IMPROVING VOLATILITY FORECASTING OF GARCH-TYPE MODELS USING INDICATOR SATURATION AND WINSORIZATION APPROACHES

SULEIMAN DAHIR MOHAMED

UNIVERSITI SAINS MALAYSIA

2025

IMPROVING VOLATILITY FORECASTING OF GARCH-TYPE MODELS USING INDICATOR SATURATION AND WINSORIZATION APPROACHES

by

SULEIMAN DAHIR MOHAMED

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

April 2025

ACKNOWLEDGEMENT

All thanks to Allah, the Most Gracious and the Most Merciful, who gave me the chance, the knowledge, and the persistence to start and finish this doctoral journey.

Secondly, I want to thank my PhD supervisor, Associate Professor Dr. Mohd Tahir Bin Ismail, for always being there for me and helping me along the way as I worked on my doctoral degree. This thesis and my growth as a researcher have been greatly influenced by your deep knowledge, priceless insights, and unshakable faith in my abilities. I'm very thankful that you've spent so many hours talking about ideas, giving me comments, and pushing me to do better. I also want to thank my co-supervisor Dr. Majid Khan Majahar Ali. Your knowledge, tough questions, and helpful comments have all been very helpful in improving and strengthening this work.

Thank you to my fellow graduate students for the interesting conversations, shared laughs, and times of friendship that have made this experience truly rewarding. The intellectual challenge and feeling of community have been very helpful. I greatly appreciate the School of Mathematical Sciences and Universiti Sains Malaysia for making this study possible by giving me the tools, facilities, and chances to do it. The financial support by Ali Muse Aden has been very important in letting me focus on my studies and research.

Lastly, I want to express my gratitude to my mother Zainab Hassan Malin in particular, whose continuous support and persistent prayers have given me strength and inspiration during this project.

I sincerely appreciate all of you. I appreciate your participation in this amazing experience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ARCH	Autoregressive Conditional Heteroscedasticity
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
EGARCH	Exponential GARCH
GJRGARCH	Glosten Jagannathan Runkle GARCH
FIGARCH	Fractionally Integrated GARCH
Win	Winsorizaiton
IS	Indicator Saturation
IIS	Impulse Indicator Saturation
SIS	Step Indicator Saturation
TIS	Trend Indicator Saturation
Win-IS	Winsorized IS Approach
Win-IIS	Winsorized IIS Estimator
Win-SIS	Winsorized SIS Estimator
Win-TIS	Winsorized TIS Estimator
IS-GARCH	Indicator Saturation GARCH
IIS-GARCH	Impulse Indicator Saturation GARCH
SIS-GARCH	Step Indicator Saturation GARCH
TIS-GARCH	Trend Indicator Saturation GARCH
BP	Bai and Perron Test
LL	Loglikelihood Value
AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
BTC	Bitcoin
ETH	Ethereum
XRP	Ripple
LTC	Litecoin
USDT	Tether USD
norm	Normal Distribution
std	Student-t Distribution
ged	Generalized Error Distribution
TB	Trend Breaks

SB	Structural Breaks
OL	Outliers

LIST OF APPENDICES

PENAMBAHBAIKAN RAMALAN KEMERUAPAN MODEL JENIS GARCH MENGGUNAKAN PENDEKATAN PETUNJUK KETEPUAN DAN WINSOR

ABSTRAK

Pendekatan petunjuk ketepuan tradisional (IS) dalam analisis data siri masa kewangan adalah baik dalam mengesan putusan, putusan trend dan nilai terpencil secara serentak. Untuk meningkatkan pengesanan pendekatan IS, kaedah Winsor (Win) dicadangkan, yang mengurangkan cerapan hujung yang melampau. Tesis ini pada mulanya mencadangkan pendekatan hibrid yang dipanggil strategi Win-IS, yang menangani pengaruh nilai terpencil melampau dalam hujung dan mengenal pasti putusan, putusan trend, dan nilai terpencil dalam mata wang kripto. Kesan winsor bergantung pada persentil yang dipilih dan atribut set data. Sebaliknya, model GARCH kerap digunakan untuk ramalan kemeruapan. Walau bagaimanapun, model GARCH menganggap kenormalan dan tidak mengambil kira ciri data asas seperti putusan berstruktur, nilai terpencil dan putusan trend. Oleh itu, kajian ini juga bertujuan untuk meningkatkan ketepatan ramalan model jenis GARCH untuk kemeruapan dengan menyepadukannya dengan pendekatan petunjuk ketepuan (IS) dan winsor (Win). Penyelidikan menggunakan Bitcoin, Ethereum, Litecoin, Tether dan Ripple sebagai data kewangan dari November 2014 hingga Jun 2023, yang mempamerkan turun naik harga yang ketara dan kerap, untuk menilai keupayaan ramalan model jenis GARCH yang dibangunkan. Kajian mendapati bahawa mata wang kripto mengandungi putusan, putusan trend dan nilai terpencil untuk nilai pulangan log, mempamerkan nilai purata positif dan negatif, kurang kestabilan, sangat tidak menentu, menyimpang daripada normal, mempunyai sisihan piawai yang tinggi, mempamerkan kedua-dua pencongan positif dan negatif, mempunyai pulangan kurtosis yang berlebihan, dan mempamerkan gelagat pengelompokan kemeruapan. Kajian mendapati pendekatan penunjuk ketepuan adalah pilihan yang lebih baik untuk dihibridkan dengan model jenis GARCH, kerana ia secara serentak dapat mengesan putusan, putusan trend dan data terpencil. Untuk menambah baik ramalan kemeruapan model jenis GARCH, analisis dijalankan secara berperingkat. Pertama, kajian ini mempertimbangkan empat model jenis GARCH dan tiga taburan ralat: norma, std, dan ged, dan menganggarkan dan meramalkan 60 model penanda aras jenis GARCH. Keputusan menunjukkan kewujudan kesan asimetri, kesan keumpilan yang positif, nilai kegigihan yang sangat tinggi (sehingga 0.999), dan kesan ingatan panjang dalam kemeruapan bersyarat pasaran mata wang kripto. Kedua, kajian menghibridkan model jenis GARCH dengan kaedah Win, IS dan Win-IS untuk menangani nilai terpencil dihujung, putusan, putusan trend dan nilai terpencil. Ini membawa kepada penciptaan 372 model jenis GARCH hibrid yang berbeza. Hasilnya menunjukkan bahawa kegigihan dan nilai ingatan jangka panjang berkurangan, dan pendekatan penghibridan Win-IS dalam model jenis GARCH mengatasi prestasi model penanda aras apabila membandingkan prestasi dalam dan luar sampel mereka. Dengan mengandaikan bahawa taburan hujung yang berat juga menyumbang, model hibrid yang dicadangkan muncul sebagai optimum. Dengan mengambil kira pelbagai set data kewangan, kerja ini membuka jalan untuk penyelidikan masa depan yang menggabungkan algoritma pembelajaran mesin dan model perubahan Markov.

