INTERMINABLE LONG MEMORY MODEL AND ITS HYBRID FOR TIME SERIES MODELING

JIBRIN SANUSI ALHAJI

UNIVERSITI SAINS MALAYSIA

2019

INTERMINABLE LONG MEMORY MODEL AND ITS HYBRID FOR TIME SERIES MODELING

by

JIBRIN SANUSI ALHAJI

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

August 2019

ACKNOWLEDGEMENT

All praise to Allah, the master of the universe, for giving me the opportunity to study in Malaysia and in particular School of Mathematical Sciences, Universiti Sains Malaysia (USM). First and foremost, I would like to thanks my supervisor, Dr Rosmanjawati Abdul Rahman for her endurance, encouragement, advice, support and professional guidance during this study. I have benefited from her experience and knowledge in the field of econometrics, time series analysis and research in general. I personally thank her for recommending me to the school for sponsorship to attend international conferences and in particular the very important conference that advanced my knowledge in R software. Today I developed an R package for our research and future time series analysis. My supervisor is more than a supervisor, she supported me as one of her family members during our formal meetings. The almighty Allah will reward you abundantly.

I would like to thank the Dean of the School of Mathematical Sciences, USM, lecturers and all postgraduate students for supporting me directly or indirectly. I am passionately indebted to my mentors and friends who have helped me in many capacity, some of which include, Yakubu Musa, S. U. Gulumbe, Sani Salihu Abubakar, Adamu Umar, Al Mansour (Kwankwasiyya), Musa Lawan, Alhassan Adamu, Nura Baba, Ibrahim Abdallah Inuwa, Suleiman Kawuwa, Gagman, Alh Rayyanu (Wanban Labar), Mal Bello,Muhammadu Sani, Mai Dodo Adam Maiyaki, Musa Marut and Buba Audu.

I would like to thank my late parents, beloved wife, Maryam Haruna Sanusi and my daughter, Halimatu Saadiyya and all members of my family. Their care, support, prayers, encouragement and sacrifices are beyond expectations. Therefore, I would like to dedicate this work to them and all Muslims ummah around the world. Finally, I would like to thank Kano University of Science and Technology (KUST) Wudil, Kano state Nigeria. It would have been impossible to complete this research work without the support given to me and I will be grateful to them for the rest of my life. May Allah continue to bless Nigeria, Ameen.

TABLE OF CONTENTS

CHAPTER 3 - THE PROPOSED FRACTIONAL FILTER, MODELS AND R PACKAGE

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ACF	Autocorrelation function
ADF	Augmented Dickey Fuller
ADRs	American depository receipts
AGARCH	Adaptive-GARCH
AIC	Akaike information criteria
APARCH	Asymmetric power ARCH
AR	Autoregressive
ARCH-LM	Autoregressive conditional heteroscedastic-lagrange multiplier
ARFIMA	Autoregressive fractional integral moving average
ArfimaMLM	Arfima-Multilevel modelling
ARFURIMA	Autoregressive fractional unit root integral moving average
ARIMA	Autoregressive integral moving average
ARMA	Autoregressive moving average
ARTFIMA	Autoregressive tempered fractional integral moving average
CUSUM	Cumulative sum
DFA	Detrended fluctuation analyses
EHAR	Extended HAR
EMH	Efficient Market Hypotheses
ETNs	Exchange-traded notes
EU	European union
FDR	Fractional difference return
FELW	Fully Extended Local Whittle
FI	Fractional integral

FIGARCH	Fractional integrated GARCH
FURI	Fractional unit root integral
GARCH	Generalized autoregressive conditional heteroscedastic
GHE	Generalized HE
GPH	Geweke and Porter-Hudak
HAR	Heterogeneous AR
HE	Hurst Exponent
ILM	Interminable LM
KLCI	Kuala Lumpur Composite Index
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LM	Long memory
LMMMs	Long memory mean models
LWE	Local Whittle Estimator
MA	Moving average
MAE	Mean absolute error
MAPE	Mean absolute percentage error
ME	Mean error
MRS	Modified Rescaled Range
MSE	Mean Square Error
NMGARCH	Normal mixture GARCH
OLS	Ordinary least square
OPEC	Organization of Petroleum Exporting Countries
PACF	Partial ACF
RMSE	Root mean square error
RS	Regime switching

RV	Realized volatility
R/S	Rescaled range statistics
SARFIMA	Seasonal autoregressive fractional integral moving average
SEMIFAR	Semiparametric fractional autoregressive
SFI	Seasonal FI
TFI	Tempered FI
TSPI	Tehran stock price index
TVARFIMA	Time varying ARFIMA
TVP-EHAR	Time Varying Parameter HAR
UK	United Kingdom
UKP	UK Pound
UMTS	Universal Mobile Telecommunication System
US	United State
USD	US Dollar
USM	Universiti Sains Malaysia
VSE	Vietnam stock exchange

