Spring 2021

Modeling time series with conditional heteroscedastic structure

Ratnayake Mudiyanselage Isuru Panduka Ratnayake

Department: Mathematics and Statistics

MODELING TIME SERIES WITH CONDITIONAL HETEROSCEDASTIC STRUCTURE

by

RATNAYAKE MUDIYANSELAGE ISURU PANDUKA RATNAYAKE

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

MATHEMATICS WITH STATISTICS EMPHASIS

2021

Approved by:

Dr. V.A. Samaranayake, Advisor
Dr. Akim Adekpedjou
Dr. Gayla Olbricht
Dr. Robert Paige
Dr. Gregory Gelles

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following four papers, formatted in the style used by the Missouri University of Science and Technology:

Paper I found on pages 24-52, will be submitted to a refereed archival journal. A subset of the results was published in the proceedings of the Joint Statistical Meetings held in Boulder, CO, in August 2019.

Paper II, found on pages 53-98, will be submitted to a refereed archival journal. A subset of the results was published in the proceedings of the Joint Statistical Meetings held virtually in August 2020.

Paper III, found on pages 99-136, will be submitted to a refereed archival journal.

Paper IV, found on pages 137-180, will be submitted to a refereed archival journal. A subset of the results was published in the proceedings of the Joint Statistical Meetings held in Baltimore, MD, in August 2017.

ABSTRACT

Models with a conditional heteroscedastic variance structure play a vital role in many applications, including modeling financial volatility. In this dissertation several existing formulations, motivated by the Generalized Autoregressive Conditional Heteroscedastic model, are further generalized to provide more effective modeling of price range data well as count data. First, the Conditional Autoregressive Range (CARR) model is generalized by introducing a composite range-based multiplicative component formulation named the Composite CARR model. This formulation enables a more effective modeling of the long and short-term volatility components present in price range data. It treats the long-term volatility as a stochastic component that in itself exhibits conditional volatility. The Generalized Feedback Asymmetric CARR model presented in this dissertation is a generalization of the Feedback Asymmetric CARR model, with lagged cross-conditional range terms added to allow complete feedback across the two equations that model upward and downward price ranges. A regime-switching Threshold Asymmetric CARR model is also proposed. Its formulation captures both asymmetry and non-linearity, which are two main characteristics that exist in the price range data. This model handles asymmetry and non-linearity better than its range-based competitors, based on the Akaike's Information Criteria. In addition to the above models, a Time Varying Zero Inflated Poisson Integer GARCH model is introduced. This model enables the modeling of time series of count data with excess number of zeroes where this excess varies with time. In this model, the zero inflation component is modeled either as a deterministic function of time or as a vector of stochastic variables.

ACKNOWLEDGMENTS

Throughout the duration of my dissertation research and its writing, I have received a great deal of support and assistance from multiple parties. I would like to thank all those who have supported me during this time.

In particular, I would like to thank my supervisor Dr. V.A. Samaranayake for his continued support, encouragement and patience. I would also like to thank my graduate committee for their valuable inputs that helped me to successfully complete my dissertation. I extend my sincere gratitude to all the faculty members and the staff of the department of Mathematics and Statistics, Missouri University of Science and Technology. The opportunities and guidance provided helped me to grow both academically and personally. In addition, I would like to thank my parents and friends for their unwavering support. Last but not least, I would like to thank my wife, Gayani, for her continued support and encouragement. I could not have completed this task without her assistance, tolerance, and enthusiasm.

TABLE OF CONTENTS

Page

PAPER

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

Time series analysis plays a vital role in modeling empirical data, especially in areas of econometrics and finance, but its usefulness also extends to fields such as biology and engineering. There are two primary methodological areas time series analysts focus on, namely the modeling of the underlying data generating process and forecasting future values. The former involves fitting statistical models to temporally observed data for extraction of meaningful information and uncover important features embedded in the time series. Such endeavors may focus on uncovering seasonal components or determining if the underlying process has a unit root in its autoregressive polynomial. The aim is on building explanatory models that provides insight into the underlying data generating process. In forecasting, the primary focus is on obtaining accurate forecasts rather than uncovering insights into the data generating mechanism. One may, however, argue that such insights would lead to better predictive models. The models proposed in this dissertation are an attempt to provide more flexibles models to explain the underlying data generating process, but they can also be used for forecasting future values, as illustrated through real-life examples.

To implement effective time series modeling strategies, practitioners need to pay attention to the nuanced features found in empirical series. For example, the variability associated with a time series can change over time and therefore it is unreasonable to assume homogeneity of variance over the observed time period. This phenomenon, known as heteroscedasticity, can refer to the unconditional variance of a time series or the conditional variance. While variance is one measure of the stochastic variability of the random variables that compose a time series, other measures such as the range is also used. The term conditional heteroscedasticity refers to the changing variability of the time series at a given time point, conditional on past observations. In the area of finance, conditional heteroscedasticity is useful in explaining the clusters of high variability and low variability periods, which are common in stock market data. High variability is associated with periods of high financial uncertainty and low variability reflects a period of less uncertainty, with both these phenomena triggered by recent economic news and market conditions. By contrast, unconditional heteroscedasticity explains the general structural changes in the variability which are not related to events during the prior period.

Volatility is another term use to refer to the uncertainty, or the variability, that exists in the financial market and reflected in time series data. In general, volatility is defined as the degree of dispersion of a random variable over time and it is usually quantified either by standard deviation, variance, or the range of the random variable of interest. Volatility modeling is employed in many areas such as economics and finance, but such models have applications in fields such as engineering and biology. Econometricians define volatility as the risk related to the value of the assets changing over time. The volatility is discussed broadly as historical and implied volatilities. The historical volatility estimates the changes of an asset by measuring the price changes over the predetermined periods while implied volatility is derived from the market price of a market traded derivative.

Traditional time series models assume that the variance is constant, meaning the statistical dispersion remains unchanged across different time periods. However, it is not uncommon for empirical time series to exhibit volatility clusters. In other words, when the volatility is high it is likely to remain high and when it is low and it is likely to remain low.

Clearly, the homogeneity of variance assumption is violated in such situations. Therefore, it is fundamentally inaccurate to fit the statistical models that assume constant variance over time. To address this issue, especially in the case of modeling financial volatility, Engle (1982) introduced the Autoregressive Conditional Heteroscedastic (ARCH), model. The ARCH model explains the future volatility based on current observables. The ARCH process describes the variability as a weighted average of previously estimated squared errors from historical data, and these weights provide more influence to the recent information and less to that of the distant past. Furthermore, an ARCH process can handle a higher number of extreme values than what is expected from a standard normal distribution, hence it is more applicable during high volatility periods.

Tim Bollerslev (1986) expanded on the ARCH model and proposed a class of heteroscedastic models called the Generalized Autoregressive Conditional Heteroscedastic (GARCH) models. Subsequently, the GARCH formulation was rapidly expanded to include what is known as the GARCH family of models. The GARCH model constitutes of three variance components: a constant variance portraying the long-run average, the variance forecast from the previous period, and the variance arising from the new information. The weights of the last two forecasts govern the fluctuations to the long-run average returns due to the arrival of new information and the volatility observed in the immediate past.

The ARCH and the GARCH family models became popular among researchers due to their versatility in modeling financial data. Motivated by the fact that the ARCH and the GARCH models are useful tools to explain the real-world phenomena and successfully forecast future volatiles, researchers added variations to the standard ARCH and GARCH models. A wide variety of ARCH and GARCH type models were proposed, including the Exponential GARCH (EGARCH) by Nelson (1991), Threshold ARCH (TARCH) and Threshold GARCH (TGARCH) by Zakoian (1994), GJR-GARCH by Glosten, Jagannathan and Runkle (1993), Quadratic GARCH (QGARCH) by Sentana (1995), Integrated GARCH (IGARCH) from Engle and Bollerslev (1986), Fractionally Integrated GARCH (FIGARCH) of Baillie, Bollerslev, & Mikkelsen (1996), and the Fractionally Integrated EGARCH (FIEGARCH) by Bollerslev & Mikkelsen (1996). For many other important developments see Engle (2003). These models try to incorporate characteristics such as nonlinearity, asymmetry, and long memory properties of volatility, utilizing a variety of parametric and non-parametric approaches. Another closely related, but econometrically distinct, volatility model called the Autoregressive Conditional Duration or ACD model was proposed by Engle and Russell (1998). Since transactions greater than a certain magnitude can be linked to financial volatility, where more frequent transactions reflect higher volatility, ACD formulations are used as an alternative way to estimate volatility through the modeling of durations between transactions.

The ARCH and the GARCH family models are mainly focused on modeling and forecasting financial volatility and risk based on price returns. In many financial applications, the standard deviation is the most common measure of stock return volatility. Therefore, they can be identified as examples of return-based volatility models. Since the concept of volatility was introduced, researchers have sought alternative measures of financial volatility. One such alternative is the range. Parkinson (1980) argued that volatility measures could be calculated by considering the daily high, daily low, and opening price of a stock in addition to the traditional closing prices. He also compared traditional measures of volatility that were calculated simply by using closing prices, with extreme value methods by taking high and low prices of an asset. He concluded that the range-based method is far superior to the standard methods based on returns. Beckers (1983) tested the validity of different volatility estimators. This paper concludes that the range of a stock price yields far more important and fresh information. Beckers also stated that using the range of a stock price is better than using close-to-close changes. Kunitomo (1992) improved Parkinson's original result and proposed a new range-based estimator, which, according to the author, is ten times more efficient than the standard volatility estimator. In another study, Alizadeh, Brandt, and Diebold (2002) proved that the range-based volatility estimators are highly efficient when compared to the classical volatility proxies that were based on log absolute returns or squared returns and showed that log range is approximately normal. Hence, the range of an asset price for a given period can be used as a more informative proxy variable to measure an asset's volatility for a well-defined period such as a day. Therefore, the range of an asset price for a given period can be used as a more informative proxy variable to measure the volatility of the asset during that period. Researchers studied this alternative approach to volatility modeling and developed new theoretical range-based models with comprehensive empirical examples illustrating their utility.

Chou (2005) introduced a range-based volatility model called the Conditional Autoregressive Range or the CARR model. The CARR model is primarily an ACD type formulation. It is employed to explain the price volatility of an asset by considering range of the log prices for a given fixed time interval while the ACD process is used to model the time intervals between events. The CARR model is quite similar to the standard volatility models such as the GARCH model. One distinct difference between the two models is that the GARCH model uses the rate of return as its volatility measure, and the CARR model uses the range as its volatility measure. The CARR model proposed by Chou is a simple and efficient tool for analyzing the volatility clustering property when compared to the GARCH models. For example, Chou (2005) showed that the effectiveness of volatility estimates produced by the CARR models is higher than the estimates of standard return-based models such as GARCH models. Due to the growing interest in the CARR model, variations of it, such as Exponential CARR, Weibull CARR, CARR-X (Chou 2005), Asymmetric CARR (Chou 2006), Lognormal CARR (Chaing 2016), Gamma CARR (Xie and Wu 2017), and Feedback Asymmetric CARR (Xie 2018) were developed. The range-based models introduced in this dissertation adds to this family of range-based volatility models.

In the next section a review of the ARCH, GARCH, and CARR model is provided.

1.1. THE AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTIC (ARCH) MODEL

The Autoregressive Conditional Heteroscedastic (ARCH) formulation was first proposed by Engle (1982) to model the time dependent variance of a time series. The ARCH process is used to model the conditional variances which is also referred to as conditional volatility, and it is expressed as a linear function of the squared errors. The large squared returns may signal a relatively high volatile period while series of small squared returns may signal a relatively low volatile period.

1.1.1. The ARCH (*p*) Model. Let ε_t be the real valued discrete-time stochastic process, and F_{t-1} denotes the sigma field generated from the information set up to time *t*-1. Then the ARCH model of order *p* is formulated as:

$$
\begin{aligned}
&\varepsilon_t = \sigma_t z_t; \quad z_t \sim i.i.d.\ N(0,1),\\
&\varepsilon_t \mid \mathrm{F}_{t-1} \sim N\left(0,\sigma_t^2\right),\\
&Var\left(\varepsilon_t \mid \mathrm{F}_{t-1}\right) = E\left(\varepsilon_t^2 \mid \mathrm{F}_{t-1}\right) - \left[E\left(\varepsilon_t \mid \mathrm{F}_{t-1}\right)\right]^2 = E\left(\varepsilon_t^2 \mid \mathrm{F}_{t-1}\right) = \sigma_t^2,\\
&\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2,\\
&\alpha_0 > 0, \quad \alpha_i \geq 0, i = 1,\ldots,p.
\end{aligned} \tag{1.1}
$$

When $p=1$, the ARCH (*p*) model can be rewritten as the ARCH (1) process. Sometimes it is important to rearrange the ARCH (1) model as an AR (1) process. To do so, define the serially uncorrelated zero-mean stochastic process $\{\eta_t\}$ such that: $\eta_t = \varepsilon_t^2 - \sigma_t^2$. After replacing σ_t^2 in equation (1.1) by $\varepsilon_t^2 - \eta_t$, the ARCH process of order 1 can also be specified as the AR (1) model: $\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \eta_t$. Here, $E\left[E\left(\varepsilon_t^2 \mid \mathrm{F}_{t-1}\right)\right] = E\left(\sigma_t^2\right) = \dfrac{\alpha_0}{1-\alpha_1} > 0$ and $0 \leq \alpha_1 < 1$. In a separate study done by Ling and McAleer (2002), the condition $0 \leq \alpha_1 < 1$ is shown to be the necessary and sufficient condition for the weak stationarity of an ARCH (1) process.