IMPROVING VOLATILITY FORECASTING OF GARCH-TYPE MODELS USING INDICATOR SATURATION AND WINSORIZATION APPROACHES

ABSTRACT

The traditional indicator saturation (IS) approach in financial time series data analysis is good in detecting breaks, trend breaks, and outliers simultaneously. To improve the detectability of the IS approach, the Winsorization (Win) method is proposed, which mitigates extreme tail observations. This thesis initially proposes a hybrid approach called the Win-IS strategy, which addresses the influence of extreme outliers in the tail and identifies breaks, trend breaks, and outliers in cryptocurrencies. The effect of winsorization depends on the chosen percentile and dataset attributes. On the other hand, GARCH models are frequently employed for volatility prediction. However, standard GARCH models assume normality and do not account for the underlying data features such as structural breaks, outliers, and trend breaks. Therefore, this study also aims to improve the predictive accuracy of GARCH-type models for volatility by integrating them with indicator saturation (IS) and winsorization (Win) approaches. The research uses Bitcoin, Ethereum, Litecoin, Tether, and Ripple as financial data from November 2014 to June 2023, which exhibit significant and frequent price fluctuations, to evaluate the predictive ability of the developed GARCH-type models. The study found that cryptocurrencies contain breaks, trend breaks, and outliers in their log returns, exhibit positive and negative mean values, lack stability, are highly volatile, deviate from normality, have a high standard deviation, exhibit both positive and negative skewness, have excessive kurtosis returns, and exhibit volatility clustering behavior. The study found that the

indicator saturation approach is a better choice to hybridize with GARCH-type models, as it can simultaneously detect breaks, trend breaks, and outliers. To improve the volatility prediction of GARCH type models, analysis is carried out in stages. Firstly, the study considered four GARCH-type models and three error distributions: norm, std, and ged, and estimated and forecasted 60 benchmark GARCH-type models. The results demonstrate the existence of an asymmetric effect, a positive leverage effect, an exceptionally high persistence value (up to 0.999), and a long-memory effect in the cryptocurrency market's conditional volatility. Second, the study hybridized GARCHtype models with Win, IS, and Win-IS methods to deal with tail outliers, breaks, trend breaks, and outliers. This led to the creation of 372 different hybridized GARCH-type models. The results demonstrate that persistence and long-term memory values decrease, and the hybridizing Win-IS approach in GARCH-type models outperforms the benchmark models when comparing their in- and out-of-sample performances. The assumed heavy-tailed distributions also contributed the optimality of the proposed hybrid models. Considering various financial datasets, this work paves the way for future research that incorporates machine learning algorithms and markov switching models.

CHAPTER 1

INTRODUCTION

1.1 Study Background

1.1.1 The History of Cryptocurrency Market

Over the last ten years, cryptocurrency has gained widespread recognition globally, witnessing a substantial surge in popularity. The increasing interest stems from the fact that digital currencies, functioning as pioneering payment systems, have introduced a novel form of currency (Ali et al., 2014). Furthermore, the market's rise can be attributed to factors such as the adoption of advanced technology, the emergence of the fourth industrial revolution, the acceptance of cryptocurrencies as legal tender in several nations, and the influence of large corporations (Yousaf & Ali, 2020). The first cryptocurrency, Bitcoin, created by Nakamoto (2008), has emerged as a captivating technological advancement of the twenty-first century. Due to its remarkable growth over time, it has become the largest cryptocurrency in existence. There exist alternative cryptocurrencies apart from Bitcoin that offer different functionalities and utilize blockchain technology to validate transactions. Notable cryptocurrencies that have emerged include Litecoin (LTC), Ethereum (ETH), Ripple (XRP), USD Tether (USDT), and several others.

Table 1.1 presents a concise summary of the historical development of the cryptocurrency market and the significant fluctuations in the price of Bitcoin up until now. Table 1.2 briefly overviews each of the cryptocurrencies considered here, along with its market capitalization, owner, and the dates of its initial release and launch.

Table 1.1 Cryptocurrency Market History

Period	Market Status	Major Activities	Price Fluctuations
2008-2010	Beginning	- BTC was created. - Blockchain functionality was tested. - 2008-2009 BTC was underway	BTC price 14–36 cents 2010
2010-2014	Market formation	- LTC was released (2011). - XRP and Doge were released (2012)	BTC price $1.06 - $27 (2011) $27-$4.77 (2011) $1,163 (2013)
2014-2016	Scammers Dominations	- BTC filed for bankruptcy and lost 850,000 BTC (Mt. Cox). - Because it was progressively stolen. - Software wallets are recommended. - USDT and Stellar (XLM) were released in 2014. - ETH was launched in July 2015.	BTC Price $700 - $900
2016-2018	Worldwide	- The software and security were enhanced. - ETH was able to start its own chain. - Binance (BNB) was released in 2017.	BTC price $434-$998 (2016) $998-$20,000 (2017)
2018-present	Recovery	- In pandemic, cryptocurrency has gained popularity. - Market capital passed $3 trillion. - SOL and SAND were launched 2020.	BTC price $3700 (2018) $69,000 (2020-2021)

Sourced from: https://www.cryptovantage.com/

Table 1.2 Overview of Cryptocurrencies

Topic	BTC	LTC	XRP	ETH	USDT
Initial release	Jan 9, 2009	Oct 7, 2011	June, 2012	July 30, 2015	Oct 6, 2014
Launch	July 13, 2010	Apr 28, 2013	Aug 4, 2013	Aug 7, 2015	Feb 25, 2015
Owner	Satoshi Nakamoto	Charlie Lee	David Schwartz, Jed McCaleb and Arthur Britto	Vitalik Buterin.	Brock Pierce, Reeve Collins and Craig Sellars.
Market cap.	$597B	$8B	$25B	$232B	$83B

Sourced from https://coinmarketcap.com/ at July 5, 2023

Historically, the crypto market undergone extreme fluctuations. It highlights that these fluctuations occur simultaneously with real financial events. The first time oil prices rose was in 2015-2016, when they fell from 106 USD to 45 USD (Ahmed & Bouri 2023). The start of ETH and Bitcoin in 2017 was a good time. The crypto market dropped 80% in 2018 and reached a low point (Harbe *et al.* 2022). The first cryptocurrency hack and initial coin offering (ICO) occurred in New Zealand correspond in 2019 (Ahmed & Bouri 2023). The COVID-19 outbreak from March 11, 2020, to January 11, 2021, led to significant drops in Bitcoin prices in June and November 2022, primarily due to bad news and threats. The study suggests that identifying changes in volatility and identifying real events can help manage risk in the volatile cryptocurrency market.