MODEL MEMORI PANJANG BERJELA DAN HIBRIDNYA UNTUK PEMODELAN SIRI MASA

ABSTRAK

Indeks kewangan dan ekonomi adalah tidak malar, saling berhubung kait dalam tempoh yang panjang dan bersifat turun naik (ketidaktentuan). Ini merupakan masalah yang serius, disebabkan ianya masing-masing memberi kesan kepada ketepatan, kesahan dan kebolehpercayaan suai padan model dan ramalan siri yang dikaji. Oleh itu, kajian ini mencadangkan turas pecahan bagi menguraikan siri masa yang tidak malar dan bermemori panjang berjela (ILM) dengan nilai pecahan terbezakan dalam selang $1 < d < 2$ kepada proses hingar putih. Mulanya, model ILM, diberi nama Autoregressive Fractional Unit Root Integral Moving Average (ARFURIMA) dibangunkan. Seterusnya, setiap ciri asas dan asimptot bagi ARFURIMA yang dicadangkan dan nilai turas pecahannya masing-masing dijanakan. Ini diikuti dengan cadangan turas pecahan terbezakan pulangan dan membangunkan model ARFURIMA-GARCH, yang mana komponen GARCH akan menangkap kesan turun naik di dalam siri tersebut. Kesannya, pakej **arfurima** and **arfurimafdrgarch** dalam R telah dibangunkan untuk larian turas pecahan yang dicadangkan, dan untuk padan suai model ARFURIMA dan ARFURIMA-GARCH yang telah dibangunkan. Turas yang dicadangkan (turas pecahan yang mana juga dipanggil turas bermemori panjang) dibandingkan dengan kedua-dua turas terbezakan bagi ARIMA (yang mana bukan turas pecahan) dan dua turas pecahan dalam model ARFIMA dan ARTFIMA. Keputusan mendapati turas yang dicadangkan lebih baik dari segi ukuran variabiliti dan nilai AIC. Juga, model ARFURIMA dibandingkan dengan model-model ARIMA, ARFIMA dan ARTFIMA dengan membuat padanan model terhadap sepuluh data harian kewangan dan indeks ekonomi yang berlainan. Keputusan menunjukkan model ARFURIMA yang dicadangkan adalah lebih bagus dari segi kedua-dua ukuran statistik (AIC, log-kebolehjadian, ME, RMSE, dan MAE) dan ujian signifikan (DM, diagnostik dan ujian bias tanda). Juga, didapati bahawa ARFURIMA-GARCH iaitu model hibrid yang dicadangkan adalah lebih bagus daripada model ARFIMA-GARCH, dari segi padanan dan ujian diagnostik dan menghasilkan ramalan yang lebih baik seperti yang disahkan oleh ujian Diebold and Mariano. Dengan ini, kesimpulan yang dibuat ialah model hibrid yang dicadangkan merupakan model yang baik bagi pemodelan dan peramalan model min-volatiliti bagi sesuatu data ekonomi dan kewangan. Kajian ini menunjukkan turas yang dicadangkan, model min dan model min-volatiliti yang digunakan adalah lebih baik bagi pemodelan dan peramalan siri masa dengan kesan jangka panjang berjela (ILM). Implikasi dari kajian ini ialah, dengan penambahbaikan pemodelan dan peramalan suatu model bermemori panjang, ianya akan memanfaatkan institusi kewangan, pelabur dan pedagang pasaran saham bagi mengawal kesan kerugian dalam dagangan saham. Tambahan pula, ia akan membantu penggubal polisi bagi membuat keputusan yang tepat yang akan memberi kesan kepada pertumbuhan ekonomi dan akhirnya, kajian ini turut menyumbang kepada karya berkaitan pemasalahan bermemori panjang di dalam data siri masa.

INTERMINABLE LONG MEMORY MODEL AND ITS HYBRID FOR TIME SERIES MODELING

ABSTRACT

The financial and economic indices are nonstationary, long-range dependence and volatile. These are very serious problems because each affect the accuracy, validity and reliability of model fitting and the forecasting of the studied series. In view of this, our current study proposes fractional filter to decompose the nonstationary and Interminable Long Memory (ILM) time series with fractional differencing value in the interval of $1 < d < 2$ into a white noise process. First, the ILM model, named Auto Regressive Fractional Unit Root Integral Moving Average (ARFURIMA) is developed. Next, each of the basic and asymptotic properties of the proposed ARFURIMA and its fractional filter were derived respectively. This follows by proposing the fractional differenced return filter and developing the ARFURIMA-GARCH model, where the GARCH component will adequately capture the volatility in the series. Consequently, the **arfurima** and **arfurimafdrgarch** packages in R were developed to run the proposed fractional filters, and to fit the ARFURIMA and ARFURIMA-GARCH models developed. The proposed filter (a fractional filter which can also be called a ILM filter) is compared with both the first differenced filter for ARIMA (which is not a fractional filter) and two fractional filters in ARFIMA and ARTFIMA models. Results found that the proposed fractional filter is better both in terms of minimum measures of variability and AIC values. The ARFURIMA models are compared with the ARIMA, ARFIMA and ARTFIMA models by fitting to ten different financial and economic index. Results show that the proposed ARFURIMA is better in terms of both statistical measures (AIC, Log-likelihood, ME, RMSE, and

MAE) and significant tests (DM, Diagnostic and Sign Bias). The proposed ARFURIMA-GARCH (the hybrid model) is better than the ARFIMA-GARCH, in terms of fit and diagnostic tests and produces better forecast which is confirmed by the Diebold and Mariano tests. With this, we conclude that the suggested hybrid model is good candidate for modeling and forecasting the mean-volatility of financial and economic data. This study shows that the proposed filter, the mean model and the mean-volatility model used are better in modeling and forecasting time series with ILM. The implications of this study are by improving modeling and forecasting of long memory, it will benefit financial intitutions, stock market investments and traders to control their loss in stock tradings. Moreover, it will help the policy makers to make the right decisions that will affect the economic growth and eventually, this study contributes to the literatures of long memory problems in time series data.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Financial indices and economic growth are highly related. Financial stress influences economic activities in countries and eventually affected the economic growth. In modeling financial indices or data, time series analyses usually are used. It comprises of methods for analysing time series data in order to extract meaningful statistics and other characteristics of the data (Box and Jenkins, 1976). Eventually, time series forecasting is used to predict future values based on previously observed values.

One of the characteristics shown in time series data is Long Memory (LM). LM occurs when autocorrelations in time series are different from zero, large and occurs for many lags (Fleming and Kirby (2011) and Ho *et al.* (2013)). LM, also called long-range dependence, is a statistical property that may arise in the financial and economic index. It was first discovered in geophysical data by Hurst (1951). Whittle (1956), Mandelbrot (1972) and Mcleod and Hipel (1978) were few among early studies on this notable property. According to Qu (2011), if the spectral density of a scalar K is proportional to K^{-2d} as K tend to zero, the process is said to have LM with d as the LM parameter and also called as the degree of fractional differencing. Other explanations of this LM both in time and frequency domain can be found in Granger and Joyeux (1980), Hosking (1981), Geweke and Porter-Hudak (1983), Hamilton (1989), Chen and Tiao (1990), Hurvich *et al.* (1998) and Bai and Perron (2003).

Brockwell and Davis (2016) and Haldrop and Vera Valdes (2017) suggested that time series to be considered for modelling and forecasting should be stationary.