A main application of the ARCH model is to forecast future conditional variances. Assume σ_{t+1}^2 is the one step ahead conditional variance in the ARCH (1) model. Then one can write:

$$\sigma_{t+1}^2 = E\left(\varepsilon_{t+1}^2 \mid \mathrm{F}_t\right) = \alpha_0 + \alpha_1 \sigma_t^2.$$

1.1.2. Parameter Estimation of ARCH Models. Engle (1982) stated that the parameters can be estimated by the Maximum Likelihood Estimation (MLE) method. Let Θ be the parameter vector such that $\Theta = (\alpha_0, \alpha_1, \ldots, \alpha_p)'$ and p is the order of the ARCH model. Then under the normality assumption, the likelihood function for the ARCH (p) is formulated as follows:

$$f(\varepsilon_1, \ldots, \varepsilon_n \mid \Theta) = f(\varepsilon_1, \ldots, \varepsilon_p \mid \Theta) \times \prod_{t=p+1}^{n} f(\varepsilon_t \mid \varepsilon_1, \ldots, \varepsilon_{t-1} : \Theta).$$

Here, $f(\varepsilon_1, \ldots, \varepsilon_p \mid \Theta)$ is the joint pdf function of $\{\varepsilon_t\}_{t=1}^{p}$, where conditional pdf of ε_t given sigma field generated by all the information set up to time t-1, F_{t-1} is:

$$\begin{aligned} f(\varepsilon_t \mid \mathrm{F}_{t-1}) &= f(\varepsilon_t \mid \varepsilon_1, \ldots, \varepsilon_{t-1} : \Theta), \\ &= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right), \end{aligned}$$

where, σ_t^2 is defined as given in the equation (1.1). As pointed out by Engle (1982), the exact form of the joint pdf of $f(\varepsilon_1, \ldots, \varepsilon_p \mid \Theta)$ is complicated, and therefore it is replaced by a joint distribution obtained by conditioning on the first p observations. The resulting conditional likelihood function of ARCH (p) can be written as:

$$\prod_{t=p+1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right).$$

Finally, the conditional log likelihood function can be presented as:

$$l\left(\Theta \mid \{\varepsilon_t\}_{t=p+1}^{n}\right) = -\sum_{t=p+1}^{n} \left[\frac{1}{2}\log(2\pi) + \frac{1}{2}\log(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2}\right].$$

Engle (1982) also proved that the parameters are asymptotically independent hence, they can be maximized separately using available numerical optimization methods.

Figure 1.1 shows the simulated ARCH (1) time series plot of size n =500, with the parameter vector $\Theta = (0.10, 0.90)'$. The simulated data exhibited volatility clustering, which means larger (smaller) changes in volatility are followed by larger (smaller) changes in volatility and they group together. Since the order of the ARCH process is 1, it has the ability to adjust quickly after a large fluctuation in volatility.

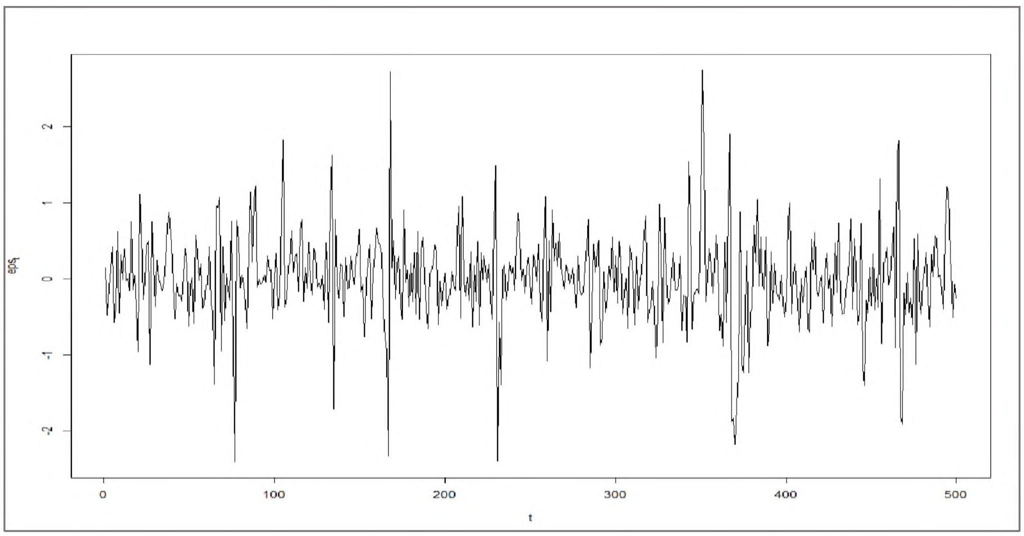

Figure 1.1: Simulated ARCH (1) time series with $\Theta = (0.10, 0.90)'$

Figure 1.2 exhibits the Autocorrelation (ACF) and the Partial ACF (PACF) of the simulated ARCH (1) time series, mentioned above. According to the ACF and the PACF plots, only lag 8 had a mildly significant result. Except for the lag 8 all the other lags in the ACF and the PACF plots, failed to show significant correlations. Therefore, the simulated data are serially uncorrelated for all practical purposes.

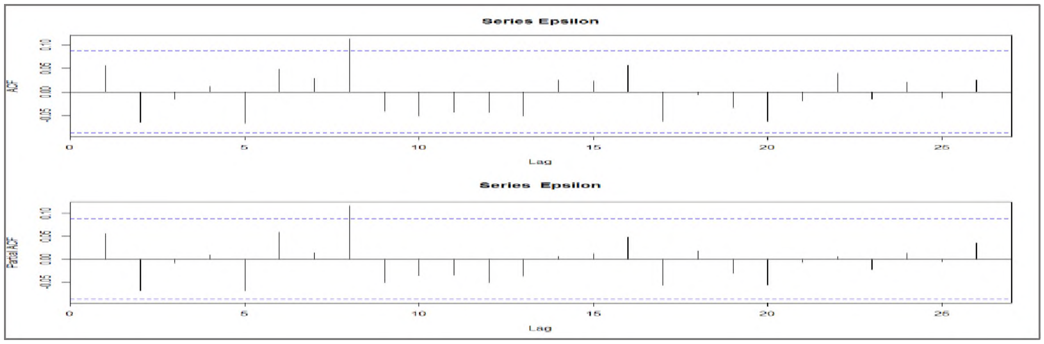

Figure 1.2: Sample ACF and PACF plots of simulated ARCH (1) process data with $\Theta = (0.10, 0.90)'$

Figure 1.3 presents the ACF and PACF plots of the squared values for the simulated ARCH (1) process. The PACF plot shows that lag 1 was significant, hence, it is clear that the ARCH (1) process is appropriate for the squared series.

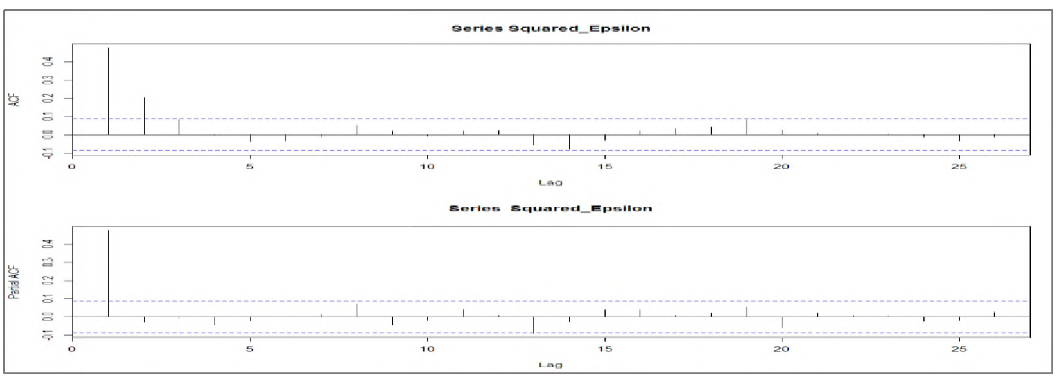

Figure 1.3: Sample ACF and PACF plots of squared simulated ARCH (1) process data with $\Theta = (0.10, 0.90)'$

1.2. THE GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTIC (GARCH) MODEL

Bollerslev (1986) first proposed the Generalized ARCH (GARCH) model, which is a natural generalization of the ARCH model introduced by Engle (1982). As discussed in the previous section, the ARCH processes model the conditional variance returns as a linear function of past lags of actual squared returns. Conversely, the GARCH method is an extension of the ARCH method, which models the conditional variance of the returns as a linear function of the past values of true squared returns as well as the conditional error variances of previous returns.

1.2.1. The GARCH (*p, q*) Model. Let ε_t be the real valued discrete-time stochastic process, and F_{t-1} denotes the sigma field generated from the information set up to time *t*-1. Then the GARCH model of order p and q is expressed as:

$$\begin{aligned}&\varepsilon_t = \sigma_t z_t, \qquad z_t \sim i.i.d.\,N(0,1),\\ &\varepsilon_t \mid \mathrm{F}_{t-1} \sim N\left(0,\sigma_t^2\right),\end{aligned} \tag{1.2}$$

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,\\ &= \alpha_0 + A(L)\varepsilon_t^2 + B(L)\sigma_t,\end{aligned} \tag{1.3}$$

where,

$$\begin{aligned}&p > 0, \quad q \ge 0,\\ &\alpha_0 > 0,\ \alpha_i \ge 0,\ i = 1,\dots,p,\\ &\beta_j \ge 0,\ j = 1,\dots,q,\\ &A(L) = \sum_{i=1}^{p} \alpha_i L^i,\\ &B(L) = \sum_{j=1}^{q} \beta_j L^j.\end{aligned}$$

Here, σ_t^2 is the conditional variance of the ε_t, given all the information, set up to time *t*-1 such that, $Var(\varepsilon_t \mid \mathrm{F}_{t-1}) = E\left(\varepsilon_t^2 \mid \mathrm{F}_{t-1}\right) - \left[E(\varepsilon_t \mid \mathrm{F}_{t-1})\right]^2 = E\left(\varepsilon_t^2 \mid \mathrm{F}_{t-1}\right) = \sigma_t^2$. The Lag operator, L, is defined by $L^k y_t = y_{t-k}$ for all $k \in \mathbb{Z}$. When the order of the GARCH model

q is equal to 0, then the GARCH process reduces to the ARCH process with order p (ARCH (p)).

The following theorem from Bollerslev (1986) provided the necessary and sufficient conditions for weak stationarity of the GARCH (p, q) process.

Theorem 1. The GARCH (p, q) process defines in (1.2) and (1.3) is weakly stationary with $E(\varepsilon_t)=0$, $\operatorname{var}(\varepsilon_t)=\alpha_0\left(1-A(1)-B(1)\right)^{-1}=\alpha_0\left(1-\sum_{i=1}^{p}\alpha_i-\sum_{j=1}^{q}\beta_j\right)^{-1}$ and $\operatorname{cov}(\varepsilon_t,\varepsilon_s)=0$ for $t\neq s$, if and only if $\left(1-\sum_{i=1}^{p}\alpha_i-\sum_{j=1}^{q}\beta_j\right)<1$.

Sometimes, it is advantageous to express the GARCH process as an ARMA time series. Let η_t be a stochastic process defined as:

$$\eta_t=\varepsilon_t^2-\sigma_t^2=\sigma_t^2 z_t^2-\sigma_t^2=\left(z_t^2-1\right)\sigma_t^2,\ \text{with}\ z_t\sim i.i.d.\ N(0,1). \tag{1.4}$$

From equation (1.4) it can be verified and $\{\eta_t\}$ is a serially uncorrelated series with $E(\eta_t)=0$. Substituting the equation (1.4) to the GARCH (p, q) process as defined in (1.2) - (1.3), then the GARCH (p, q) process can be rewritten as:

$$\varepsilon_t^2-\eta_t=\alpha_0+\sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2+\sum_{j=1}^{q}\beta_j\left(\varepsilon_{t-j}^2-\eta_{t-j}\right),$$

$$\varepsilon_t^2=\alpha_0+\sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2+\sum_{j=1}^{q}\beta_j\left(\varepsilon_{t-j}^2-\eta_{t-j}\right)+\eta_t,$$

$$\varepsilon_t^2=\alpha_0+\sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2+\sum_{j=1}^{q}\beta_j\varepsilon_{t-j}^2+\eta_t-\sum_{j=1}^{q}\beta_j\eta_{t-j}. \tag{1.5}$$

Therefore, the GARCH process of order p, q in equation (1.5) can be interpreted as ARMA process in ε_t^2 of orders $m=\max(p,q)$ and q.