1.1.2 The Characteristics of Cryptocurrency Market

Several characteristics of financial data return series have been extensively documented in recent times. Cryptocurrencies exhibit most characteristics that are found in financial data, such as volatility clustering, fat tails indicating a departure from normalitliy, asymmetry, and persistency (Phillip et al., 2018; Kaseke et al., 2022). According to a study by Chan et al. (2017), the returns of major cryptocurrencies deviate from a normal distribution and do not follow a normal distribution. A notable characteristic of the cryptocurrency market is the abrupt and extraordinary fluctuations in price and volatility that occur within a brief time span. Jiang et al. (2023) assert that the cryptocurrency market is characterized by extreme volatility and frequent shifts in market conditions. As shown in Table 1.1, the cryptocurrency market experienced unforeseen expansion in 2017 and once more in 2020 and 2021. Wars, political conflicts, and financial crises are some of the important events that have been linked to time-varying surges in this market (Dutta & Bouri, 2022). The link between time-varying jumps and the existence of outliers in this market and consequential occurrences such as wars, political conflicts, espionage, and financial crises has been noted (Dutta & Bouri, 2022). As demonstrated in Figures 1.1, 1.2, 1.3, 1.4, and 1.5, certain characteristics of daily log return prices, can be extracted from cryptocurrency return plots. The biggest returns that went up and the biggest returns that went down happened around the same time in 2017, 2020, and 2021. Figures 1.1-1.5 also show major outliers (e.g., substantial increases in 2018 and 2020), structural breaks with higher volatility between 2019 and 2020, and a stabilising trend from 2021 onwards. However, Outliers in time series data sets can be discovered apart from visualization using statistical approaches such as Z-scores or Interquartile Range (IQR), whereas structural breaks can be recognised using methods such as the CUSUM test, or breakpoint analysis. These techniques provide a thorough grasp of significant patterns and properties in the dataset. On the otherhand, Several viable strategies exist to reduce the impact of extreme tail outliers. Winsorization substitutes extreme values with defined percentiles (for example, the first and 99th percentiles), but robust statistical approaches like median-based regression or M-estimators lessen susceptibility to outliers. Transformations like logarithmic or square root can reduce the impact of extreme numbers. Trimming removes a specified proportion of the top and lowest values, whereas clipping caps extreme values at predefined maximum and minimum thresholds. These strategies ensure that the analysis is credible by minimising the disproportionate influence of outliers on the results.