Also, as trend and seasonal variation are adjusted or removed by using transformation methods, the LM is another emerging variation that is important to be eliminated from any economic and financial data that exhibits dependency. There are exist transformation filters or methods that were developed for obtaining the stationary component of the non-stationary economic and financial data (see Granger and Newbold (2014)). In the literatures, differencing, de-trending and transformation are some of the procedures used for eliminating variability and noise signal. Therefore, it is pertinent to note that for a non-stationary series, eliminating or reducing excess noise signal is a significant stage of every modelling and forecasting of time series.

According to Granger and Joyeux (1980) and Hosking (1981), when time series exhibit LM behavior, fractional differencing is the appropriate transformation approach that can avert over differencing. In view of this, they introduced a fractional filter for differencing dependence or Fractional Integeral (FI) series. The filter is suitable for eliminating LM component of a data that has memory parameter in the interval of $0 < d < 1$ and stationary series is obtained due to the fractional differencing. Also, Porter-Hudak (1990) introduced the seasonal fractional filter for denoising Seasonal FI (SFI) series that exhibits LM and seasonal trends while Beran (1999) suggested a filter for obtaining stationary series from deterministic and stochastic trend specifically if there was no prior information about the series. Furthermore, Meerschaert *et al.* (2014) developed filter for decomposing Tempered FI (TFI) series that exhibits irregular trend. All the filters were introduced to form different Long Memory Mean Models (LMMMs) such as the Auto Regressive Fractional Integral Moving Average (ARFIMA), the Seasonal Auto Regressive Fractional Integral Moving Average (SARFIMA), Semiparametric Fractional Auto Regressive (SEMIFAR) and the Auto Regressive Tempered Fractional Integral

Moving Average (ARTFIMA) models respectively. Similarly, Pumi *et al.* (2019) developed Beta ARFIMA (βARFIMA) for studying continuous random variables or simply time series that exists in the unit interval $(0,1)$.

In literatures, it is shown that financial and price index exhibited LM. This type of series are also called the FI process. Also, the residuals of the models fitted to the financial and economic indices are affected by heteroscedasticity. In view of this, Baillie *et al.* (1996) extended the ARFIMA process and the Generalized Autoregressive Conditional Heteroscedastic (GARCH) to study both the LM and time-dependent heteroscedasticity in inflation series. Meanwhile, Cheung and Chung (2009) have used the ARFIMA model with Normal Mixture GARCH (NM-GARCH) process, called the ARFIMA-NM-GARCH model to study the series that exhibited LM and volatility. Leite *et al.* (2009) applied the ARFIMA-GARCH model to account for both LM and conditional volatility in heart rate variability records. However, Fofana *et al.* (2014) have formed a hybrid specification called the Regime Switching ARFIMA-GARCH (RS-ARFIMA-GARCH) models to account for structural change in a series that exhibited both LM and volatility.

1.2 Problem Statement

The Markovian theory, geometric Brownian motion, random walk and Efficient Market Hypothesis (EMH) emphasized that market indices including financial and price indices are memoryless, independent, unpredictable and adjust quickly to new information in the markets respectively. However, it is a known fact that this established theories are constantly in conflict with established behavior of the market indices as shown by Cont (2005), Baillie and Kapetanios (2008) and Arouri *et al.* (2012). Their studies highlighted that most of the daily market indices (financial and economic) specifically those from emerging countries exhibits nonstationary, LM, volatility and sometime are dominated by two or all the mentioned features. Therefore, these lead to the current developments and includes the following:

1. The recent studies of LM series were posed with series that exhibited LM parameter value of $1 < d < 2$. One of the suggested solution is to obtain the first difference then apply the condition of fractional d. Even though the approach works to some certain extent, efficient estimation is still not guaranteed. Therefore, in this study, a fractional differencing filter is proposed to handle type of Interminable LM (ILM) for nonstationary time series. The proposed filter is presented first for fractional unit root differencing of Fractional Unit Root Integral (FURI) series. This filter is used in this study to fractionally difference a series with fractional differencing value in the interval of $1 < d < 2$. It can be used to obtain stationary series with minimum values of certain statistics such as variance, standard deviation and autocovariance. Also, the proposed filter would entirely eliminate linear trend, removes the ILM component and provides a filtered or stationary series for the parameter estimation of the Autoregressive Moving Average (ARMA) models.
2. Next, the conditional mean specification for ILM named Autoregressive Fractional Unit Root Integral Moving Average (ARFURIMA) is presented, and to be used for modeling nonstationary univariate time series with fractional differencing value in the interval of $1 < d < 2$. The proposed ARFURIMA model is intended to effectively handled any time series with fractional differencing value greater than unity and in particular $1 < d < 2$. The ARFURIMA model will be better in capturing some patterns in the FURI process which may not be handled adequately by ARFIMA model. Essentially,

this may lead to a free error statistical tests and a choice of stable and efficient model(s). Therefore, this current study applies the ARIMA, ARFIMA, ARTFIMA and ARFURIMA models to study the financial and economic data that exhibits ILM and selects the best model(s) based on information criteria, log-likelihood, accuracy measures and forecast accuracy test.

3. Also in literatures, it was showed that the residuals of LMMMs regularly shows the presence of heteroscedasticity and this may always lead to bizarre forecast. Combining the mean (ARFIMA) and variance (GARCH) model to form hybrid models and study the LM and heteroscedasticity simultaneously provides a substantial improvements in terms of fit and diagnostics test (see Ballie *et al.* (1996), Cheung and Chung (2009) and Fofana *et al.* (2014)). However, the ARFIMA-GARCH models were designed to study time series with fractional differencing value less than unity ($0 < d < 1$) only. Our study intend to introduce the hybrid model, called the ARFURIMA-GARCH which is significant to study the ILM-volatile process with fractional difference value between $1 < d < 2$.