1.2.2. The GARCH (1, 1) Model. The GARCH (1, 1) process is the most frequently used GARCH process in empirical studies and it can be derived by setting p=1 and q=1 in the equation (1.3):

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \ \alpha_0 > 0, \ \alpha_1 \geq 0, \ \beta_1 \geq 0. \tag{1.6}$$

Based on Theorem 1, the GARCH (1, 1) process is stationary if $\alpha_1 + \beta_1 < 1$.

1.2.3. Parameter Estimation of the GARCH Model. In this section, the Maximum Likelihood Estimation (MLE) method to estimate the GARCH model parameters is discussed. For the illustrative purposes, we consider the GARCH (1,1) process with normal innovations was considered. Generalization to the GARCH (p, q) process is straightforward.

Let Θ be the parameter vector such that $\Theta = \left(\alpha_0 \quad \alpha_1 \quad \beta_1\right)'$ and $\varepsilon_t \mid \mathrm{F}_{t-1} \sim N\left(0, \sigma_t^2\right)$, where σ_t^2 is defined as in equation (1.6). The conditional pdf and the joint pdf of ε_t given information set up to time t-1 is expressed as:

$$f\left(\varepsilon_t \mid \mathrm{F}_{t-1}\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right),$$
$$f\left(\varepsilon_n, \ldots, \varepsilon_1\right) = f\left(\varepsilon_{n-1}, \ldots, \varepsilon_1\right) f\left(\varepsilon_t \mid \varepsilon_{n-1}, \ldots, \varepsilon_1\right). \tag{1.7}$$

Therefore, from equation (1.7) the conditional log likelihood of the data $l\left(\Theta \mid \left\{\varepsilon_t\right\}_{t=1}^n\right)$ can be derived as follows:

$$l\left(\Theta \mid \left\{\varepsilon_t\right\}_{t=1}^n\right) = -\frac{n}{2}\log\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^n \left[\log\left(\sigma_t^2\right) + \frac{\varepsilon_t^2}{\sigma_t^2}\right]. \tag{1.8}$$

Since there are no closed form solutions for the maximum likelihood estimators for the parameters α_0, α_1 and β_1, they are estimated by maximizing the equation (1.8) numerically. For more derivations of the MLE method see Bollerslev (1986). Note that Nelson and Cao (1992) proved the necessary and sufficient conditions to guarantee the non-negativity of the GARCH (*p, q*) process.

Figure 1.4 shows the simulated GARCH (1, 1) time series plot of size *n*=500, with the parameter vector $\Theta = (0.01, 0.05, 0.90)'$. According to the time series plot, there are periods with high (low) volatility followed by high (low) volatile periods. The ACF and the PACF of the time series data are given in the Figure 1.5. According to the ACF and the PACF plots of the time series, there are no significant correlation lags, hence the time series looks to be uncorrelated.

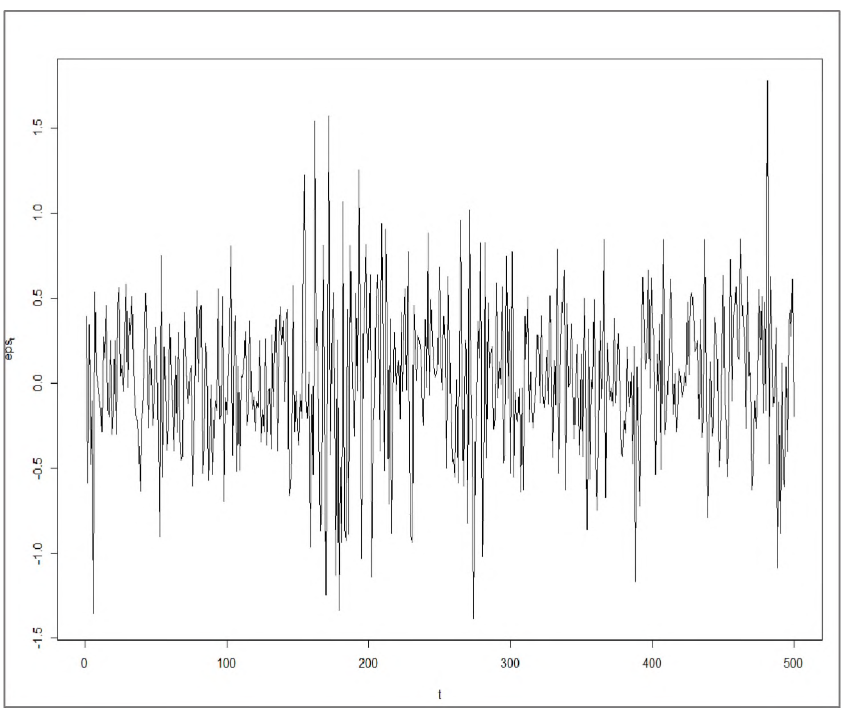

Figure 1.4: Simulated GARCH (1, 1) time series with $\Theta = (0.01, 0.05, 0.90)'$

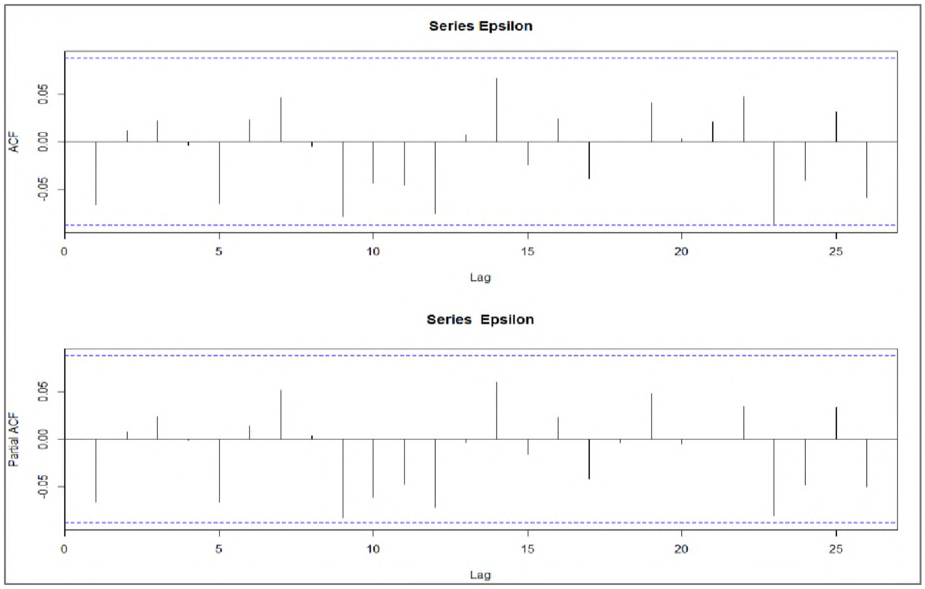

Figure 1.5: Sample ACF and PACF plots of simulated GARCH (1, 1) process data with $\Theta = (0.01, 0.05, 0.90)^{'}$

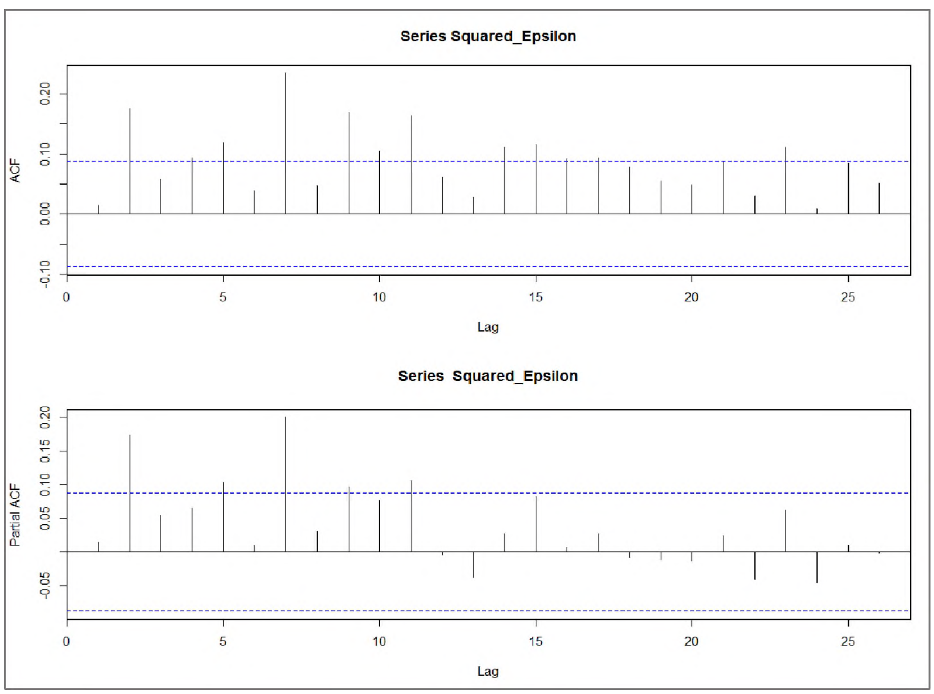

Figure 1.6: Sample ACF and PACF plots of squared simulated GARCH (1, 1) process data with $\Theta = (0.01, 0.05, 0.90)^{'}$

Figure 1.6 shows the sample ACF and PACF of the squared values of simulated GARCH (1, 1) data. These plots indicated existence of significant autocorrelation patterns; hence the squared series is serially correlated.

1.3. CONDITIONAL AUTOREGRESSIVE RANGE (CARR) MODEL

Chou (2005) introduced the CARR which is primarily a range-based model. The CARR model is used to fit the price volatility of an asset by considering range as a measure of price volatility. Let P_s be the logarithmic price of an asset at a time point $s \in (t-1, t]$ and R_t be the price range defined over the fixed time period $(t-1, t]$, and it is formulated as: $R_t = P_s^{high} - P_s^{low} \quad s \in (t-1, t]$.

The CARR model of order (*p*, *q*) is presented as CARR (*p*, *q*) and defined as follows:

$$\begin{aligned}
&R_t = \lambda_t \varepsilon_t, \\
&\mathrm{E}(R_t \mid \mathrm{F}_{t-1}) = \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i R_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \\
&\varepsilon_t \sim f(.), \, i.i.d., \, E(\varepsilon_t) = 1, \\
&\omega > 0, \alpha_i \geq 0, \beta_j \geq 0.
\end{aligned} \tag{1.9}$$

Here, λ_t is the conditional expectation of the price range based on the sigma field F_{t-1} generated by all information set up to time, *t*-1. The non-negative disturbance term, also known as the standardized range, is denoted by ε_t, which is independent and identically distributed with probability density function $f(.)$ with non-negative support and a unit mean. Moreover, Chou (2005), proved that if $0 \leq \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$, holds then the CARR model presented in equation (1.9) is weakly stationary. Since ε_t has a non-negative support, standardized residuals can be explained by using an exponential distribution with unit mean. This results in an exponential CARR, which is abbreviated as the ECARR model.

1.3.1. Properties of the ECARR Model. The ECARR model with order p =1 and q =1 is a widely used CARR model in financial time series and ECARR (1, 1) model is represented as:

$$R_t = \lambda_t \varepsilon_t,$$
$$E(R_t \mid \mathrm{F}_{t-1}) = \lambda_t = \omega + \alpha R_{t-1} + \beta \lambda_{t-1}.$$

The unconditional expectation of the ECARR (1, 1) model can be obtained by:

$$E(R_t) = E(\lambda_t \varepsilon_t) = E\left[E(R_t \mid \mathrm{F}_{t-1})\right] = E(\lambda_t),$$
$$E(\lambda_t) = \omega + \alpha E(R_{t-1}) + \beta E(\lambda_{t-1}),$$

considering the weak stationary assumption, $E(R_t) = E(R_{t-i}) = \mu,$ so that:

$$\mu = E(R_t) = E(\lambda_t) = \frac{\omega}{1-\alpha-\beta}. \text{ Here, } 0 \leq \alpha + \beta < 1.$$

The unconditional variance of the ECARR (1, 1) model can be derived from:

Since $\varepsilon_t \sim \exp(1)$ with $E(\varepsilon_t) = 1 \text{and} E(\varepsilon_t^2) = 2,$ we have:

$$E(R_t^2) = E(\lambda_t \varepsilon_t)^2 = E(\lambda_t^2) E(\varepsilon_t^2) = 2E(\lambda_t^2).$$

Under the weak stationarity assumption:

$$
\begin{aligned}
E\left(\lambda_t^2\right) &= \frac{\mu^2\left[1-(\alpha+\beta)^2\right]}{1-2\alpha^2-\beta^2-2\alpha\beta}, \\
\operatorname{var}\left(R_t\right) &= E\left(R_t^2\right)-\left[E\left(R_t\right)\right]^2, \\
&= 2E\left(\lambda_t^2\right)-\mu^2, \\
&= \frac{2\mu^2\left[1-(\alpha+\beta)^2\right]}{1-2\alpha^2-\beta^2-2\alpha\beta}-\mu^2, \\
&= \frac{\mu^2\left(1-\beta^2-2\alpha\beta\right)}{1-2\alpha^2-\beta^2-2\alpha\beta}.
\end{aligned}
$$

From the above results, to have a finite variance for the ECARR (1, 1) process parameters in the model, it must satisfy the following condition: $2\alpha^2+\beta^2+2\alpha\beta<1$. Let $R_f(1)$ be the one step ahead forecast value R_{t+1} where $R_{t+1}=\lambda_{t+1}\varepsilon_{t+1}$, with $\lambda_{t+1}=\omega+\alpha R_t+\beta\lambda_t$. Then, $R_f(1)=E\left(R_{t+1}\mid \mathrm{F}_{t-1}\right)=\lambda_{t+1}=\omega+\alpha R_t+\beta\lambda_t$.