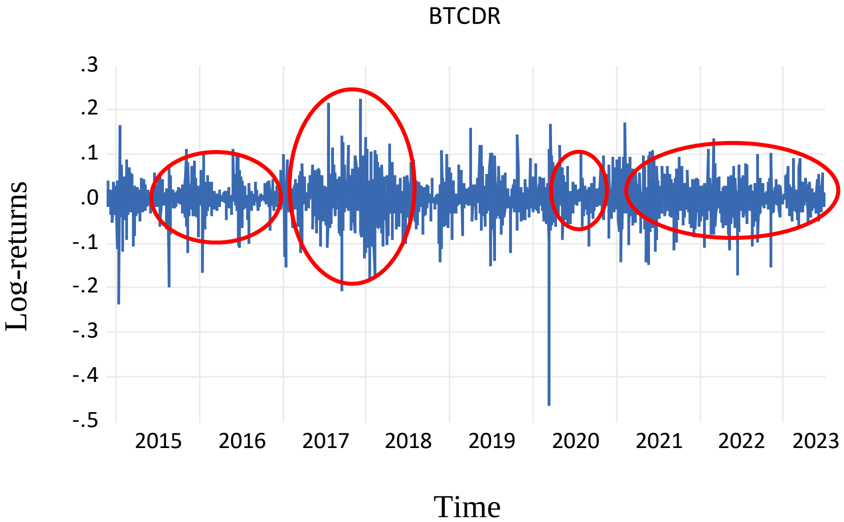

Figure 1.1 BTC daily log-returns

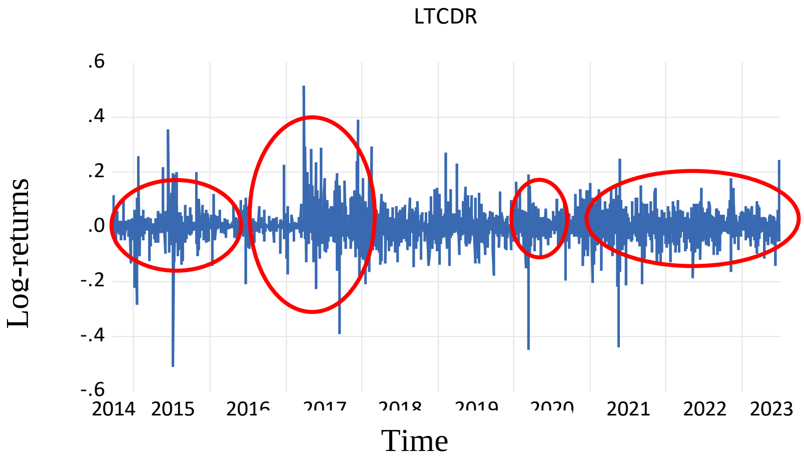

Figure 1.2 LTC daily log-returns

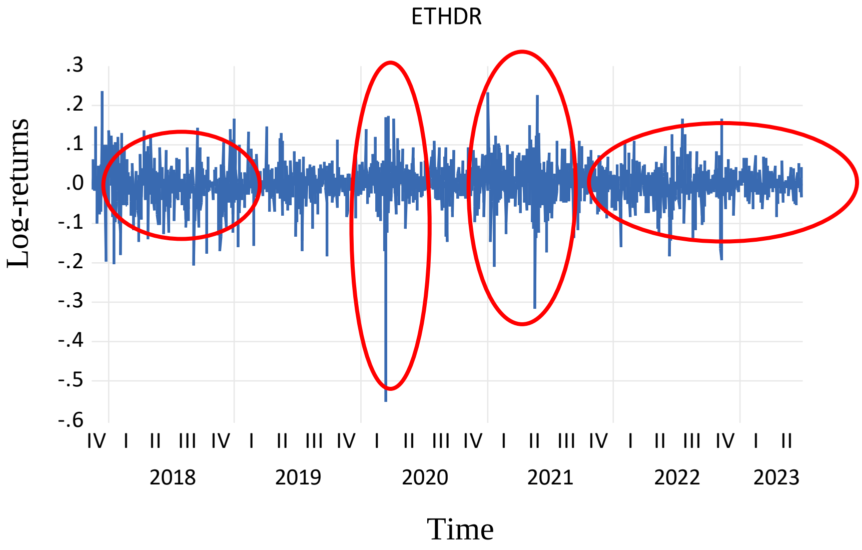

Figure 1.3 ETH daily log-returns

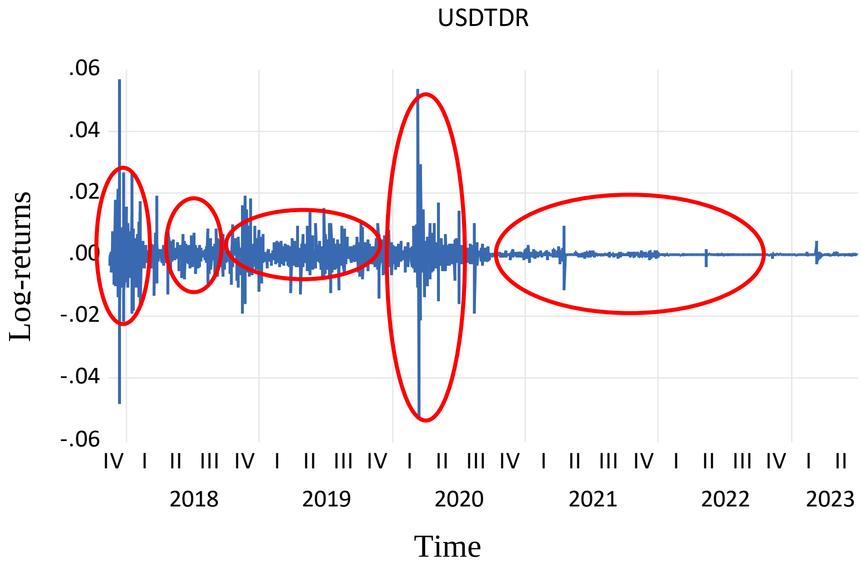

Figure 1.4 USDT daily log-returns

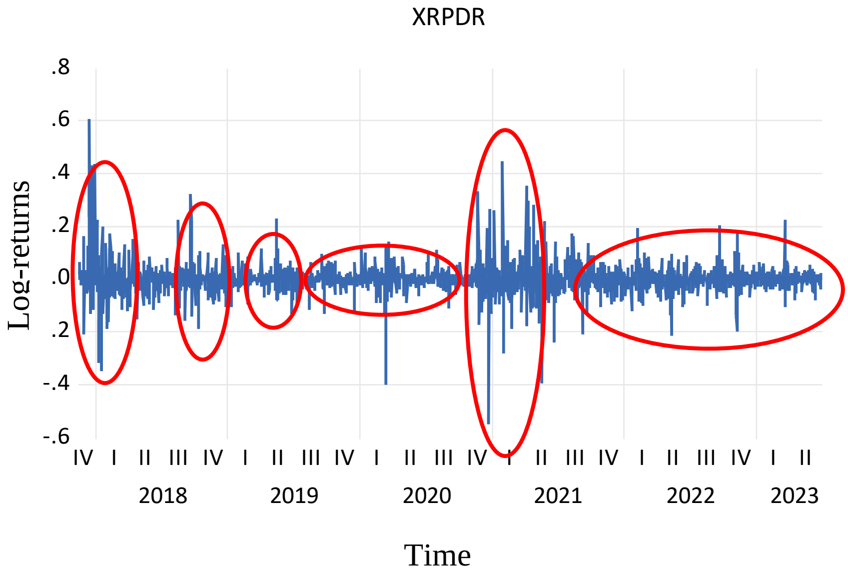

Figure 1.5 XRP daily log-returns

These graphical representations provide valuable insights into the dynamics of each return. Visually, each plotted figure represents a specific cryptocurrency and displays several stylized facts that offer insights into their distinct characteristics within financial markets. Firstly, the return clusters in each figure of both high and low volatility periods demonstrate volatility clustering. Additionally, as observed in each figure, extreme events occur more frequently than expected, indicating fat tails in the return distribution. The persistence of past returns influencing future movements is also visually evident. It is also noteworthy that there is asymmetry in returns, with sharp declines outweighing equivalent gains. The significant changes in each coin's behavior over time indicate shifts. Furthermore, the varying return volatility across different market conditions illustrates heteroskedasticity. The stylized facts observed in the cryptocurrency market and the emergence of this market require careful attention and create opportunities for precise modeling and prediction. Academics, investors, and financial experts have recently become more interested in studies on modelling and predicting the volatility of cryptocurrencies due to the significant and frequent volatitliy fluctuations. Volatility of financial returns such as cryptocurrency is essential in several key areas, including risk management, asset pricing, and portfolio allocation (Caporale and Zekokh, 2019; Yu, 2019; Naimy et al., 2021). Volatility, which is the standard deviation of logarithmic returns, is the most used risk indicator (Ané et al., 2008). It is widely recognized that volatility is a major risk factor because it is directly linked to the deviation of returns from the average (Othman et al., 2019). Risk and return are two essential factors for making financial decisions and forecasting (Sosa et al., 2019).