4. Finally, one of the major issue with modelling LM series is the availability of packages to solve them especially in the environment of R statistical software. There are few packages that implement the fractional autoregressive model. Most of the existing packages only implement the situation where d is bounded between 0 and 1. The package **fracdiff** developed by Fraley *et al.* (2012) only considered the case of $0 < d < 1$. It is specifically restricted to $0 < d < 0.5$ where it has been established that the LM series will be stationary. For $d > 0.5$, the LM series is non-stationary and the covariance will not be estimable. Thus, the package **fracdiff** cannot be used for the case $d > 0.5$. The package **forecast**

by Hyndman *et. al.* (2018) depends on **fracdiff** and thus carried over its deficiencies. The package **rugarch** by Ghalanos (2018) is also limited to $0 < d < 1$ and also restricted to $0 < d < 0.5$. Therefore, we present the R package **arfurima** and **arfurimafdrgarch** which are used for implementing the fractional filters, ARFURIMA and ARFURIMA-GARCH models in this study.

1.3 Objectives of the Study

This study focuses on the LM and the hybrid models. In doing so, the objectives are:

1. To propose the $\left[(1-L)\left(1-d^*(1+L)\right)\right]Y_t$ filter in ARFURIMA models and simulates nonstationary FURI series.
2. To derive the basic and asymptotic properties of the proposed ARFURIMA model and the fractional filter respectively, and later to compare the numerical statistical properties of the filter by using the first difference and fractional difference filters.
3. To compare the ARFURIMA model in 1 with the ARIMA, ARFIMA and ARTFIMA models.
4. To propose the ARFURIMA-GARCH models (hybrid models) and compare with the ARFIMA-GARCH.
5. To built an R package **arfurima** and **arfurimafdrgarch** to work with the proposed fractional filters, ARFURIMA and ARFURIMA-GARCH models.

1.4 Significance of the Study

The problems in modeling financial and price data are the existence of large noise, trend, persistence and volatility. However, the residuals analysis such as

Portmanteau, Jarque-Bera, and Autoregressive Conditional Heteroscedastic-Lagrange Multiplier (ARCH-LM) tests only expose few pattern in the original series which are not fully captured by the chosen and fitted models. In spite of this, the existing LM filters is not more suitable to study the ILM in nonstationary series. In view of this, it is sufficient to introduce a fractional filter for differencing the nonstationary and ILM series.

The proposed fractional filter shall produces a white noise series with statistical properties similar to stationary and ergodicity process. It produces a white noise series by eliminating huge noise signal as often observed in financial and economic indices. It also overcomes the problems of over and under-differencing as highlighted by Hurvich and Chen (2000) and Nau (2014) and Wei (2006) respectively. Besides, filter shall capture adequately the dependence in financial and economic indices. Therefore, the proposed ILM fractional filter will entirely eliminates linear trend, removes the LM component and provides a filtered or stationary series for estimating AR and MA components. The combined estimations of the AR and MA components by using the fractional filter form the ARFURIMA models. This ARFURIMA models will not be handled adequately capture patterns in the nonstationary and ILM process which may not be handled adequately by other LM models including the ARFIMA and ARTFIMA models. This will leads to the estimation of parsimonious, stable, efficient and stationary models.

It is also known that studying the LM and volatilities of time series by using a hybrid models ensured adequate removal of the noise signals that affect modeling procedures. Also, hybrid models are very significant to handle LM-volatile process with fractional difference value in the interval of $1 < d < 2$. Having said that, the hybrid ARFURIMA-GARCH models were proposed in this study to increase the precision of fitting and diagnostic tests and also to assist in obtaining reliable forecast results.

1.5 Limitation of the Study

This thesis focuses on modeling and forecasting the nonstationary crude oil prices, financial and economic indices with high degree of dependency by developing new filtering methods for fractional filter, ARFURIMA and ARFURIMA-GARCH model. The daily crude oil prices used in this study are the Brent, Dubai, West Texas Intermediate (WTI) and Tapis, each between the 26/01/2004 and 31/12/2018. Also, the financial and economic indices from developing economies namely Greece between 03/10/1988 and 31/12/2018, Nigeria between 17/01/2000 and 31/12/2018, Saudi Arabia between 04/06/2007 and 31/12/2018, Malaysia between 03/10/1988 and 31/12/2018, Kuwait between 01/02/1999 and 31/12/2018, and Uruguay between 01/02/1999 and 31/12/2018 are used. This study uses the first and fractional differencing methods of ARIMA, ARFIMA, ARTFIMA and ARFIMA-GARCH models. These existing methods are used as benchmarks for assessing the performance of the newly developed fractional filters, ARFURIMA and ARFURIMA-GARCH models.

1.6 Daily Financial and Economic Time Series Data

The list of names of the ten financial and economic daily time series data, with its abbreviations and sizes are dispayed in Table 1.1. The first four out of the seven financial data are called Brent, Dubai, WTI and Tapis crude oil prices and they are referred to as BRCOP, DBCOP, WTICOP and TPCOP respectively. The remaining financial data are ATHEX composite index, Nigeria all share index and Saudi Arabia DOM Islamic price index referred to as ACINDEX, NGINDEX and SISINDEX respectively. Other economic indices used in this study are Malaysia 3-months interest rate for deposit, Kuwait dinar to United State dollar (USD) exchange rate and Uruguay peso to United Kingdom pound (UKP) exchange rate and they are named as M3IRD, KWUSDEX and URUKPEX respectively.

Table 1.1: Daily Time Series Used for Analysis

S/No.	Type of Economics and Financial	Abbreviation	Sample Size	Date
1	Brent Crude Oil Prices	BRCOP	3896	26/1/04 - 31/12/18
2	Dubai Crude Oil Prices	DBCOP	3896	26/1/04 - 31/12/18
3	WTI Crude Oil Prices	WTICOP	3896	26/1/04 - 31/12/18
4	Tapis Crude Oil Prices	TPCOP	3896	26/1/04 - 31/12/18
5	ATHEX Composite Index	ACINDEX	7891	3/10/88 - 31/12/18
6	Nigeria all share index	NGINDEX	4946	17/1/00 - 31/12/18
7	Saudi Arabia DOM Islamic price index	SISINDEX	3021	4/6/07 - 31/12/18
8	Malaysia 3 Months Interest Rate for Deposit	M3IRD	7891	3/10/88 - 31/12/18
9	Kuwait Dinar to USD Exchange Rate	KWUSDEX	5196	1/2/99 - 31/12/18
10	Uruguayan Peso to UKP Exchange Rate	URUKPEX	5196	1/2/99 - 31/12/18