Some properties of the ECARR model can be inferred from the ARMA process. Therefore, it is important to know how to formulate the CARR model as an ARMA process. Define stochastic difference random variable η_t such that:

$$
\eta_t = R_t - E\left(R_t \mid \mathrm{F}_{t-1}\right) = R_t - \lambda_t. \tag{1.10}
$$

Here, $E\left(\eta_t\right)=E\left(R_t-\lambda_t\right)=0$, and $\operatorname{cov}\left(\eta_t,\eta_s\right)=E\left(\eta_t\eta_s\right)-E\left(\eta_t\right)E\left(\eta_s\right)=E\left(\eta_t\eta_s\right)=0$.

From equation (1.10), the ECARR (*p*, *q*) model was rewritten as:

$$
R_t=\omega+\sum_{j=1}^{g}\left(\alpha_j+\beta_j\right)R_{t-j}+\eta_t-\sum_{j=1}^{q}\beta_j\eta_{t-j},
$$

where, $g=\max\left(p,q\right)$ and $\alpha_j=0: j>p$, and $\beta_j=0: j>q$. Therefore, equation (1.9) is in the form of an ARMA process with order *g* and *q*.

1.3.2. Parameter Estimation of the ECARR Model. In this section, the use of the Maximum Likelihood Estimation technique to estimate the parameters of the ECARR (*p*, *q*) model is discussed. Let Θ be the parameter vector such that $\Theta = \left(\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q\right)'$ and $g = \max\left(p, q\right)$. The likelihood function of the data is:

$$f\left(R_{1,\ldots,}R_n \mid \Theta\right) = f\left(R_{1,\ldots,}R_g \mid \Theta\right) \times \prod_{i=g+1}^{n} f\left(R_i \mid R_{1,\ldots,}R_{i-1} : \Theta\right).$$

Here $\left\{R_{1,\ldots,}R_n\right\}$ the past realization of the range series data and conditional pdf of R_t given the sigma field F_{t-1} generated from all the information set up to time *t*-1 denoted by:

$$f\left(R_t \mid \mathrm{F}_{t-1}, \Theta\right) = \frac{1}{\lambda_t} \exp\left(-\frac{R_t}{\lambda_t}\right).$$

Then the conditional log likelihood function of the data is formulated by:

$$l\left(\Theta \mid R_{1,\ldots,}R_n\right) = -\sum_{t=g+1}^{n} \left[\log\left(\lambda_t\right) + \frac{R_t}{\lambda_t}\right].$$

Since there are no closed form equations for the parameter estimates, an optimization algorithm can be used to estimate the model parameters in Θ. Chou (2005) mentioned that usual asymptotic theories of maximum likelihood estimates hold for the ECARR (*p*, *q*) process when $\{R_t\}_{t=1}^{n}$ is weakly stationary.

Figure 1.7 shows the simulated range data from the ECARR model with order (1, 1). The height of the spikes indicates the volatility, and higher heights imply higher volatility periods. The simulated data clearly show volatility clustering. Figure 1.8 shows the ACF of the simulated ECARR model. The ACFs are highly significant. Therefore, the serial dependence of the data is verified.

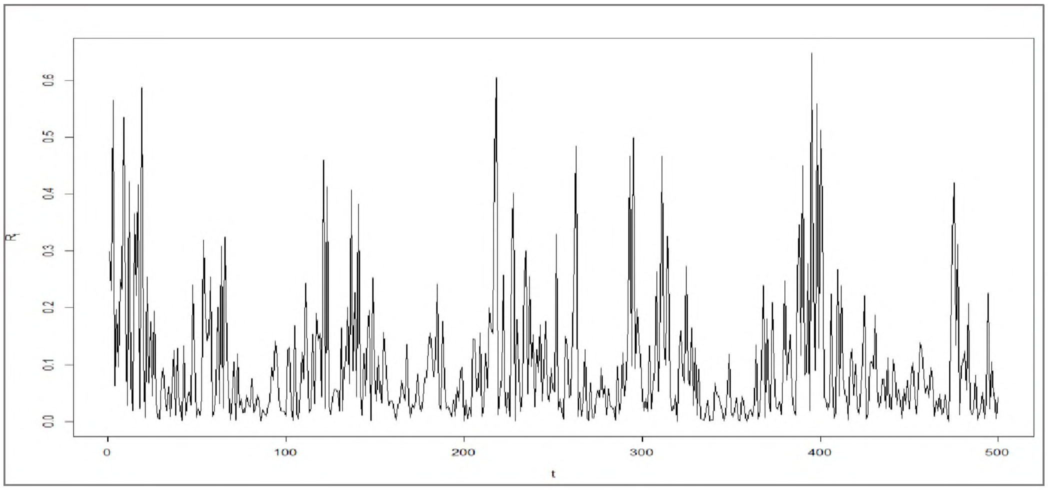

Figure 1.7: A simulated ECARR (1, 1) series with $\Theta = (0.01, 0.20, 0.70)'$

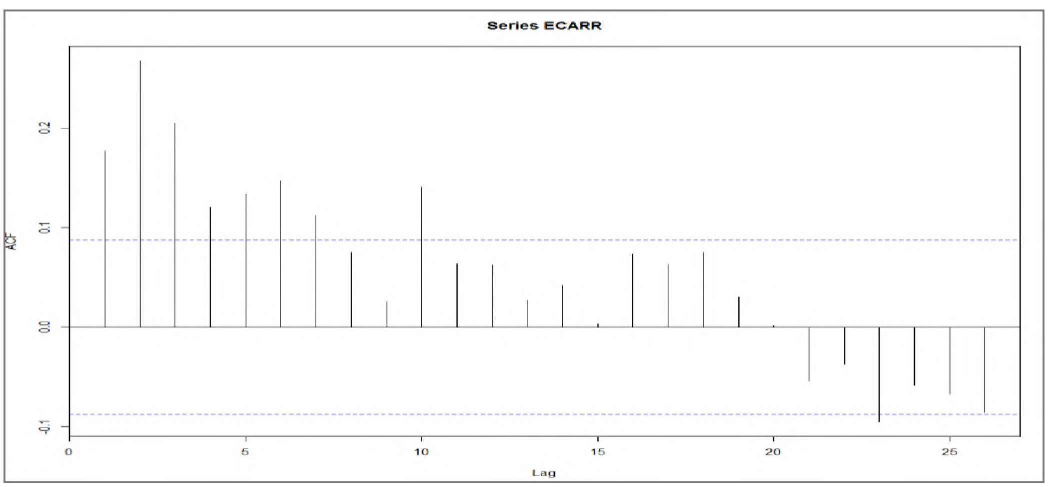

Figure 1.8: The ACF function of simulated ECARR (1, 1) series with $\Theta = (0.01, 0.20, 0.70)'$

The results presented in this dissertation extends the CARR and the Feedback Asymmetric CARR (FACARR) models in three papers. In addition, another paper presents a generalized zero-inflated Poisson model whose formulation parallels the GARCH model. The basic structure of this model for count data is based on what is known as the Autoregressive Conditional Poisson (ACP) model introduced by Heinen (2001, 2003). The ACP model is connected to the ACD models mentioned earlier in the same way a Poisson process is related to a recurrent event process where the durations between events are exponentially distributed. Unlike the regular homogenous Poisson processes and recurrent event processes with independent durations between events, the ACP and the ACD allow serial correlation. Note that the ACP model is the same as the Integer GARCH (INGARCH) process independently proposed by Ferland, Latour, and Oraichi (2006) if the underlying distribution of the count process is Poisson.

1.4. DISSERTATION RESEARCH

This dissertation contains of four papers. Three of these are generalizations of the CARR type models and one is derived from the Autoregressive Conditional Poisson (ACP) model which is related to the ACD model.

Paper I: In this paper a new class of composite range-based component model for volatility to analyze long-term and short-term volatilities in daily price range data is developed. The proposed Composite CARR (CCARR) is a multiplicative component model similar to the Spline-GARCH model of Engle and Rangel (2008). The long-term volatility is modeled using a stochastic volatility component, which itself exhibits conditional volatility. The long-term and the short-term components in the CCARR model are driven by the past realizations of the range model. The application of the proposed model is illustrated by using S&P 500 and FTSE 100 stock indices.

Paper II: This paper generalizes the Feedback Asymmetric CARR (FACARR) model of Xie (2018) and introduce the Generalized FACARR (GFACARR) model. The FACARR model limits the cross feedback to past ranges and does not include past conditional means. The proposed Generalized Feedback Asymmetric Conditional Autoregressive Range Model (GFACARR) removes this limitation and allows the upward range model to include past upward and past downward ranges, along with their respective conditional means. A similar model was defined for modeling a downward range as well. The proposed model is more aligned with the multivariate CARR model. The use of the GFACARR model is illustrated by its application to several price series, including S&P 500, CAC 40, and NIKKEI 225 stock indices.

Paper III: In this paper a Threshold Asymmetric Conditional Autoregressive Range (TACARR) model for the price ranges of financial assets is proposed. The disturbance term of the range process is assumed to follow a threshold distribution with positive support. The study assumes that the conditional expected range process switches between two market regimes. The two market regimes are namely the upward market and the downward market. This model addresses several inefficiencies found in previous price range models including Conditional Autoregressive Range (CARR), Asymmetric CARR (ACARR), Feedback ACARR (FACARR), and Threshold Autoregressive Range (TARR) models. The performance of the TACARR model is assessed using IBM index data. Empirical results show that the proposed TACARR model is useful in-sample prediction and out of sample forecasting of volatility.

Paper IV: This paper introduced a time varying zero-inflated Poisson process to model time series from count data with serial dependence. The model assumed that the intensity of the underlying Poisson process evolves according to a generalized conditional heteroskedastic (GARCH) type model. The proposed model is a generalization of the Zero Inflated Poisson Integer GARCH (ZIP-INGARCH) model proposed by Fukang Zhu in 2012, which, in return, was considered a generalization of the Integer GARCH (INGARCH) model proposed by Ferland, Latour, and Oraichi in 2006. The proposed model is built on these previous formulations and it incorporate the flexibility for the zero-inflation parameter to vary over time, according to a deterministic function or be driven by exogenous variables. Two applications based on the real-world data are discussed and the proposed time varying ZIP-INGARCH (TVZIP-INGARCH) model fitted better with the data compared to Zhu's constant ZIP-INGARCH model.

PAPER

I. MODELING AND FORECASTING FINANCIAL VOLATILITY USING COMPOSITE CARR MODELS

ABSTRACT

The literature showed that forecasting realized volatility and the Conditional Autoregressive Range (CARR) models that utilize the daily range of a commodity price, they outperform the traditional the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) approach that models the daily returns. The CARR models, however, assume that the unconditional mean range is constant over time, which hold only if the unconditional volatility remains fixed throughout the study period. As several researchers reported, there is strong empirical evidence suggesting the feasibility of modeling a slow-varying change in the unconditional volatility over the study period using a long-term volatility component. This paper proposed a new composite range-based component model to analyze both long-term and short-term volatility components in daily price range data. The proposed CCARR models long-term volatility changes as a stochastic component which itself exhibits conditional volatility and the application of the proposed model was illustrated by using the S&P 500 and the FTSE 100 stock indices.

Key Words: CARR Models, Range Estimators, Financial Time Series, Market Volatility, Duration Models

1. INTRODUCTION

Financial volatility is a measure of the dispersion of returns for a given asset. It is the conventional measure in assessing the risk of speculative assets. In general, the riskiness of the market is directly proportional to the volatility. Volatility is closely linked with the stability of the financial market and plays a vital role in determining the level of economic activity. It is also a key input for asset pricing. Thus, financial volatility is an essential factor that policy makers and regulators should consider before any form of financial decision making. Moreover, modeling volatility is crucial in understanding the nature of the dynamics of the finical market.

Modeling the financial volatility of asset prices was discussed extensively in the financial and econometric literature. One of the most successful volatility models used by researchers to model time series volatilities is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model introduced by Bollerslev (1986). This paper was based on the ideas put forth in the seminal paper by Engle (1982), which proposed the Autoregressive Conditional Heteroscedasticity (ARCH) model to address the complexities of time-varying volatility and volatility clustering in the financial time series. The ARCH approach models the error variance as a function of actual errors of the previous periods, while the GARCH method, which is an extension of the ARCH method, models the variance of the error as a function of error terms and its conditional variance.

