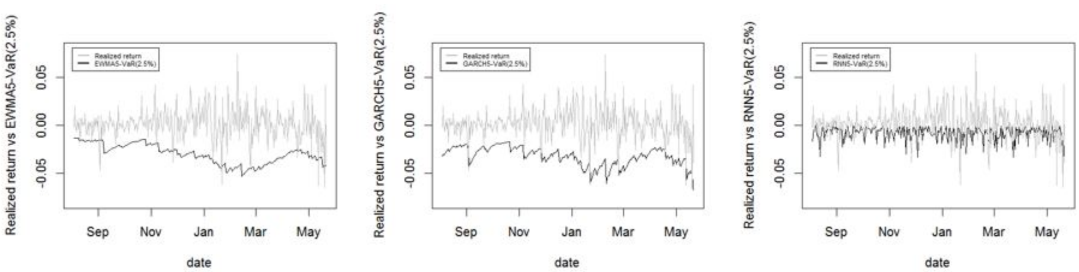


Figure 8: Out of sample realized return vs 5 days ahead VaR(2.5%) of EWMA, GARCH and RNN model

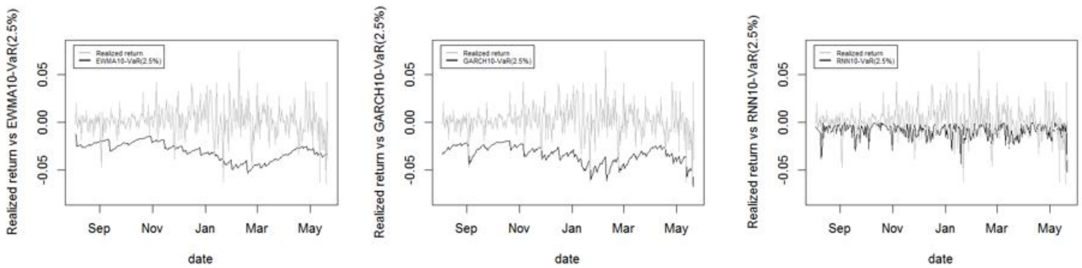


Figure 9: Out of sample realized return vs 10 days ahead VaR(2.5%) of EWMA, GARCH and RNN model

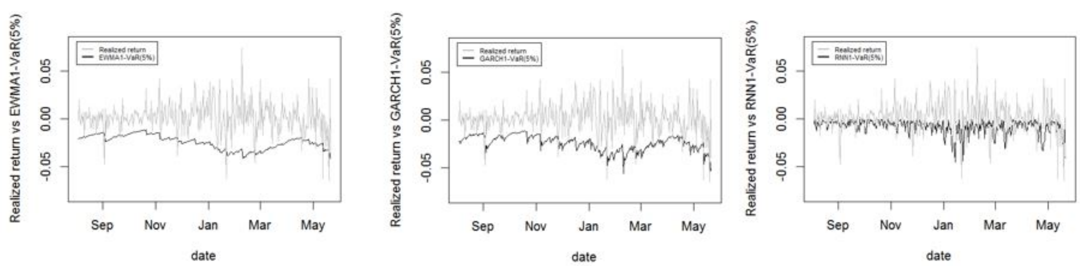


Figure 10: Out of sample realized return vs 1 day ahead VaR(5%) of EWMA, GARCH and RNN model

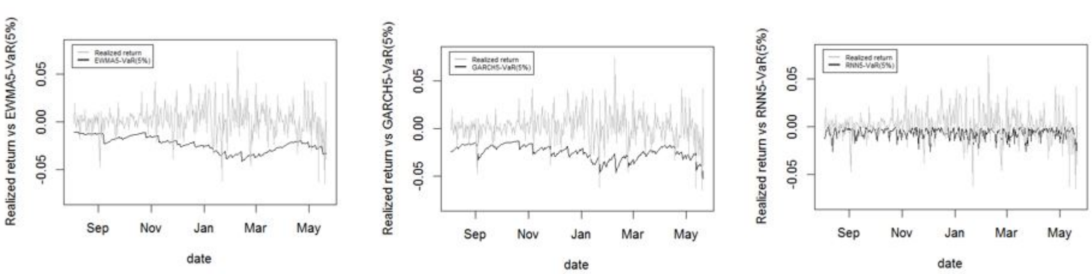


Figure 11: Out of sample realized return vs 5 days ahead VaR(5%) of EWMA, GARCH and RNN model

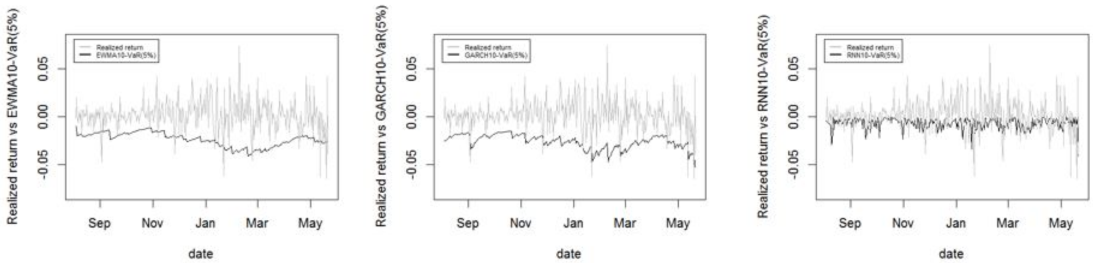


Figure 12: Out of sample realized return vs 10 days ahead VaR(5%) of EWMA, GARCH and RNN model