Source: Datastream

1.7 Organization of the Thesis

This thesis is organized in the following way. The first part of this chapter is an introduction of LM, LMMMs and hybrid models followed by the problem statement, the research objectives and the signficance of the research. In Chapter 2, a review of related literatures on the LM and FI methods are presented. Chapter 2 also consists of an introductory and literature appraisal for LM diagnostic tools, FI process and types of LM as well as ARIMA and ARFIMA methods. The discussion on the volatility and hybrid models and R packages for LM are also reviewed in chapter 2. Chapter 3 explains the various steps which make up the new proposed models. Chapter 4 describes the models performance and their applications while Chapter 5 consists of the conclusion and recommendation(s) for future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter presents a clear review of LM estimation, modeling and forecasting methods. The reviews cover the application of LM methods in the area of financial and economics (such as stock price, exchange rate, crude oil prices, inflation) and hydrology and climate data. Similarly, studies by using hybrid of mean and volatility methods are also reviewed.

2.2 Long Memory Estimation Methods

Methods for identifying, testing and estimation of LM parameters are classified as heuristic, nonparametric and semiparametric. Hurst (1951), Mandelbrot and Wallis (1968), Teverovsky and Taqqu (1997) introduced the rescaled range statistics (referred to R/S), a heuristic and graph based methods for detecting dependence in time series. This graphical LM detection methods were used in many early studies of LM. In the graphical approach, although slope of regression lines can be used to obtain a rough estimate of the Hurst or memory parameter, H, the graphical approach failed in: providing information about the type of lag association or autocorrelation between time series, also in testing and obtaining the accurate estimates of the LM parameters. To overcome the weaknesses of the heuristic method, nonparametric methods were introduced. Some of the methods include the modified R/S test of Lo (1991), Detrended Fluctuation Analysis (DFA) of Peng *et al.* (1994) and Kolmogrov test of Kulperger and Lockhart (1998). For other nonparametric methods for testing LM, see Phillips and Shimotsu (2004), Abadir *et al.* (2007) and Boutahar *et al.* (2007).

However, the nonparametric methods have shortcomings in estimating the degree of LM and providing additional information about the spectral density of the time series.

Meanwhile, the semiparametric methods for estimating LM parameters are based on the assumption that the spectral density of the time series is nonstationary. The log periodogram regression of Geweke and Porter-Hudak (referred to GPH (1983)), Local Whittle Estimator (LWE) credited to Kunsch (1987) and Robinson (1995), and Hurst Exponent (HE) are some of the semiparametric estimation methods. The Baillie and Kapetanios (2008) and Asai *et al.* (2012) are some of the studies that employed the semiparametric approach for estimating fractional differencing parameters. See Zevallos and Palma (2013) and Busch and Sibbertsen (2018) for other recent LM estimation in the frequency and time domain respectively.

2.3 Application of Long Memory Models

Recent works had used ARFIMA models to study both the historical and return series. Analyzing stock index, exchange rate, crude oil prices and specifically low frequency series, the degree of fractional differencing, d, had been estimated to be in the interval $0 < d < 1$ (see Arouri *et al.,* (2012), Charfeddine and Ajmi (2013) and Ballie *et al.,* (2014)) and some are in the interval of $1 < d < 2$ (see Dalla (2015), and Gil-Alana *et al.,* (2018)).

For example, Erfani and Samimi (2009) studied the daily Tehran Stock Price Index (TSPI) using the modified R/S and HE methods. They found evidence of LM in TSPI by applying Peters (1991) approach for estimating fractional differencing parameter. They estimated the fractional differencing value as equal to 0.4767. The comparison of stationary ARIMA and ARFIMA models in their study revealed that the ARFIMA (2, 0.4767, 18) is the best model to study the daily TSPI. Also, in-sample and long range out-of-sample forecast further supported the choices of the LM models to study the TSPI data.

Onali and Goddard (2011) investigated LM in the returns of stock index of eight European countries and US by using the Rescaled Range Analysis. They found high degree of LM for the stock prices of Italy and Czech Republic contrary to the claim that the market is efficient. However, the remaining six European markets were showed no evidence or a very low LM which was supporting the arguments that stock markets are efficient.

Smith (2012) examines the LM of fifteen developing and developed daily stock markets by using the rolling window variance ratio tests. The study revealed that the degree of fractional differencing values for the UK, Polish, Hungarian and Turkish markets are very small while the Estonian, Maltese and Ukrainian stock markets are high. Similarly, he emphasizes that 2008 financial crisis is the major reason for the LM in the returns of the UK, Portuguese, Hungarian, Slovakian, Croatian and Polish stock markets. However, the same crisis did not have impact in influencing the LM in stock markets of Russia, Romania, Greece, Turkey and Latvia.

Sensoy and Tabak (2013) compares the degree of LM in all stock markets in European Union (EU) member countries by using the Generalized Hurst Exponent (GHE) method. They claimed that the choice of GHE over LM estimation methods to estimate the fractional differencing values of the EU nations was due to its sensitivity and insensitivity to LM and outliers respectively. Their findings show that there are different degrees of LM among the EU countries with a few countries are having efficient stock markets due to the maturity of the market level. The United Kingdom (UK) and France markets are found to be inefficient indicating high arbitrage opportunities compared to young markets of the union. Also, they found the value of HE is greater than 0.5 among several EU nations which indicates the level of LM in the markets.

The paper of Martinez *et al.* (2018) analyzes the existence of memory in weekly EU countries cooperate bonds and stock index using Detrended Fluctuation Analysis (DFA). The results show that the type of LM in bond differ from the type in the stock markets. Caporale *et al.* (2017) study LM in financial time series by using three type of frequency data namely; daily, weekly and monthly. The fractional differencing values was estimated by using the R/S and fractional integration method. High degree of LM and fractional differencing values were discovered and estimated respectively in low frequencies stock markets for both developed and emerging economies. The presence of LM implies a high chance of predicting the stock price values which is in opposite direction with the efficient market hypothesis that market prices are independent. Also, the magnitude of the LM reveals the potential to make abnormal profits in this type of market setting.