Owing to the significance of modeling and forecasting asset price volatilities, a wide range of empirical and theoretical investigations were completed within the context of econometric literature to select the ideal model. Akgiray (1989), mentioned that GARCH (1, 1) models fit the daily return series data reasonably well after considering the evidence from the time series behavior of stock prices. The GARCH model used in this paper employed rate of return to study the volatility and found that daily return series demonstrated a significant level of second-order dependence, which cannot be modeled using merely a linear white noise process.

Due to the growing interests and developments in financial time series during the 1990s, some researchers became invested in modeling the time intervals between events. The first durational model was proposed by Engle and Russell (1998). In their publication, they introduced a new statistical model that is capable of analyzing irregularly spaced financial transaction data and they named the model the Autoregressive Conditional Durational (ACD) Model. Since then, multiple authors have proposed related versions of ACD models such as the Logarithmic ACD (LACD) model by Bauwens and Giot (2000), Nonlinear ACD by Zhang, Russell and Tsay (2001), and Box-Cox ACD by Hautsch (2002).

In many financial time series applications, the standard deviation is the most common measure of stock return volatility because it not only calculates the dispersion of returns, but it also summarizes the probability of seeing extreme values in returns. Researchers have also focused their attention on finding alternative measures of financial volatility, such as range. It is well-known in statistics that the range is a measure of the variability of a random variable. Parkinson (1980) argued that volatility measures could be calculated by considering the daily high, daily low, and opening price of a stock, in addition to the traditional closing prices. Parkinson also compared traditional measures of volatility that are calculated by using the closing prices, with extreme value methods by taking high and low prices of an asset. The study concluded that a range-based method is far superior to the available standard methods. Beckers (1983) tested the validity of different volatility estimators and states that the range of a stock price has more important and fresh information. It is also claimed that using the range of a stock price is better than using the close-to-close changes. Hence the range of an asset price for a given period can be used as a more informative proxy variable to measure the volatility of the asset during that period. Researchers who studied this alternative approach to volatility modeling developed new theoretical range-based models with comprehensive empirical examples. For example, Brandt and Jones (2006) fitted effective Exponential GARCH (EGARCH) models to range data from the S&P 500 index.

Chou (2005) first introduced the Conditional Auto Regressive Range (CARR) model, which is primarily an ACD model. While the ACD model is used to model the time intervals between events with positive observations, the CARR is employed to model the price volatility of an asset by considering the range of the log prices for a given fixed time interval. The CARR model is similar to the standard volatility models such as the GARCH model. However, one distinct difference between the two models is that the GARCH model uses the rate of return as its volatility measure, and the CARR model uses the range as its volatility measure. The CARR model proposed by Chou is a simple but efficient tool for analyzing the volatility clustering property, when compared to the GARCH models. This was illustrated empirically via out-of-sample forecasting for S&P 500 data. Chou showed that the effectiveness of volatility estimates produced by the CARR models was higher than the estimates of standard return-based models, such as GARCH models. Zou (2014) used the CARR model and the GARCH model to forecast the volatility of the stock index in the Shanghai stock market. This paper used the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE), which were also used by Chou (2005), to compare the Weibull-CARR model with the GARCH-t models. Zou concluded that the Weibull CARR model outperformed the ARCH-t model in forecasting ability. Quiros (2011) discussed volatility forecasting with range models. He improved previous work done by Chou (2005) by extending the time-period studies and analyzing the performance of the CARR model in contrasting situations such as in periods with upward trends and in periods with downward trends. He proposed various range estimators to analyze the forecasting performance, and he further stated that Parkinson's (1980) model is preferable to the CARR model during periods with upward trends, while the CARR is recommended for periods with downward trends.

Chaing (2016) proposed the Lognormal Logarithmic Conditional Auto Regressive (Lognormal Log CARR) model intending to examine the volatility outliers and improve the accuracy of forecasting. This model was influenced by the Logarithmic Autoregressive Conditional Duration (Log ACD) model of Bauwens and Giot (2000). One major advantage to using either a Log ACD or a Log CARR model is that these models relax positivity restrictions on the parameters of the conditional expectation function. Fernandes, Mota, and Rocha (2005), proposed the multivariate extension of the CARR model, derived the conditions for the existence of statistical properties, such as the first moment, stationarity of the model.

The broad scope of volatility models proposed by various researchers provide copious opportunities to model volatility as a single component. Recent studies carried out on the subject, led to examination of the economic and financial variables' volatility as functions of long and short-term components.

Engle and Lee (1999) introduced an additive component to GARCH models with long-term and short-term components. The Spline GARCH model proposed by Engle and Rangel (2008) models equity market volatilities as a combination of macroeconomic activities and time series dynamics. In this same paper, Engle and Rangel named the slow-moving trend in the volatility process as low-frequency volatility and presented the functional form of the low-frequency volatility by adopting a non-parametric approach. In essence, they considered the low-frequency component was deterministic. Instead of using an additive component, a multiplicative component was used in the Spline-GARCH model to separate low and high-frequency volatilities. Therefore the 'high-frequency return volatility' is a product of a slow-moving deterministic volatility component that can be represented by an exponential spline combined with a unit GARCH model. This model was able to capture short and long-term behaviors of financial market volatilities. The slow-moving volatility component can be used to model the long-run dynamic behavior of the market while the unit GARCH model can be employed to capture short-term dynamics. Based on the Spline GARCH model Engle, Ghysels, and Sohn (2013) proposed a new component model with a direct link to the economic activities and this new class of models was named as the GARCH MIDAS. This paper explained the long-term volatility using an approach that can handle stock volatilities and economic activities recorded in different frequencies, namely: daily monthly, or quarterly. The mixed data sampling (MIDAS) technique was initially introduced by Ghysels (2006) and is used to build a link between the long-run volatility component and macroeconomic variables. The unit GARCH process was used as in the Spline GARCH approach to model the short-run volatility component. The GARCH MIDAS model is a multiplicative model with differentiated short and long-run components of volatility. The conditional volatility of returns in this model depends on macroeconomic variables and previous economic periods or lags. Engle et al. (2013) formulated long-term movement with inflation and industrial production growth. They found that including macroeconomic variables in the model outperforms the traditional time series in terms of long and short horizon forecasting. With the motivation of the GARCH MIDAS model, Swanson (2017) proposed the CARR MIDAS model. In this study, volatility was decomposed into short and long-term components, and the short-run volatility component was explained by an exponential CARR (1, 1) model. The long-run volatility component is computed by aggregating measures of scaled realized range over past k low-frequency periods.

Several other authors also utilized empirical data to illustrate the modeling of short and long- term volatility components using the Spline GARCH and the GARCH MIDAS models. Nguyen and Walther (2017) conducted an empirical study using commodity futures, which are traded in the New York Stock Exchange (NYMEX). They fitted both the Spline GARCH and the GARCH MIDAS models. They found that disentangling high and low volatility components produced better results for in-sample fit in both models.

More recent provided a basic insight into different types of volatility models including the range-based volatility models and discussed the importance of analyzing the long-term and the short-term volatility components in them. While the CARR model assumes a constant unconditional mean range over time, several other studies namely Engle et al. (2013) and Conrad, Christian, Custovic, Anessa, Ghysels, & Eric (2018) suggested with empirical evidence, that unconditional volatility in return series changed over the study periods.

In the paper, a new class of composite range-based component model for volatility to analyze long-term and short-term volatilities in daily price range data was proposed. We introduced a stochastic component to model the long-term volatility in daily price range data, which in itself exhibits conditional volatility. The long-term and the short-term components are driven by the past realization of range price series.

The remainder of the paper is organized as follows: Section 2, presents the proposed Composite CARR (CCARR) model. Thereafter, an estimator to estimate the unobserved long-term volatility component and discuss the parameter estimation procedure is introduced. Further, the Maximum Likelihood Estimation (MLE) procedure to estimate the model parameters is discussed in Section 3, which is followed by the simulation study is presented in Section 4. In Section 5, the proposed CCARR model is applied to multiple indices such as S&P 500 and FTSE 100 are used for the empirical study and compare the prediction and forecasting ability of the CCARR model against the single component CARR model. Section 6 concludes the paper.

2. THE MODEL SPECIFICATION AND DISCUSSION

2.1. THE CONDITIONAL AUTOREGRESSIVE RANGE (CARR) MODEL

Chou (2005) proposed the CARR model which is primarily a range-based model. The CARR model is employed to fit the price volatility of an asset by considering range as a measure of price volatility. A CARR model of order (p, q) is presented as CARR (p, q) and defined as follows:

$$R_t = \lambda_t \varepsilon_t,$$
$$\mathrm{E}(R_t \mid \mathrm{F}_{t-1}) = \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i R_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j},$$
$$\varepsilon_t \sim i.i.d. f(\phi, \varepsilon_t).$$

Here, λ_t is the conditional expectation of the range, based on all information up to time *t*-1. The non-negative disturbance term, also known as the standardized range, is defined by $\varepsilon_t = \frac{R_t}{\lambda_t}$, which is independent and identically distributed with probability density function $f(.)$ with a unit mean. Since R_t and λ_t are positive, the coefficients of the conditional mean range equation have the following restrictions:

$$\omega > 0, \alpha_i \geq 0, \beta_j \geq 0, \text{ for all } i = 1,2,3,...,p \text{ and } j = 1,2,3,...,q.$$

Let R_t be the price range defined over the time interval $\left[t_{open}, t_{close}\right]$ such that:

$R_t = \max(P_\tau) - \min(P_\tau)$, where $\tau \in \left[t_{open}, t_{close}\right]$. Here, we let P_τ be the price of an asset at a given time τ.

2.2. THE COMPOSITE CONDITIONAL AUTOREGRESSIVE RANGE (CCARR) MODEL

Let $P_{j,i,t}$ be the logarithmic price of a speculative asset defined at time *j* of a given short-term period (i.e., day) *t* of any arbitrary long-term period *t* such as month, quarter and year. Here $j,i,t \in [i_{open,t}, i_{close,t}]$ and $i = 1,2,3,...,N_t$ where N_t is the number of days for the given long-term period t. Here $t = 1,2,3,...,T$ where *T* be the number of long-term periods in total time span. The observed price range over the short-term time period *i* at a given long-term period t is denoted as $R_{i,t}$, and it is defined as follows.

$$R_{i,t} = \left[\max(P_{j,i,t}) - \min(P_{j,i,t})\right] \times 100. \tag{2.1}$$

The Composite Conditional Autoregressive Range (CCARR) model for the range is defined as follows:

$$R_{i,t} = \tau_t g_{i,t} \varepsilon_{i,t}, \tag{2.2}$$

where, $\varepsilon_{i,t} \square\ i.i.d.\ f(\sigma, \varepsilon_{i,t})$ with a unit mean (i.e. $\mathrm{E}(\varepsilon_{i,t}) = 1$), $\forall i = 1,2,3,...,N_t$ and $t = 1,2,3,...,T$. Observe that the daily price range ($= R_{i,t}$) is separated into short-term and long-term volatility components.

The long-term volatility component τ_t is given by,

$$\begin{aligned} \tau_t &= \omega_t \eta_t, \\ \omega_t &= \gamma_0 + \gamma_1 \tau_{t-1} + \delta \omega_{t-1}, \\ \omega_t &= \mathrm{E}\left(\tau_t \mid \mathrm{F}_{(t-1)}\right). \end{aligned} \tag{2.3}$$

Here, ω_t is the mean of the long-term volatility component conditioned on all information up to time t-1, and $\mathrm{F}_{(t-1)}$ is the sigma field generated by the information setup to long-term period t-1. The long-term disturbance term is denoted by η_t, where $\eta_t \sim i.i.d.\ f(\nu, \eta_t)$ and $\mathrm{E}(\eta_t) = 1$. The long-term volatility component τ_t is modeled as a stochastic component that, itself, exhibits conditional volatility according to the Conditional Autoregressive Range (CARR (1, 1)) process.

The short-term volatility component $g_{i,t}$ is given by,

$$g_{i,t} = (1-\alpha-\beta) + \alpha R^{*}_{i-1,t} + \beta g^{*}_{i-1,t},$$
$$R^{*}_{i-1,t} = \begin{cases} \dfrac{R_{N_{t-1},t-1}}{\tau_{t-1}} : i = 1 \\ \dfrac{R_{i-1,t}}{\tau_t} : i > 1 \end{cases},$$
$$g^{*}_{i-1,t} = \begin{cases} g_{N_{t-1},t-1} : i = 1 \\ g_{i-1,t} : i > 1 \end{cases}. \tag{2.4}$$

Here, the short-term volatility component $g_{i,t}$ is defined as obeying a unit CARR (1, 1) model similar to Engle and Rangel (2013). Following the derivation given by Engle et al. (2013) for the short-term volatility component, we can prove that unconditional expectation of short-term volatility component is $\mathrm{E}(g_{i,t}) = 1$. Both the short-term and the long-term volatility components are driven by the past realization of the range series.