Good volatility forecasts are therefore necessary (Ardia et al., 2019; Köchling et al., 2020; Naimy et al., 2021). Given the evidence of high volatility in cryptocurrency markets, choosing the right risk modeling strategies is essential when trading digital currencies (Maciel, 2021). The predictiveness of a model is influenced by various factors, such as the nature of the asset, the duration of the in and out samples, and the frequency and time interval of the data (Naimy & Hayek, 2018). When dealing with time series, large sample size can lead to structural changes that contradict the assumption of stationarity (Sen & Das, 2023). Failure to take breaks into account can lead to poor models, bad policies, managerial insights, and inaccurate theory testing (Castle and Hendry, 2019; Charles and Darné, 2019). Also, when time series data becomes contaminated by outliers, it can lead to misleading inferences, poor forecasts, and incorrect model parameters (Ané et al., 2008; Kamranfar et al., 2017, Jiao et al., 2024). In finance, the GARCH model is extensively utilized across a variety of financial data and is considered stable. However, this stability can be called into question when financial crises occur, and structural breaks and outliers emerge. Financial return series that extend across several years are frequently utilized in GARCH model estimates, which exhibit notable persistence (Hillebrand, 2005). Hwang and Valls Pereira (2008) provided empirical support for the notion that ARCH and GARCH estimations tend to forecast higher levels of persistence in situations where the persistence parameters exhibit structural breaks. Structural breaks can lead to incorrect estimations of GARCH models and unreliable volatility predictions (Caporale & Zekokh, 2019).

Conversely, outliers can contribute to excess kurtosis when fitting GARCH models to real-world data (Carnero et al., 2012). Financial data, such as that of cryptocurrencies, display heavy tails, which are supposed to result from outliers and breaks. Dutta and Bouri (2022) argue that the presence of price volatility in major cryptocurrencies underscores the necessity of incorporating time-varying jumps and significant shocks (outliers) into volatility models for cryptocurrency markets. Neglecting to consider sudden changes in the cryptocurrency market may result in wrong forecasts and an incorrect evaluation of the volatility of cryptocurrencies (Mensi et al., 2019). As stated by Kim et al. (2021), GARCH family models are employed to estimate overtime fluctuations in the volatility of cryptocurrencies. There are various fat tail distributions that have been proposed to address the problem of heavy tails in financial data. Franses and Ghijsels (1999) and Petropoulos et al. (2022) discovered that even though the GARCH model can manage heavy-tailed distributions well, the calculated residuals frequently still show excessive kurtosis due to the existence of outliers in the return observations. Leptokurtic return distributions are known for their instability and unpredictability, as they frequently include extreme returns that pose a significant risk (Tiwari et al., 2019). To obtain accurate and valid conclusions from statistical analysis, it is crucial to properly address outlier observations (Song & Kang, 2021; Tan et al., 2021).

Due to the issues caused by outliers and breaks in model estimation and prediction, many researchers have focused on improving the precision of volatility forecasting models. The goal is to enhance the accuracy of volatility predictions in GARCH models by incorporating the impact of breaks and outliers. Bhowmik and Wang (2020) argue that integrating the GARCH model with another statistical methodology can improve performance. Combining univariate models can enhance forecasting, as stated by Catania et al. (2019). However, addressing distinct structural breaks and outlier tests has received limited attention. Outliers have a significant negative impact on parameter change tests, as noted by Song and Kang (2021). Fearnhead and Rigaill (2019) point out that many methods for detecting structural breaks may be ineffective when outliers are present or when the noise follows a heavy-tailed distribution. Even if the model is accurately specified, outliers can introduce bias into the results, thereby reducing the effectiveness of the employed strategies. So, identifying certain outliers may be challenging due to the presence of a masking effect.

1.1.3 The Indicator Saturation Approach

Although there are different approaches on testing for known and unknown breaks and outliers, the indicator saturation (IS) approach by Hendry (1999) is useful when there are several unknown outliers and breaks. The IS approach is a general class of techniques known as indicator saturation estimators that seek robust inference in the presence of uncertain numbers and locations of outliers, breaks, and trend breaks by creating indicators suited to the issue (Castle et al., 2021). These procedures may handle shifts at any time in the sample (even the final observation) and do not require prior knowledge of the numbers, signs, timings, magnitudes, or durations of the breaks. The IS approach employs general-to-specific methodology by starting with a full set of indicators and leaving only the most significant ones without having to specify a minimum break duration, maximum break number, or structural breaks (Pretis et al., 2017; Petropoulos et al., 2022). There are different estimators under IS approach: impulse indicator saturation (IIS) for outliers of Castle and Shephard (2009) and Santos et al. (2008), Step indicator saturation (SIS) for location shifts of Castle et al. (2015) and trend indicator saturation (TIS) for trend breaks of Castle and Hendry (2019). The IS approach is effective and performs better in detecting breaks and outliers when handling high-frequency data with fat-tailed distributions (Castle et al., 2012). The inclusion of the IS approach in this study is motivated by its capability to detect breaks, trend breaks, and outliers in high-frequency data, such as cryptocurrencies known for their fat tails (Troster et al., 2019; Tan et al., 2021; Jiang et al, 2023). Furthermore, the use of a single test methodology strengthens its consideration. This is a direct contrast to other tests that only allow the detection of one feature, such as a break or outlier. If both features are required, two tests with different methodologies would need to be used.

1.1.4 The Winsorization Approach

Due to the high frequency and fat-tailed nature of the data used in this study, caution must be taken when employing the IS approach to simultaneously detect these features. While the IS approach is generally effective, this study improves its detectability by applying the Winsorization method to address the top 1% of extreme outliers in the data symmetrically. Winsorizing, an alternative approach to handling non-normality, maintains the original sample size by replacing the extreme values in the data instead of removing them. Charles P. Winsor (1895–1951), a biostatistician, proposed a technique called winsorization to compensate for the reduction in data produced by the trimming effect (Dixon, 1960).