Núñez *et al.* (2017) investigate the occurrence of long-range dependence by considering stocks of nineteen prominent world markets. The study was carried out to ascertain if the duration of data, interval of data or the nature of the available LM models for analysis can affect the results. Parametric and nonparametric methods for estimating LM parameters were used in addition to the ARFIMA models. Results indicated that there is LM effect in the South Korea and China index despite the type of models, interval of data collected and the size(duration) of data used. However, the findings also showed that some markets stock index used in the study exhibits short memory behavior.

The study of Ahmad (2013) investigate dependence strength of Kuala Lumpur Composite Index (KLCI) future contract and spot prices by using fractional integration and ARFIMA methods. His results too show the presence of LM in the series. In addition, the findings confirmed that it is possible to use previous prices to forecast both the future price values and dependency observed in the KLCI index. This is in line with Hyndman and Anathasopolos (2013) on obtaining or realizing a similar forecast behavior that emulates the historical series used in a study.

Gil-Alana *et al.* (2018) investigates LM in the daily Baltic stock market indices by using fractional integration methods. They investigate the presence of structural breaks based on the range of data used and associates the two breaks to bull and bear market phases. They found evidence of LM and estimate the fractional differencing values to be greater than one in each of the overall and sub-sample data. Furthermore, the volatility analysis indicates high degree of LM in the bear markets compared to the bull markets. Overall, they conclude that both the historical and returns of the Baltic indices are generated by a LM process. Nguyen and Darne (2018) used fractional integrated volatility models and Vietnam Stock Exchange (VSE) to investigate LM properties. Their results confirmed that the VSE is a process that has LM behavior.

In previous literatures, the stationary models for example ARIMA are used to describe exchange rate records (see Ayekple *et al.*, (2015), Ngan (2016), Yıldıran and Fettahoğlu (2017)). However, when the Autocorrelation Function (ACF) of currency exchange rate exhibits a slow decay, ARFIMA models is the appropriate model in describing the data. Recently, Sivarajasingham and Balamurali (2017) and Omane-Adjepong *et al.* (2018) each study the Sri Lanka Rupee and Ghana Cedi to United State (US) dollar exchange rate respectively. They shows that both the two types of exchange rates are affected by LM behavior and therefore they can be analyzed by using the ARFIMA models.

Hamzaoui and Regaieg (2017) examines the structure of the daily Euro to US dollar forward premium types of exchange rate by explaining the LM behavior. The exchange rate is based on one month, three month, six month, nine month and one year forward premium for the period of seventeen years and are analyzed by using the GPH and LWE methods. The results of the analysis confirmed the evidence of LM and fractional dynamics of the forward premium data. They estimated the ARFIMA models by applying semiparametric and Maximum Likelihood Estimation (MLE) method. Also, they conclude that the ARFIMA model adequately can recall the dynamics of the LM of the forward premium.

Bora and Kumar (2017) study the presence of LM in the Indian American Depository Receipts (ADRs) markets by using the R/S methods. The results reveal that the Indian ADRs markets exhibit LM and nonlinear behavior. They further ascertained that the outcome of their study will help in predicting risk control and markets management activities. As mentioned by Rinke *et al.* (2017), the LM of inflation rates are significant for determining or formulating the type of monetary policy of an economy. Align with this, they study the monthly consumer price index of the G7 countries by testing for spurious LM and estimating the fractional differencing parameters by using the spectral based test of Qu (2011) and modified LWE. Six of the G7 inflation rates indicate spurious LM while U.S inflation shows evidence of true LM. In analyzing the inflation rates, they ignore the spurious LM assumptions but consider the structural breaks assumptions and use the ARFIMA and SEMIFAR models. Finally, they conclude that though the ARFIMA is suitable to capture LM and describe inflation rates, the SEMIFAR models are adequate for studying the behavior of inflation dynamics.

Kurita (2010) carried out the modeling and forecasting of Japan's unemployment rate by using the annual inflation data and ARFIMA models. Applying the method of maximum likelihood, the fractional differencing value was estimated to be -0.19 indicating that the Japan inflation index was generated by intermediate memory process. The ARFIMA model was fitted to the unemployment data and a short term forecast was evaluated by using the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). According to the findings, the 2008 financial crisis as it increased the level of the Japan's unemployment rate, has less impacts on the model's forecasting performance.

Market efficiency hypothesis always hypothesized that it is impossible to predict daily interest rate. However, the availability of liquid assets, withdrawal of huge capital for business and high volume of activities in markets give rise to possible prediction of the intraday rate. In view of this, Monticini and Ravazzolo (2011) study the intraday interest rate in the whole euro area. They test for the presence of LM by using the Lo's RS statistics. The results show that the intraday interest rates are nonstationary and are generated by a persistent process called the LM. Furthermore, they estimate and compare the ARMA and ARFIMA model and results show that ARFIMA(0, d, 0) model adequately fitted the data and statistically outperforms the random walk and ARMA models. In a similar way, Couchman *et al.* (2006) study the LM of three types of real interest rates for sixteen countries using ARFIMA model. The types of interest rate are realized and two ex-ante interest rates. They concluded that most of the countries exhibit LM parameters in the interval of $0 < d < 1$.

Furthermore, Nezhad *et al.* (2016) estimates the fractional differencing values in the time series of Organization of Petroleum Exporting Countries (OPEC) oil prices by using the GPH, R/S and Modified Rescaled Range (MRS) methods. Their results confirmed that for the period of three years, that is between 15/3/2011 to 22/4/2014, the oil prices for the twelve exporting countries exhibit LM with the estimated fractional differencing values less than 0.5. They also state that the discovered LM value is significance for modelling and forecasting the returns of OPEC oil price.

Burnecki and Sikora (2017) introduces the identification and validation procedures for estimating ARFIMA process. They apply the low-variance procedures for estimating the fractional differencing parameters. They collected Universal Mobile Telecommunications System (UMTS) data from an urban area of Wroclaw, Poland. They show that the random part of the historical data can be described by the ARFIMA model. Furthermore, they emphasize the ARFIMA process can be generated by simulating the noise from its empirical distribution function. As conclusion, they recommend to extend the model by using a hybrid of ARFIMA and GARCH model.