3. ESTIMATIONS OF CCARR MODEL

3.1. ESTIMATING OF THE LONG-TERM VOLATILITY COMPONENT

Observe that the range observed on short-term period i and long-term period t is given by,

$$R_{i,t} = \tau_t g_{i,t} \varepsilon_{i,t},$$
$$\overline{R}_{.t} = \frac{\sum_{i=1}^{N_t} R_{i,t}}{N_t} = \frac{\sum_{i=1}^{N_t} \tau_t g_{i,t} \varepsilon_{i,t}}{N_t} = \tau_t \frac{\sum_{i=1}^{N_t} g_{i,t} \varepsilon_{i,t}}{N_t}.$$

Since $\dfrac{\sum_{i=1}^{N_t}\left\{E\left[E\left(g_{i,t}.\varepsilon_{i,t} \mid \mathrm{F}_{i-1,t}\right)\right]\right\}}{N_t} = 1,$ it concludes that:

$$\overline{R}_{.t} \approx \tau_t . \qquad (3.1)$$

Therefore, the long-term unobserved volatility component can be estimated using the mean range for the given fixed long-term period.

3.2. PARAMETER ESTIMATION OF THE CARR MODEL

In this section, the log likelihood function for the proposed CCARR model is derived. For this derivation, it is assumed that the model disturbance term $\varepsilon_{i,t}$ is independent and identically distributed as a lognormal distribution with mean $-\frac{\sigma^2}{2}$ and variance σ^2 such that:

$$\varepsilon_{i,t} \sim i.i.d.\ LN\left(-\frac{\sigma^2}{2}, \sigma^2\right).$$

The reason for this assumption is empirical evidence gathered from the two data sets, which are analyzed in Section 5. However, other distributions may also be utilized. Under the assumption of the lognormal distribution, $\mathrm{E}(\varepsilon_{i,t}) = 1$ and $\mathrm{var}(\varepsilon_{i,t}) = \exp(\sigma^2) - 1$. The long-term disturbance term η_t is assumed to be independently and identically distributed as a lognormal distribution with mean $-\frac{\nu^2}{2}$ and variance ν^2 such that:

$$\eta_t \sim i.i.d.\ LN\left(-\frac{\nu^2}{2}, \nu^2\right).$$

Hence, the $\mathrm{E}(\eta_t) = 1$ and variance $\mathrm{var}(\eta_t) = \exp(\nu^2) - 1$. Further we assume that $\varepsilon_{i,t}$ and η_t are independent.

We consider equation (2.2), (2.3) and (2.4) to obtain the following results:

$$
\begin{aligned}
R_{i,t} &= \tau_t g_{i,t} \varepsilon_{i,t}, \\
\ln(R_{i,t}) &= \ln(\tau_t g_{i,t} \varepsilon_{i,t}), \\
\ln(R_{i,t}) &= \ln(\tau_t) + \ln(g_{i,t}) + \ln(\varepsilon_{i,t}), \\
\ln(R_{i,t}) &= \ln(\omega_t) + \ln(\eta_t) + \ln(g_{i,t}) + \ln(\varepsilon_{i,t}).
\end{aligned}
$$

Since $\varepsilon_{i,t}$ and η_t are lognormal distributions, $\ln(\varepsilon_{i,t})$ and $\ln(\eta_t)$ are normal distributions, and $\ln(\varepsilon_{i,t}) + \ln(\eta_t)$ is normally distributed with mean $-\frac{\theta}{2}$ and variance θ where $\theta = \sigma^2 + \nu^2$. Then the conditional distribution of $R_{i,t}$ given $\mathrm{F}_{i-1,t}$ is expressed as:

$$
f(R_{i,t} \mid \mathrm{F}_{i-1,t}, \Phi) = \frac{1}{\sqrt{2\pi\theta} R_{i,t}} \exp\left[\frac{\left(\ln(R_{i,t}) - \ln(\omega_t) - \ln(g_{i,t}) + \frac{\theta}{2} \right)^2}{-2\theta} \right].
$$

Here $\Phi = (\alpha, \beta, \gamma_0, \gamma_1, \delta, \sigma^2, \nu^2)$ is the parameter vector. Thus, the conditional log likelihood function can be derived as follows:

$$
\begin{aligned}
L(\Phi \mid R) &= \prod_{\forall i,t}^{T} f\left(R_{i,t} \mid \mathrm{F}_{i-1,t}\right), \\
l(\Phi \mid R) &= \ln\left[L(\Phi \mid R)\right] = \sum_{\forall i,t}^{T} \ln\left[f\left(R_{i,t} \mid \mathrm{F}_{i-1,t}\right)\right], \\
l(\Phi \mid R) &= -\frac{1}{2}\left(\sum_{\forall i,t}^{T} \left\{ \ln(2\pi) + \ln(\sigma^2) + 2\ln(R_{i,t}) + \left[\frac{\left(\ln(R_{i,t}) - \ln(\omega_t) - \ln(g_{i,t}) + \frac{\theta}{2} \right)^2}{\theta} \right] \right\} \right).
\end{aligned}
\tag{3.2}
$$

Here,

$$\omega_t = \gamma_0 + \gamma_1 \tau^*_{t-1} + \delta\omega_{t-1},$$
$$g_{i,t} = (1-\alpha-\beta) + \alpha R^*_{i-1,t} + \beta g^*_{i-1,t}.$$
$$R^*_{i-1,t} = \begin{cases} \dfrac{R_{N_{t-1},t-1}}{\tau^*_{t-1}} : i = 1 \\ \dfrac{R_{i-1,t}}{\tau^*_t} : i > 1 \end{cases}, \quad g^*_{i-1,t} = \begin{cases} g_{N_{t-1},t-1} : i = 1 \\ g_{i-1,t} : i > 1 \end{cases} \text{ and } \tau^*_1 = \overline{R}_1 .$$

In this paper, the Maximum Likelihood Estimation (MLE) was employed to obtain the model parameters for the proposed CCARR model. To utilize the MLE method, initial parameter values must be obtained. Determination of these initial values are discussed in the following sub-section.

3.3. INITIAL VALUE ESTIMATION

First, we needed to find an estimator for the unobserved long-term volatility component ($= \tau_t$). The unobserved long-term volatility component was estimated by using monthly mean value of daily price ranges as derived in Section (3.1). Therefore, the long-term volatility component was estimated by $\overline{R}_t$($\approx \tau_t$). Then the long-term volatility component was modeled by using CARR (1, 1) process as given follows:

$$\omega_t = \gamma_0 + \gamma_1 \overline{R}_{t-1} + \delta\omega_{t-1}, \text{ where, } \omega_1 = \frac{\gamma_0}{(1-\gamma_1-\delta)}.$$

After fitting a CARR (1, 1), we found the initial values for parameters γ_0, γ_1, δ, and σ^2. Next we needed to find the initial values for the model parameters in the short-term volatility component model $g_{i,t}$. Let $R^*_{i,t}$ be the daily adjusted price range which is defined as follows:

$$R^{*}_{i-1,t} = \begin{cases} \dfrac{R_{N_{t-1},t-1}}{\tau_{t-1}} : i = 1 \\ \dfrac{R_{i-1,t}}{\tau_t} : i > 1 \end{cases},$$

$$R^{*}_{i,t} = g_{i,t}\varepsilon_{i,t}.$$

Where, $g_{i,t}$ is given by,

$$g_{i,t} = (1-\alpha-\beta) + \alpha R^{*}_{i-1,t} + \beta g^{*}_{i-1,t}.$$

$$g^{*}_{i-1,t} = \begin{cases} g_{N_{t-1},t-1} & : i = 1 \\ g_{i-1,t} & : i > 1 \end{cases}.$$

We next fitted a unit CARR (1, 1) model to the adjusted daily price range and found the initial parameter values for α , β and ν^2

4. SIMULATION STUDY

The finite sample performance of estimators was investigated using a simulation study. We used ‘nloptr’, which is a nonlinear optimization function of R software to generate the relevant data. The length of the long-term time series was set to n = 360 and n =720, and the length of the short-term period was set to c =22, which represented the one business month. Therefore, the length of the time series m $(= cn)$ and s= 500 simulation runs were completed for each parameter sample size combination. The simulation study consisted of two parts. First, the price range data for the proposed CCARR model was generated as given in the equation (2.2), (2.3), and (2.4). Then, maximized the profile likelihood function (3.2) using the constrained nonlinear optimization function ‘nloptr’ in R. The Mean Absolute Deviation Error (MADE) was utilized as the evaluation criterion.

The MADE is defined as, $\frac{1}{s}\sum_{j=1}^{s}\left|\hat{\phi}_j - \phi_j\right|$ where s is the number of replications. Simulation results are reported in Table 1.

Table 1: Means of MLE method estimates and MADE in parenthesis

Model	n	α	β	γ_0	γ_1	δ	θ
True Parameters		0.30	0.60	0.10	0.20	0.70	0.60
M1	360	0.2994 (0.0103)	0.5988 (0.0130)	0.1121 (0.0270)	0.2016 (0.0271)	0.6850 (0.0443)	0.5995 (0.0077)
	720	0.2995 (0.0070)	0.5998 (0.0089)	0.1048 (0.0157)	0.19994 (0.0192)	0.6955 (0.0277)	0.5998 (0.0055)
True Parameters		0.30	0.50	0.02	0.20	0.40	1.00
M2	360	0.2995 (0.0126)	0.4982 (0.0180)	0.0208 (0.0040)	0.2006 (0.3021)	0.3829 (0.0920)	0.9991 (0.0128)
	720	0.2994 (0.0085)	0.4996 (0.0127)	0.0205 (0.0027)	0.1993 (0.0223)	0.2913 (0.0648)	0.9997 (0.0091)
True Parameters		0.20	0.70	0.10	0.30	0.60	0.25
M3	360	0.1996 (0.0080)	0.6983 (0.0134)	0.1166 (0.0312)	0.3017 (0.0349)	0.5810 (0.0532)	0.2498 (0.0032)
	720	0.1996 (0.0055)	0.6997 (0.0094)	0.1075 (0.0188)	0.2990 (0.0252)	0.5934 (0.0351)	0.2499 (0.0023)

According to the simulation results, the MLE method can be used to estimate the parameters with higher accuracy. The accuracy of the estimates was increased when the length of the long-term volatility period was increased.

5. AN EMPIRICAL ANALYSIS

5.1. THE DATA SETS

In this study, two stock indices, namely: the Standard and Poor's 500 (S&P 500) index of United States and the Financial Times Stock Exchange 100 (FTSE 100) index on the London Stock Exchange were used. The sample periods for both S&P 500 and FTSE 100 start on January 4, 1990 and ended on December 31, 2018. Daily values for the opening price, closing price, high price, low price and adjusted price were reported over the span of the study period. The data set was downloaded from the Yahoo Finance from the web site (https://finance.yahoo.com/) using the 'quantmod' package in R software. The data set was divided in to two samples where one sample spanned from January 4, 1990 to December 29, 2017 and was used for the model parameter estimation and in-sample predictions. The out-of-sample predictions were done by using the sample from January 1, 2018 to December 31, 2018. The same sample separation procedure was used for both stock indices.

Table 1 presents the summary statistics of the daily price range series for the S&P 500 and the FTSE 100 indices. The daily price range $\left(=R_{i,t}\right)$ of a given day i on a month t was obtained as given in Equation (2.1).

The high values for Kurtosis indicated a strong deviation from the normal distribution. Both price ranges had large positive skewness and it is suggested that a positively skewed density should be used to model disturbance term. The Jarque-Bera test statistics fell far from zero and had extremely low p-values (<0.0001) leading to a rejection

Table 2: Summary statistics for daily S&P 500 and FTSE 100, January 04, 1990 – December 29, 2017 (p-value)

Summary Statistics	S&P 500	FTSE 100
Mean	1.2524	1.2926
Maximum	10.9041	10.7532
Minimum	0.1456	0.0762
Standard Deviation	0.9185	0.9042
Skewness	3.2012	2.8668
Kurtosis	18.7175	14.9450
Jarque-Bera	115093(<0.0001)	75987(<0.0001)
Ljung-Box Q-22	41938(<0.0001)	42047(<0.0001)

of the null hypotheses that the data is normally distributed. The Ljung-Box test null hypothesis was that the time series data are independently distrusted. In this study, time lags of 22 trading days, which was the approximate number of trading dates for a month, was used for the test. After 22 lags of sample autocorrelations were examined, the large test statistic values and very small p-values (<0.0001) conclude that the data exhibited a strong persistence in daily price range data. Time series plots for the daily price range data of S&P 500 and FTSE 100 over the in-sample period are given in Figure 1 and Figure 2.

Both graphs exhibit the same behavior over the period of study. Height of the spikes is an indication of price volatility and if the spikes were high during a certain period, then that period was considered to be highly volatile.