1.2 Problem Statement

There are various issues in statistics that must be addressed. One such problem is identifying the position and timing of breaks and outliers in dynamically variable time series datasets. This may necessitate presenting hybrid techniques, particularly when dealing with the masking effect. Another significant issue is the necessity to account for variability in datasets in order to enhance model predictions. Accurate modelling of financial time series volatility is critical for risk management and economic decision-making. Standard GARCH models are commonly employed to capture time-varying volatility, however they presume that financial returns have a normal distribution, despite significant evidence of heavy tails and leptokurtic behaviour. There is strong evidence to suggest that cryptocurrencies are typically leptokurtic and heavy-tailed, rather than Gaussian (Troster et al., 2019; Tan et al., 2021; Jiang et al, 2023). Li (2017) argues that assuming a normal distribution for asset returns in risk management is unsuitable, as it overlooks the fundamental characteristics of financial returns. Specifically, Chan et al. (2017) examined the distributional characteristics of cryptocurrencies and found that returns are clearly non-normal, with most cryptocurrencies displaying heavy tails. Furthermore, financial time series frequently contain structural breaks and outliers induced by market shocks, which can mislead typical GARCH models' results. Political unrest, recessions, wars, and natural disasters can cause uncertainty in the financial markets and during the stress phase, prices tend to fluctuate dramatically. On the other hand, the existence of structural breaks and outliers in financial data has been a subject of investigation throughout history. The presence of these features in financial time series data can have a substantial effect on the efficacy of GARCH models. To solve these issues, the Indicator Saturation (IS) strategy is used for log-return data since it can identify multiple characteristics at once, such as structural breaks, trend breaks, and outliers, hence boosting model resilience. The Win-IS approach, an upgraded version of IS, improves this capacity by refining break and outlier identification, resulting in more precise modifications.

Furthermore, the issue of incorporating these characteristics into GARCH models to obtain an efficient predictive model has persisted over time. This thesis creates hybrid models that combine IS and Win-IS techniques with GARCH to better structural break and outlier handling while also increasing distributional resilience. Consequently, the task at hand is to precisely identify these characteristics, which is vital for enhancing the modelling of financial market volatility dynamics and risk estimation.

To tackle this issue, researchers employed separate tests to detect these characteristics. Yet it's possible that the features needed will not be revealed when considering separate tests. Nasir and Ismail (2018) have suggested that to improve the precision of GARCH model estimations, it is possible to detect breaks and outliers concurrently by employing an indicator saturation (IS) strategy. In contrast to their research, this study broadens the identified features to include three distinct characteristics: breaks, trend breaks, and outliers.

As a result, a completely hybridised Win-IS-GARCH model is developed that takes into consideration breaks, trend breaks, outliers, tail outliers, heavy tails, and volatility clustering. By integrating these approaches, the study intends to increase the accuracy and reliability of volatility modelling in financial time series analysis.

1.3 Research Questions

The primary question in this research is how to make GARCH models better at forecasting volatility. This research will, therefore, answer the following questions:

a) Can the indicator saturation approach simultaneously detect structural breaks, trend breaks, and outliers in cryptocurrency data?

b) How to reduce extreme outliers in heavy tail distributions?

c) How can the hybrid between indicator saturation and winsorization approaches be incorporated into the GARCH model?

d) How accurate is the volatility forecasting of the extended GARCH model compared to GARCH models?

1.4 Research objectives

The primary goal is to create a hybrid GARCH model that combines the attributes of GARCH with the IS approach and the winsorization technique, to increase the accuracy of volatility forecasting. Specifically, this study tries:

a) To apply IS approach to cryptocurrency data in detecting breaks, trend breaks and outliers.

b) To enhance the detectability of the IS approach using the Winsorization technique in reducing the impact of extreme outliers in the tails.

c) To integrate the Win, IS, and Win-IS approaches to the GARCH model in examining volatility of cryptocurrency returns.

d) To compare the performance between the standard GARCH and hybrid GARCH models.

1.5 Significance, Scope, and Limitation

The principal aim of this research is to improve the accuracy of predicting the volatility of cryptocurrencies, which is an essential component for academic inquiry and real-world applications. The implication of this work is also significant for investors, traders, regulators, and the economy. It is also good for financial decisions such as portfolio selection, option pricing, risk management, and monetary policy. By combining three approaches (Win, IS, and Win-IS) with GARCH with three distributions, the proposed hybrid GARCH model may be able to overcome the shortcomings of the conventional GARCH model. This model can provide financial institutions and investors with accurate predictions of market volatility, thus aiding in the risk management and decision-making processes. Moreover, to evaluate the effectiveness of the hybrid model, it might be applied to a different real-world financial dataset.

To the best of our knowledge, Nasir and Ismail (2018) modelled volatility data using IS and GARCH. The study is nevertheless constrained by its initial emphasis on two distinct IS approaches—IIS and SIS. Furthermore, they update the GARCH model to incorporate the breaks captured by SIS and modify the outliers. Modelling the volatility of Shariah-compliant indices is the third limit, but no prediction was conducted. Therefore, a distinct reconstruction of that model is required. This study deviates from the research conducted by Nasir and Ismail (2018) in various respects. Firstly, the IS technique is used to simultaneously discover three distinct features: breaks, trend breaks, and outliers, hence expanding the identified features. Furthermore, it improves the detectability of IS approach by utilizing the Winsorization method to address extreme outliers in the tail. Furthermore, it expands upon this methodology by integrating GARCH models to estimate and forecast volatility of cryptocurrency. Finally, it uses dummy variables to account for all these properties in GARCH. In addition, this study develops hybrid GARCH-type models that are limited to the empirical application of five cryptocurrencies and consideration of four distinct types of univariate GARCH type models with three error distributions. The analysis of each coin is therefore univariate; no other financial assets were considered. The data used in this study consists of time series datasets from major cryptocurrencies, specifically Bitcoin, Ethereum, Litecoin, Tether, and Ripple as financial data from November 2014 to June 2023, with an emphasis on price variations across time. These statistics contain historical price fluctuations, trade volumes, and other important financial metrics. Cryptocurrency data is included in this study because it shows extremely volatile and unexpected patterns, making it a good subject for analysing fluctuations and developing prediction algorithms. Extreme price fluctuations bring issues such as the masking effect, necessitating the development of strong statistical and hybrid techniques for better identification and forecasting.

When developing a hybrid GARCH model that combines change point detection and extreme outlier treatment, it is important to consider certain key assumptions. These assumptions include the requirement for the time series to be stationary and the presence of conditional heteroskedasticity (Brooks, 2019). The study also assumes the existence of breaks, outliers, and volatility clustering behaviour in the log returns. This underscores another assumption that the models will perform better when breaks and outliers are accounted for. So, accurate specification of the model, precise identification of change points, and effective treatment of outliers are also required. These assumptions are essential for guaranteeing the accuracy and strength of the model. However, there are other limitations that must be taken into account. These elements include the complex structure of the model such as non linearity, presence of structural breaks and outliers, which might hinder the estimation of parameters and increase computational requirements, as well as the requirement for high-quality data.