Besides trend and seasonal type of variations, recent studies on modeling and forecasting empirically revealed the presence of LM in meteorological data. While Taylor *et al.* (2009) shows the significance of modeling series that exhibits long range dependence with LM models, Knight and Nunes (2018), Tyralis and Papacharalampous (2018) and Gil-Alana *et al.* (2018) each reports evidence of substantial LM in wind speed, streamflow and temperature data respectively. On the other hand, D'Amico *et al.* (2013) propose the using of indexed semi-Markov model in analyzing of wind speed data by including memory index. They consider the wind speed records obtained by Lasten station, Italy and sample the data at interval of ten minutes. Also, they carry out Monte Carlo simulations by using the simple semi-

Markov process. A comparison of the ACF of both the historical and simulation data reveal the influence of LM in the wind speed data. They conclude and recommend that a good model of wind speed should be able to describe the major characteristics of wind such as hushing or breaks, speed, memory and in addition, should be able to handle historical data without distributional assumptions.

Beblo and Schmid (2010) analyze daily and monthly wind speeds of Manchnow, Germany and carry out short term forecasts. They compare simple regression to the ARFIMA model. The long-range modeling method was chosen due to the LM exhibited by the wind speed data records. After model identification, estimation and testing, the ARFIMA model performed excellent in forecasting the wind speed data due to minimum accuracy measures, the root mean square error. Furthermore, they emphasize that the information provided by the forecast results will assist in planning, developing and upgrading the wind farms or factories.

Another LM mean model that is used to study time series data is the ARTFIMA model. For example, Sabzikar *et al.* (2019) analyzed geophysics, finance, turbulence and climate data by each comparing the ARTFIMA and ARFIMA model. They fitted ARTFIMA(0,0.75,0.03,0), ARTFIMA(0,0.3,0.03,0), ARFTIMA(2,1.3,0.1,0) and ARTFIMA(0,1.01,0.01,0) to the geophysical turbulence in water velocity data, adjusted closing price for AMZN stock, vorticity in a turbulent velocity flow and high resolution hydraulic conductivity data respectively. In each case, they concluded that ARFIMA model is misspecified and fitted poorly to all the data.

2.4 Volatility and Hybrid Models

The concept and modelling of volatility has been described and discussed in many studies and the recent includes that of Gyldenløve (2014), RodrÌguez (2017), Martinez *et al.* (2018) and Segnon *et al.* (2018). Moreover, some studies in financial time series such as Cont (2005) and Baillie *et al.* (2014) documented evidence of persistence, heteroscedasticity, volatility clustering and leptokurtosis. The persistence in volatility occurs, when the effects of the volatility shocks decay slowly due to the historical events has a long time effects. It describes the properties of a financial series whose autocorrelations are large and different from zero for many lags (Charfeddine and Guegan, 2012). This persistence in volatility is also called LM. Meanwhile, the occurrence of volatility clustering is when a large changes in stock values are followed by large changes and vice-versa.

Baillie and Morana (2012) estimated the LM values in the S&P500 stock index using the ARFIMA, Fractional Integrated GARCH (FIGARCH) and Adaptive GARCH (A-FIGARCH) methods. Results revealed a fractional differencing values between 0.32 and 0.33. They confirmed that the stock market index and results based on Monte Carlo simulation are both have LM. Maria *et al.* (2014) study relationships between financial crisis and occurrence of LM among European banking indices by using the HE, ARFIMA and FIGARCH models. They estimated large LM values and discovered a strong relationships among the Argentina, Mexico, Russia, Asia and global financial crisis that occurred in 2008 and 2009. The HE for STOXX600 Bank index was greater than 0.5 which implies LM during the Mexican and global financial crisis. Also, the evidence of LM was discovered in the data of MSCI European Bank index and was associated to the financial crisis of Argentina, Russian and Asia. Similarly, the FIGARCH models revealed the presence of LM in volatility of the banks' index.

The assumptions that financial return are non-normal distribution and have fat tail is referred to leptokurtosis. For example, Ding (2011) forecasts conditional volatility of three major markets which are Standard & Poor 500 stock market daily closing price index and MSCI Europe index by using Asymmetric Power ARCH (APARCH) model. He assumed the returns of the three series to be normal distribution, student's t-distribution and skewed student's t-distribution. His finding suggested that returns follows a skewed student's t-distribution due to the minimum errors and large log-likelihood. Al-Najjar (2016) found that the Jordan's Stock Exchange returns is leptokurtic and thus followed non-normal distribution. However, the GARCH specification used as the hybrid models in the work found that the returns is normally distributed.

While some authors studied the LM and volatility in the financial and economics indices independently for example, Goddard and Onali (2012) and Juchelka (2017) and Flores-Muñoz *et al.* (2018), but there are some who looks into the market shocks impact on the LM and volatilities concurrently; see Kang and Yoon (2013), Kasman *et al.* (2009) and Almeida *et al.* (2017). Their studies showed that hybrid models are significance for investigating the interaction between LM and volatility. In view of this, Baillie *et al.* (1996) lament that market shocks possess some degree of control over returns and volatility at the same time.

Furthermore, several studies have concurrently studied the LM and heteroscedasticity. For example, Ishida and Watanabe (2009) use the GPH and Robinson estimators to estimate the degree of LM in the Nikkei 225 future Realized Volatility (RV) data. The results show that the RV was generated by intermediate LM process. In addition, ARFIMA model was fitted and the residuals of the model indicates evidence of heteroscedasticity. Therefore, ARFIMA-GARCH was formed to capture persistence in the volatility of the Nikkei 225 RV in respective of whether the data used are historical, square-root or log-transformed series. Eventually, they conclude that the ARFIMA-GARCH is excellent to capture the heteroscedasticity in the Nikkei 225 RV. Other studies related to LM and heteroscedasticity are Koopman *et al.* (2007) and Fofana *et al.* (2014).