5.2. ESTIMATION OF CARR MODEL

Initially, a single component CARR model to daily price range data was fitted to explain price volatility over the study period. We assumed the disturbance term ε_t in CARR $(p=1, q=1)$ model specified in the Equation (2.1) follows the Exponential (ECARR), Weibull (WCARR) and the Lognormal (LNCARR) distributions.

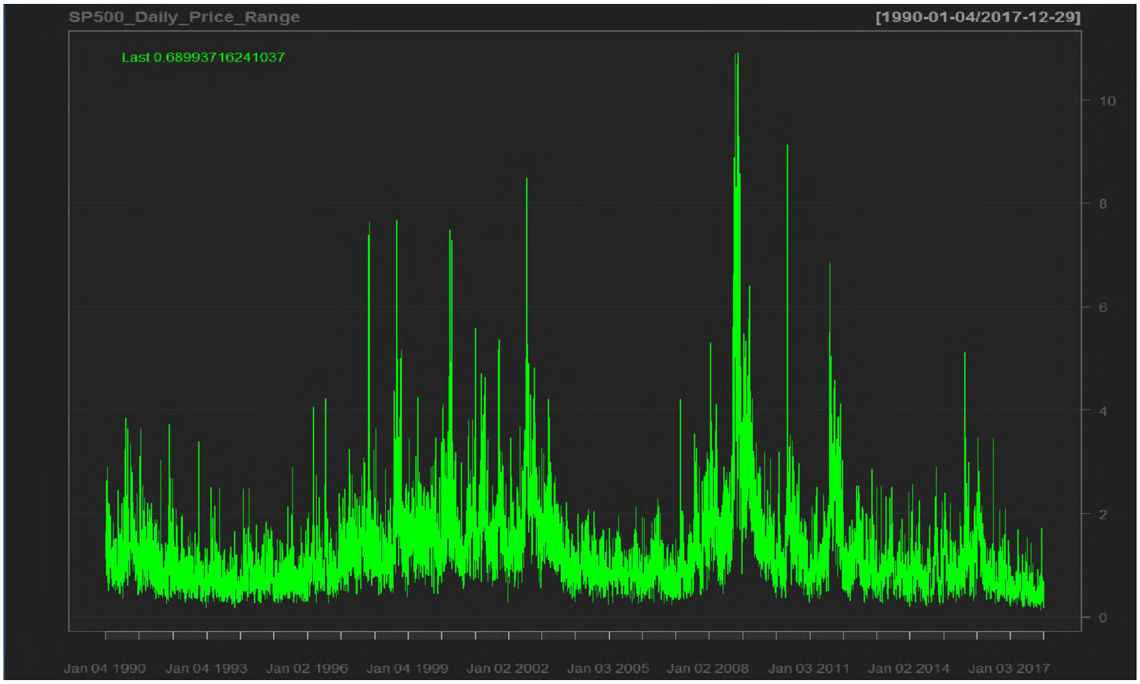

Figure 1: S&P 500 daily price range from 01/04/1990 to 12/29/2017

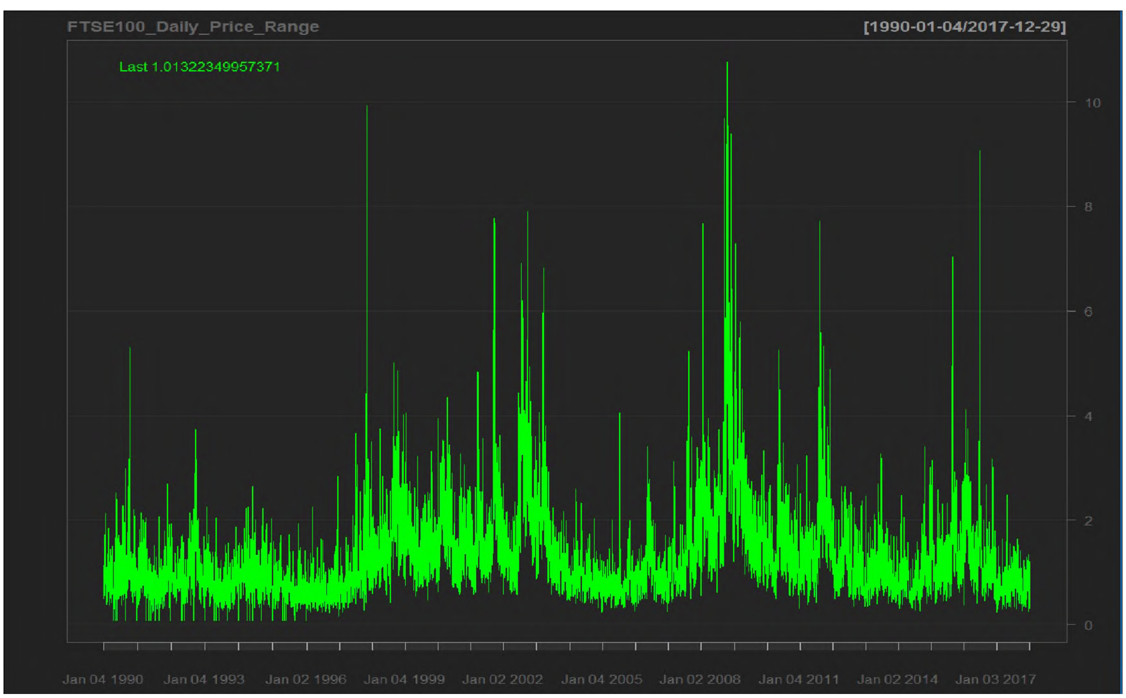

Figure 2: FTSE 100 daily price range from 01/04/1990 to 12/29/2017

Table 3: Estimation of CARR (1, 1) model using daily S&P 500 index data (p-value)

	ECARR (1,1)	WCARR (1,1)	LNCARR (1,1)
γ_0	0.0193 (<0.0001)	0.0286 (<0.0001)	0.0149 (<0.0001)
γ_1	0.1679 (<0.0001)	0.1780 (<0.0001)	0.1653 (<0.0001)
δ	0.8163 (<0.0001)	0.7979 (<0.0001)	0.8228 (<0.0001)
AIC	15870.63	9868.95	**8507.40**

Table 4: Estimation of CARR (1, 1) model using daily FTSE 100 index data (p-value)

	ECARR (1,1)	WCARR (1,1)	LNCARR (1,1)
γ_0	0.0212 (0.0060)	0.0334 (<0.0001)	0.0140 (<0.0001)
γ_1	0.1715 (<0.0001)	0.1953 (<0.0001)	0.1677 (<0.0001)
δ	0.8116 (<0.0001)	0.7772 (<0.0001)	0.8223 (<0.0001)
AIC	16507.90	10050.41	**9424.89**

Since, daily price range data had large positive skewness, positively skewed distributions like the Exponential, Weibull or the lognormal should be used to model the data. According to the AIC values given in Table 3 and Table 4, the LNCARR (1, 1) had a lower AIC value, hence it fitted the daily price range data better for both stock indices.

5.3. ESTIMATION OF CCARR MODEL

The proposed CCARR process models daily price volatility by using short-term and long-term volatility components. In this study, day was considered as a short-term time period, while the month was used as the long-term period of interest. Initially we need to find an estimator for the unobserved long-term volatility component, τ_t, and it was estimated by using the monthly mean as previously derived in Section (3.1). Figures 3 and

4 present the comparison of daily price range data and monthly mean as a long-term volatility component for each of the indices. According to the Figures 3 and 4 monthly mean closely followed the long-term changes in price volatility and it did a quite good job capturing the periods with high volatility.

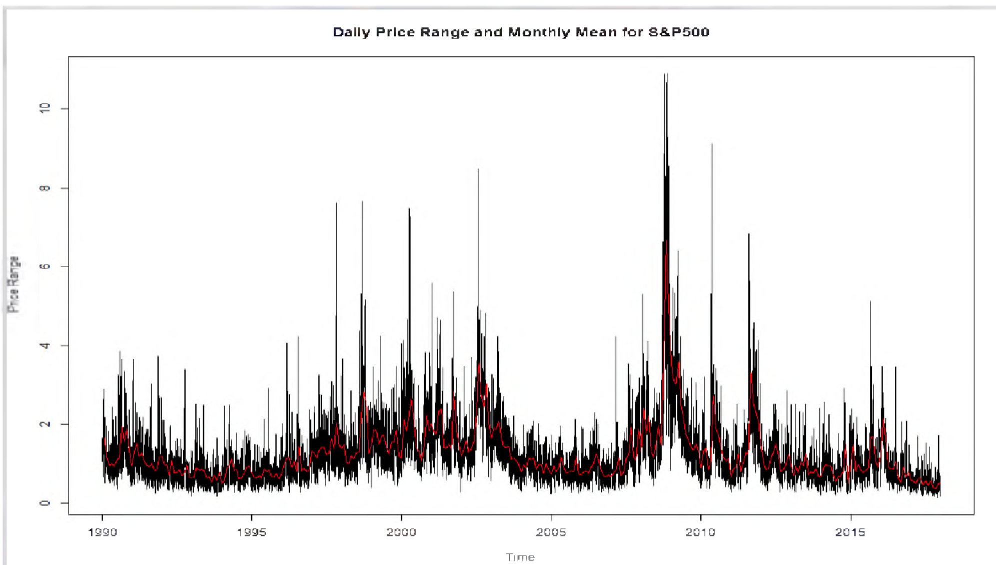

Figure 3: Daily price ranges (black) and monthly observed mean for S&P 500 (red) from 01/04/1990 to 12/29/2017

Based on the method describe the Section 3.3 we estimated the initial values for both indices. Initial values for the S&P 500 parameters were (0.20, 0.63, 0.27, 0.20, 0.70, 0.25) and that of the FTSE 100 were (0.13, 0.64, 0.25, 0.20, 0.68, 0.26). After determining the initial values for the model parameters, the conditional log likelihood function (3.2) was maximized by using ‘nloptr’ package which is a nonlinear optimization algorithm in R. Table 5 presents the MLE results for the CCARR model.

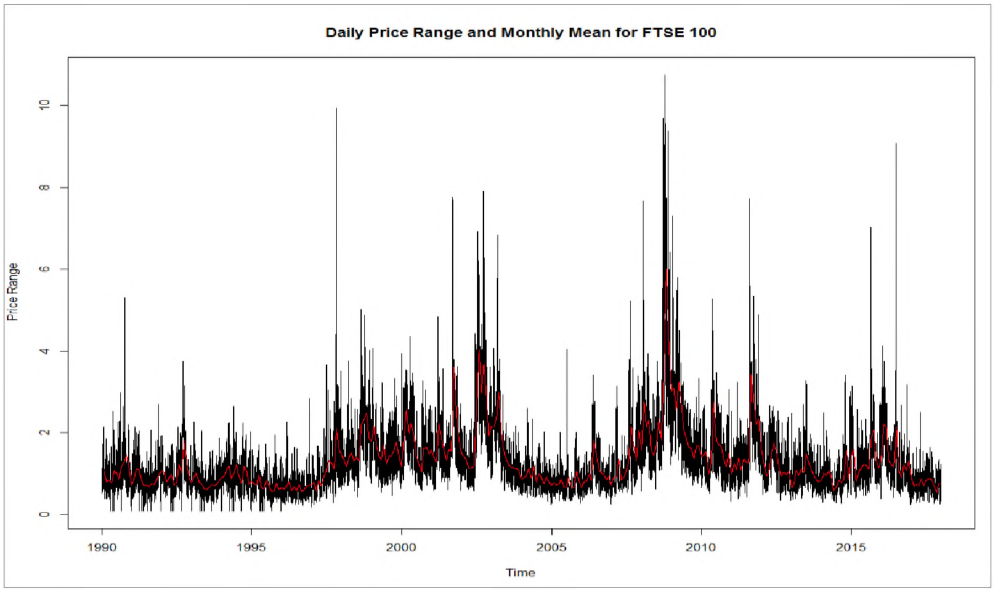

Figure 4: Daily price ranges (black) and monthly observed mean for FTSE 100 (red) from 01/04/1990 to 12/29/2017

Table 5: Estimation of the CCARR model using daily S&P 500 index data and FTSE 100 index data (p-value)

	S&P 500 Estimated Coefficients	FTSE 100 Estimated Coefficients
γ_0	0.0353	0.0359
γ_1	0.1595	0.2214
δ	0.8053	0.7536
α	0.1753	0.1855
β	0.7758	0.7566
θ	0.1820	0.1900
Ljung-Box Q-22	33.048 (0.0619)	24.442 (0.3245)

The Ljung-Box Q test was used to assess whether the residual series were independently distributed. Large values for Ljung-Box Q-22 test for the price range indicated that there was a significant persistence in the volatility. However, the residual series for the fitted CCARR model demonstrated a significant reduction in the Ljung-Box

Q-22 statistics with p-values exceeding 0.05, suggesting the absence of serial autocorrelation up to 22 trading days.