Therefore, we identify outliers, breaks, and trend breaks using the IS technique and then we include them into our analysis using dummy variables. Regarding outliers, we have categorized them as either extreme or non-extreme outliers. Multiple criteria may be used to identify extreme outliers. The Interquartile Range (IQR) method finds data points that are more than three times the IQR below the first quartile (Q1) or above the third quartile (Q3) as severe outliers. The Grubbs' Test is used to identify exceptional outliers by comparing the test statistic to the crucial value at a certain level of significance. In the Boxplot approach, data points that exceed a distance of 3 times the interquartile range (IQR) from the neighboring quartile are considered severe outliers. These methodologies provide robust frameworks for finding extraordinary outliers in various data analysis contexts. Here, we treat extreme outliers using the Winsorization method. Thus, the 1st and 99th percentiles were used to determine whether an observation is classified as an extreme outlier or not. Therefore, we have addressed outliers by identifying them through IIS and including them into the analysis using dummy variables, and by handling tail outliers by Winsorization.

1.6 Research Contribution

The primary contribution of this study is the creation of hybrid GARCH-type models. Given the highly volatile nature of the cryptocurrency market and its tendency to experience significant price changes, the conventional GARCH model, which assumes a Gaussian distribution, may not accurately reflect the occurrence of extreme events and the clustering of volatility in this market. This requires contributing to research by constructing a hybrid GARCH model that combines the capabilities of the Win, IS, and the enhanced IS approach (Win-IS). The Winsorization strategy is employed to mitigate extreme tail outliers and to enhance the detectability of the IS approach. As a result, the study introduced an improved version of IS approach referred as (Win-IS) and proposed a grand total of 432 models, with 60 using the GARCH-dist theme as benchmark models across the study, 60 using the Win-GARCH-dist theme, 180 using the IS-GARCH-dist theme, and the remaining 132 using the Win-IS-GARCH-dist theme. This contribution focuses on of constructing hybrid GARCH models developed for cryptocurrency markets that can also be applied to other financial datasets, providing applications in risk management, volatility forecasting, and portfolio optimisation across other asset classes.

1.7 Research Structure

The current study comprised seven chapters. Chapter 1 encompasses the introduction, problem statement, research aims, scope and constraints of the study, and research contributions. Chapter 2 provides a thorough analysis of volatility and its characteristics in relation to cryptocurrencies. It explores the complexities of understanding cryptocurrency and the current methods used to handle disruptions and anomalies in financial data. Moreover, it delves into the advancements achieved so far in the field of volatility enhancements. Chapter 3 outlines the proposed research methodology used to achieve the study's goals, the theoretical basis of the models and approaches used, and the details of the datasets used. Chapter 4 provides results and a detailed analysis of the empirical application of the IS and Win-IS approaches to identifying breaks, trend breaks, and outliers in cryptocurrency data. Chapter 5 presents the results of the in-sample performances of the 432 benchmark and hybrid GARCH type models, as well as the diagnostics and selection criteria for each. Chapter 5 also presents the optimal in-sample models for each theme. Chapter 6 presents the out-of-sample performances of the models presented in Chapter 5, a comparison of models for predicting volatility in the cryptocurrency market, and optimal out-of-sample models. Chapter 6 answers the question of which treatments yield the best predictive performance of the treatments applied, including Winsorizing, accounting for breaks, accounting for trend breaks, and accounting for outliers. Chapter 7 concludes the study and presents a thorough overview of the research findings and limitations, along with recommendations for further research.

1.8 Chapter Summary

This chapter presents a synopsis of the overall concept and establishes the framework for the present study. The chapter commences by presenting a comprehensive outline of the cryptocurrency market, including its attributes, historical background, and significance as financial volatility in diverse domains. It then delves into the ramifications of disregarding breaks and outliers in volatility estimations, the methodologies that will be employed to enhance cryptocurrency volatility, and the overarching objectives that will steer the entirety of this investigation. Additional details pertaining to contemporary literature will be presented in the subsequent chapter.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter provides a review of the volatility models and distributions that have been used historically. It explores the impact of outliers and structural breaks on the estimation and forecasting of volatility models. Additionally, it discusses strategies for improving the volatility forecasting of GARCH models in the cryptocurrency market. The chapter also examines existing approaches for detecting breaks and outliers, as well as the cryptocurrency breaks and outliers that have been reported in the literature. Finally, the chapter concludes with a review of using an indicator saturation approach and Winsorization techniques in financial applications.

2.2 Volatility Classification and Definition

Volatility modelling and forecasting are important areas of study that have both theoretical and practical implications. These concepts have been widely applied in financial contexts. According to Brooks (2019), volatility is defined as the extent to which a series is highly inconsistent over time, as measured by its standard deviation or variance. In financial applications, volatility is typically categorized into three groups based on the data sources used. Tsay (2012) states that the volatility of an asset can be evaluated using the prices of its stock, derivatives, or both, including the daily return for each trading day. Therefore, it can be derived the following three different measures of volatility for an asset.

Conditional Volatility: This kind of volatility is defined as the conditional standard deviation of daily returns.

Implied Volatility: Is calculated using prices from the options markets and a pricing model, such as the Black-Scholes pricing formula. Implied volatility is frequently criticized for being model dependent. The volatility index (VIX) of Chicago Board Options Exchange (CBOE) is an implied volatility.

Realized Volatility: This can be estimated as the daily volatility using intraday returns, such as 5-minute returns, since high-frequency financial data is readily available.

This thesis employs the concept of conditional volatility. Volatility is estimated using the sample standard deviation of log-returns in cases where computation and modelling are involved. However, it is also common for practitioners to interpret volatility as unconditional variance (Stawiarski, 2015).

2.2.1 Volatility Stylized facts

There are certain stylized facts that are examined to gain insights into the unique characteristics and dynamics of volatility in financial data, including cryptocurrency markets. By analyzing these stylized facts, analysts can develop a clearer understanding of the patterns, trends, and risks associated with price fluctuations in volatile assets. Here are some of the common stylized facts of volatility that have been observed in financial data.