Meanwhile, Zhou and He (2009) recommends the hybrid of ARMA and APARCH model for forecasting S&P 500 stock index when the errors are assumed to be skewed-t distribution. Duppati *et al.* (2016) studied the persistence in Asian stock markets using low-frequency data and ARFIMA-APARCH models. Their finding confirmed the presence of LM in volatility of Asian equity markets based on five minutes intra-day returns.

On the other hand, Arouri *et al.* (2012) studied the dynamic and LM of stock prices of gold, silver, platinum and palladium. They use both parametric and semiparametric methods for testing and estimating LM parameters. Their results show that there are high dependence in the daily returns of the precious metal commodities. Also, the fitted hybrid model indicates the appropriateness of using LM methods to study the type of series considered in the study.

Past studies also had used volatility and hybrid models to describe the movement of the crude oil prices. Danielson (2011) explained that a combination of LM and volatility model can be used to describe both the LM and variability in financial and oil price time series. Also, Kang and Yoon (2013) examines the ability of three different volatility models such as GARCH, FIGARCH and IGARCH as added to the ARFIMA specification for the conditional mean. The LM in both returns and volatility were studied and the results were inconclusive as they failed to indicate the best model combination(s) to describe the petroleum future contract. For Tapis prices evaluation and modeling see Manera *et al.* (2004) and Akron and Ismail (2017).

Ambach and Ambach (2018) introduced periodic ARFIMA-GARCH process to study the LM and volatile behaviors of WTI oil price that cannot be handled by short memory models. The preliminary analyses confirmed evidence of long-range dependence due to large autocorrelation values. Also, they found evidence of high conditional heteroscedasticity presence in the prices. After model identification and estimation the best order of the periodic ARFIMA-GARCH model was selected based on the minimum values of information criteria. They concluded that although the proposed hybrid model perform well in describing the WTI oil price, the model should be extended to capture other components of the volatility. Therefore, they suggested the asymmetric and threshold fractional integrated volatility models to be considered in the future study.

Similarly, Masa and Diaz (2017) models and forecasts the dependence of Exchange-traded Notes (ETNs) by using daily closing prices of equity, commodity and currency in which represents three classification of the ETNs. In the study, they employed a hybrid of both LM mean and fractional integrated volatility methods and ARFIMA-GARCH model. Their results revealed that LM in the ETNs series is significant and further confirmed that the findings oppose the theory of weak-form efficiency hypothesis. Also, they stressed that the intermediate memory that was determined in the ETNs is a signal for potential investors to take appropriate measures to save their investment.

However, Belkhouja *et al.* (2008) extent the ARFIMA to the ARFIMA-GARCH models with a time varying GARCH specification, in examining the pattern of monthly inflation rate of eight European countries. The long-range test and LM estimation revealed that, all the inflation rates for the eight countries have evidence of the LM properties except Canada in which indicated a non-stationary but mean reverting. Furthermore, the analyses, on inflation rates of Denmark, Finland, Italy, Spain, Portugal and Japan revealed that they can be adequately described by the ARFIMA-STVGARCH hybrid model.

Ioremberg *et al.* (2018) evaluate the LM in the quarterly inflation rate for Nigeria using Augmented Dickey Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) known as KPSS (1992) test. However, they modelled the economic indices by using a combination of ARFIMA-GARCH model. According to the analyses, they found that the quarterly inflations exhibit slow decay in ACF. Also, the test result shows that the series are mean reverting which is an evidence of LM. They concludes that the LM variation presents in the Nigeria quarterly inflation as a results of shock could be the reason for the unexpected increase in prices of energy products. However, this will not cause a permanent change in general price level because it was expected to stabilize at its average price level.

Besides that, Baillie *et al.* (2019) tests the claim that RV were generated by LM process. The fractional differencing values were estimated by using the Fully Extended Local Whittle (FELW), LWE and ARFIMA models. The data represents the five minute low-frequency intraday returns series of five spot exchange rates against U.S dollar and S&P 500 stock index consisting of five minutes tick interpolated prices. The RV were analyzed by using the Heterogeneous AR(HAR), ARFIMA(*p*,*d*,0), Extended HAR(EHAR), ARFIMA-HAR, ARFIMA-EHAR, Time Varying Parameter HAR (TVP-HAR), and TVP-EHAR models. The Bayesian Schwarz BIC methods were considered for models comparison. The conclusion was that modelling the LM behavior is significant for describing RV. For other studies on modeling of exchange rate using the ARFIMA and hybrid of conditional mean and volatility models, see Floros (2008), Karemera and Cole (2010), Chortareas *et al.* (2011) and Kumar (2014).

Meanwhile, Kane and Yusof (2013) investigates the presence of LM in the rainfall data of Chui Chak, a station in Peninsular Malaysia. They estimates the LM parameter by using the GPH and ARFIMA(0,*d*,0) methods. The results indicate fractional differencing values is 0.839. Furthermore, a comparison of the hybrid of ARIMA and ARFIMA with GARCH, revealed that the ARFIMA-GARCH outperforms the ARIMA-GARCH models. The conclusion was based on the fact that the degree of the estimated LM indicates the degree of dependence in the rainfall data, which further shows that the time series variability is a mean-reversion. Further, Ambach (2016) examines the autocorrelation and heteroscedasticity of wind speed and recommends the use of ARFIMA-APARCH model for the prediction of both wind speed and wind power.

Massei (2013) studies the daily turbidity records collected from the karst spring used for water supply of the city of Le Havre, Upper Normandy, France. The study considered the short and LM model and fitted the ARIMA and ARFIMA models. The short and LM models were fitted to the data. The serial correlation analysis of the residuals was carried out by using both the Ljung-Box and McLeod-Li tests. The outcome shows evidence of heteroscedasticity. Also, after applying the hybrid model, results show that the ARIMA-GARCH is adequate to study the daily turbidity data. The procedures indicate an excellent short-term turbidity prediction. Other studies that have applied the hybrid ARIMA-GARCH methods, include Mohamadi *et al.* (2017) and Dritsaki (2018).

2.5 Spurious Long Memory

It is a known fact that structural breaks can be disguised as LM in the time series data. Therefore, the importance of testing for structural breaks in the conditional