2.2.2 Volatility Clustering

Volatility clustering is a common phenomenon in financial time series, such as stock prices, interest rates, currency exchange rates, inflation rates, and cryptocurrencies. According to Brooks (2019), volatility clustering, also known as volatility pooling, refers to the tendency for volatility to occur in clusters. In other words, small returns are likely to be followed by small returns, while large returns are likely to be followed by large returns. These clusters of information arrivals that drive price changes are not evenly distributed over time. This characteristic is likely to be present in all assets return series in finance. Maciel (2021) further asserts that volatility clustering is a stylized fact observed in the log-returns of digital coins, like what is commonly observed in the log-returns of traditional financial asset prices.

2.2.3 Leverage affects.

Black (1976) introduced the leverage effect, which is considered one of the most significant stylized features of financial time series. The leverage effect is caused by price changes and refers to the phenomenon where unexpected negative shocks have a greater impact on volatility than expected positive shocks. Brooks (2019) defines the leverage effect as the tendency for volatility to increase more after a significant price decline compared to a price rise of the same magnitude. In other words, the leverage effect describes how asset volatility tends to increase more after a significant decline in asset price than after a gain of the same size. In GARCH models, the terms "symmetric" and "asymmetric" are used to describe how volatility responds to both positive and negative shocks. Engle and Ng (1993) used the terms asymmetric and leverage effect interchangeably, while some researchers, like Caporin and Costola (2019), argue that they are two distinct concepts. However, Engle and Patton (2007) emphasized that asymmetry is sometimes attributed to the leverage effect and other times to the effect of the risk premium. Therefore, it is common to use leverage as a synonym for asymmetry from an empirical standpoint. To determine whether an asymmetric model is necessary for a particular series or if a symmetric GARCH model is sufficient, it is important to utilize the sign and size bias tests introduced by Engle and Ng (1993).

2.2.4 Leptokurtic

As stated by Brooks (2019), leptokurtosis is often used to describe the phenomenon where financial asset returns have distributions with fat tails and excess mean peaking. In comparison to the normal distribution, the fat tails in financial time series are thicker. Doornik and Ooms (2005) also mention that financial data frequently displays thick tails and volatility clustering. To tackle this issue, two possible approaches come to mind: opting for a distribution with fatter tails, like the student-t distribution, or dealing with outliers.

2.2.5 Persistence

The persistence of a GARCH model is determined by the decay of large volatilities after a shock. According to Kaseke et al. (2022), the volatility of a GARCH (1,1) model is measured by the sum of α and β, which should not exceed one. A higher value indicates a higher level of variance/volatility with increased persistence.

2.2.6 Long Memory

The long memory property refers to the tendency of financial time series data to exhibit prolonged periods of either high or low volatility. This property greatly influences market dynamics and long-term risk management strategies. The parameter d in the FIGARCH model of Baillie et al. (1996) is primarily used to measure the degree of long memory.

2.2.7 Half Life

In a volatility model, the "half-life" of volatility is used as an additional indicator of persistence. The half-life (HL) of a volatile refers to the time it takes for high volatility to decrease to half of its original value. Engle and Patton (2007) defined the half-life as the duration required for volatility to reach halfway back to its unconditional mean. The formulas below can be used to calculate the half-life of volatility for different GARCH models. The half-life of GARCH (1,1) model is $HL = \frac{ln\, 0.05}{ln(\alpha_1+\beta_1)}$, the half-life of EGARCH (1,1) model is $HL = \frac{ln\, 0.05}{ln(\beta_1)}$, and the half-life of GJR-GARCH (1,1) model is: $HL = \frac{ln\, 0.05}{ln\left(\alpha_1+\beta_1+\frac{1}{2}\gamma\right)}$.

Thus, as we explained, notable stylized facts have been observed in financial data. It is worth noting that these characteristics have also been observed and reported in data pertaining to the cryptocurrency market. Hence, it can be argued that digital coins demonstrate similar stylized facts as other financial assets. Numerous researchers have highlighted these facts in the context of the cryptocurrency market. For example, Phillip et al. (2018), Othman et al. (2019), and Alqaralleh et al. (2020) have explored the time-varying nature of certain cryptocurrencies and discovered the presence of long memory and leverage effects. Mensi et al. (2019) have documented the dual long memory property of BTC and ETH. Segnon and Bekiros (2020) have reported the asymmetry and long memory properties of BTC. Dutta and Bouri (2022) have shown that BTC, ETH, and LTC exhibit high persistence values. John et al. (2019) conducted a study on the half-life property of BTC, XRP, and LTC, revealing that these cryptocurrencies have a short half-life. In addition, nearly all papers demonstrate that cryptocurrency returns exhibit fat tails, volatility clustering behavior, excess kurtosis, and high volatility. However, it should be noted that not all these characteristics may be present in every cryptocurrency. According to Tiwari et al. (2019), BTC and LTC, for example, do not exhibit significant leverage effects and do not behave like stock prices. To summarize, this study examines six specific volatility patterns observed in cryptocurrencies, including volatility clustering, leverage effects, leptokurtosis, persistence, long memory property, and half-life.

2.3 Volatility Models

2.3.1 Classification of Volatility Models

There are numerous models available in the literature to represent the stylized characteristics of volatility. Researchers have classified these models in various ways. For instance, Engle and Patton (2007) classified volatility models into two general classes. The first class includes models such as ARCH and GARCH, which explicitly represent the conditional variance as a function of observables. The second class consists of volatility models that go beyond observables-based functions, often referred to as latent volatility models or, more precisely, stochastic volatility models. In other classifications, volatility forecasting models have been widely grouped into historical, stochastic, and implied volatility models (Naimy & Hayek, 2018). These models were developed based on the observed patterns in the data, resulting in a wide range of models available for practitioners. Among these models, ARCH-type models have been particularly popular (Engle and Patton, 2007) due to their ability to capture and represent these observed patterns. However, evaluating the effectiveness of a volatility forecast is not straightforward since volatility itself is not directly observed. In this study, we will focus on different types of ARCH class conditional volatility models. The following section discusses the historical development of ARCH class models.

2.3.2 Overview of (G)ARCH Class Models

Engle (1982) introduced a new class of stochastic processes called autoregressive conditional heteroscedastic (ARCH) processes. This was done to address the questionable assumption made by conventional econometric models, which assumes a constant forecast variance throughout a given period and to capture leptokurtic and volatility clustering. ARCH processes have a mean of zero, are serially