5.4. COMPARISON BETWEEN LNCARR MODEL AND CCARR MODEL

In this section, we tested the in-sample prediction and out-of-sample forecasting ability of the proposed CCARR model. To test the differences in prediction and forecasting power between the CCARR and the LNCARR, we conducted in-sample prediction and out-of-sample forecasting. To test how well the proposed CCARR models performed in extreme situations, such as a recession period, we conducted the analysis for the period from December 2007 to June 2009 for S&P 500 and April 2008 to June 2009 for FTSE 100.

The in-sample prediction for the LNCARR was its conditional mean range and that of the CCARR is the product of estimated long-term and short-term volatility components. To compare the in-sample prediction and out-of-sample forecasting ability, we calculated AIC and Mean Absolute Error (MAE) statistics for the CCARR and the LNCARR. The MAE is calculated as follows:

$$\text{MAE} = \left(\frac{\sum_{\forall i,t} \left(\left| MV_{i,t} - PV_{i,t} \right| \right)}{N} \right).$$

The unobserved real volatility is represented by $MV_{i,t}$, and here we use price range $(R_{i,t})$ as proxy variable for real volatility. Predicted values $PV_{i,t}$ are the fitted values for price range $\left(R_{i,t}\right)$.

Table 6 presents the model comparison between the LNCARR (1, 1) model and the CCARR model. The CCARR model showed a better performance over the LNCARR model for all periods. During the full period of in-sample and the time of recession the CCARR models had smaller MAE values and lower AIC values, when compared to the LNCARR models for both stock indices. Furthermore, diagnostic test results for the residuals indicated that they were independently and identically distributed in the CCARR model. However, there was a clear rejection of the null hypothesis in the LNCARR model where errors showed high persistence in the residual of price range data. We calculated the one step ahead out-of-sample forecasted values for both stock indices. Based on the out-of-sample statistics, the CCARR model dominated over the LNCARR model with respect to the MAE statistics.

Table 6: Model comparison between LNCARR (1, 1) and CCARR for S&P 500 and FTSE 100 (p-value)

Sample Period		S&P 500		FTSE 100	
		LNCARR (1,1)	CCARR	LNCARR (1,1)	CCARR
In-Sample	MAE	0.422	0.420	0.415	0.410
	Standardized Residuals Q (22)	51.724 (0.0003)	33.048 (0.0612)	50.084 (0.0050)	24.442 (0.3245)
	AIC	8507.395	8466.001	9424.89	9375.982
Recession	MAE	0.88	0.87	0.93	0.92
Out-of-sample	MAE	0.47	0.46	0.34	0.33

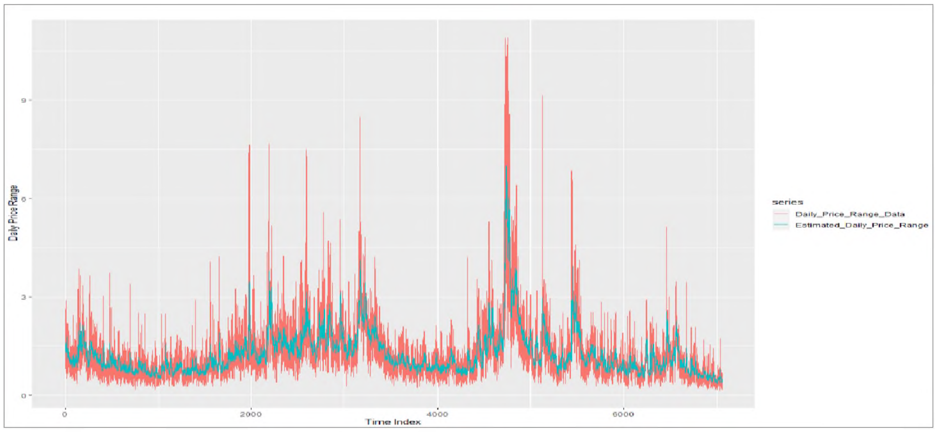

Figure 5: In-sample prediction by CCARR model for S&P 500

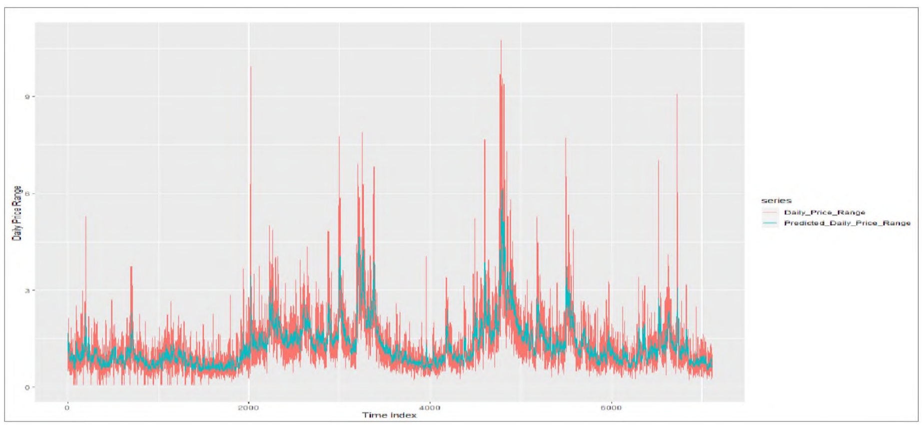

Figure 6: In-sample prediction by CCARR model for FTSE 100

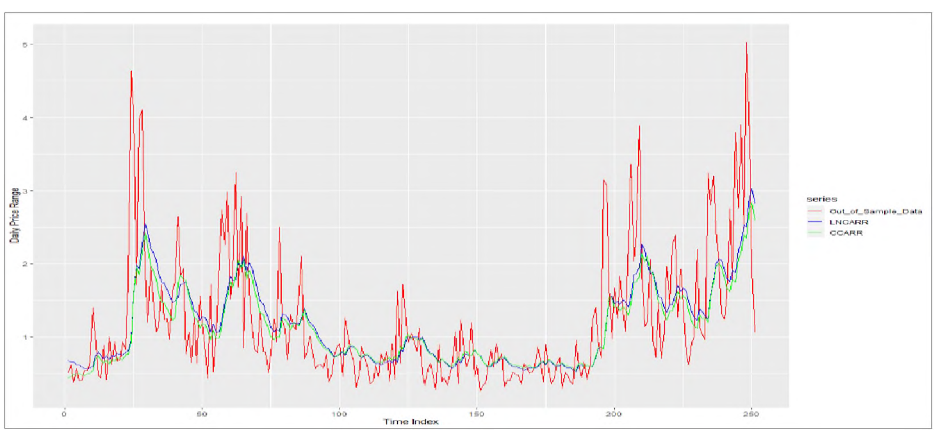

Figure 7: 1-step ahead forecasted value comparison between LNCARR (1, 1) and CCARR for S&P 500

Figure 8: 1-step ahead forecasted value comparison between LNCARR (1, 1) and CCARR for FTSE 100

Figure 7 and Figure 8 show how well the proposed model performed in 1-step ahead prediction. It can be seen in the figures that the CCARR model picked high volatility periods (high spikes), as the LNCARR did however, the CCARR quickly captured the low volatile periods (short spikes) while the LNCARR did not have the flexibly to adapt to such situations.

6. CONCLUSIONS

In this study, we proposed a composite range-based model to estimate the long-term and the short-term volatility components. The proposed methodology modeled the long-term volatility by using a stochastic process, which exhibited conditional volatility. Furthermore, both the short-term and the long-term volatility components are driven by the past realization of price range data. The empirical results based on the MAE and the AIC values showed that the CCARR model dominated the LNCARR model (which was selected based on performance out of other CARR models) in performance, especially during the recession periods. The proposed CCARR model also did better than the single component LNCARR model with respect to the residual diagnostics.

REFERENCES

Akgiray, V. (1989). Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts. The Journal of Business, 62(1), 55-80.

Bauwens, L., and Giot, P. (2000). The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks. Annales D' Économie ET De Statistique, (60), 117-149.

Beckers, Stan (1983). Variance of Security Price Return Based on High, Low and Closing Prices. Journal of Business 56, 97–112.

Ben Taieb, S., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. Expert Systems with Applications, 39(8), 7067-7083.

Bollerslev, Tim (1986). Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics 31, 307–327.

Brandt, M.W. and C.S. Jones (2006). "Volatility Forecasting with Range-Based EGARCH Models". Journal of Business and Economic Statistics, 24, 470-486.

Chiang, M., Chou, R. Y., & Wang, L (2016). Outlier detection in the lognormal logarithmic conditional autoregressive range model. Oxford Bulletin of Economics and Statistics, 78(1), 126-144.

Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model. Journal of Money, Credit and Banking, Vol. 37, pp. 561-582.

Conrad, Christian, Custovic, Anessa and Ghysels, Eric, (2018), Long- and Short-Term Cryptocurrency Volatility Components: A GARCH-MIDAS Analysis, Journal of Risk and Financial Management, 11, issue 2, p. 1-12.

Engle, Robert (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation. Econometrica 50, 987–1008.

Engle, R. F., Ghysels, E., & Sohn, B (2013). Stock market volatility and macroeconomic fundamentals, Review of Economics and Statistics, 95(3), 776–797.

Engle, R. & Lee, G. (1999). A long-run and short-run component model of stock return volatility. In Engle, R. & White, H. (eds.), Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger, pp. 475–497. Oxford University Press.

Engle, R. & Rangel, J. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. Review of Financial Studies 21, 1187–1222.

Engle, Robert, & Jeffrey Russell (1998). Autoregressive Conditional Duration: A New Model for Irregular Spaced Transaction Data. Econometrica 66, 1127–1162.

Fernandes, M., Mota, B. and G. Rocha (2005), "A multivariate conditional autoregressive range model," Economics Letters, 86, 435-440.

Ghysels, E., Sinko, A., & Valkanov, R. (2006) MIDAS regressions: Further results and new directions. Econometric Reviews 26, 53–90.

Hautsch, Nikolaus (2002) .Modelling Intraday Trading Activity Using Box-Cox-ACD Models, CoFE Discussion Paper, No. 02/05, University of Konstanz, Center of Finance and Econometrics (CoFE), Konstanz, http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-7845.

Nguyen, Duc Khuong & Walther, Thomas (2017*): Modeling and forecasting commodity market volatility with long-term economic and financial variables.

Parkinson, Michael (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. Journal of Business 53, 61–65.

Quirós, J.L., & Izquierdo, J.D. (2011). Volatility forecasting with range models: An evaluation of new alternatives to the CARR model.

Swanson, S. (2017). Range-Based Component Models for Conditional Volatility and Dynamic Correlations (Master's thesis). Retrieved fromhttps://digitalcollections.trentu.ca/islandora/object/etd:549

Zhang, M.Y., Russell, J. R., & Tsay, R.S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. J. Econometrics, 104(1):179–207.

Zou, W. & GU, M. (2014). Using CARR model and GARCH model to forecast volatility of the stock index: Evidence from China's Shanghai stock market. *2014* International Conference on Management Science & Engineering 21st Annual Conference Proceedings, Helsinki, pp. 1106-1112.

II. A GENERALIZED FEEDBACK ASYMMETRIC CONDITIONAL AUTOREGRESSIVE RANGE MODEL

ABSTRACT

The Conditional Autoregressive Range (CARR) model is an alternative to the Generalized Autoregressive Conditionally Heteroscedastic (GARCH) approach of modeling volatility. The former models the price range and the latter focuses on modeling the price returns. The Asymmetric CARR (ACARR) model was introduced for separate modeling of upward and downward ranges observed within each day, with the actual range expressed as the sum of these two components. This formulation, however, ignores feedback from one type of range to another. The Feedback Asymmetric Conditional Autoregressive Range (FACARR) was introduced in 2018 to remedy this drawback. The FACARR, however, limits this cross feedback to past ranges and does not include past conditional means. The proposed Generalized Feedback Asymmetric Conditional Autoregressive Range Model (GFACARR) removes this limitation and allows the upward range model to include past upward and past downward ranges, along with their respective conditional means. A similar model is defined for modeling downward range as well. The proposed model is more aligned with the multivariate CARR model. The use of the GFACARR model is illustrated by its application to several price series, including the S&P 500.

Key Words: Volatility Modeling, CARR Models, ACARR, Price Range, Time Series.

1. INTRODUCTION

Financial volatility is a factor that policy makers and investors should consider prior to any form of financial decision making. Modelling volatility is crucial to understanding the nature and dynamics of the financial market. Financial volatility of asset prices has been discussed extensively in past financial and econometric literature. One of the most successful volatility models used in a time series setting is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model introduced by Bollerslev (1986). Engle (1982) proposed the Autoregressive Conditional Heteroscedasticity (ARCH) model to address the complexities of time varying volatility and volatility clustering in the financial time series. In the ARCH formulation, the conditional volatility is modeled as a function of past returns. The GARCH model is an extension of the ARCH formulation and models the conditional volatility as a function of lagged squared returns, as well as past conditional variances. Since all aforementioned models focus on exhibiting price returns, they can be identified as examples of return-based volatility models.

In many financial time series applications, standard deviation is the most common measure of the stock return volatility because it not only calculates the dispersion of returns but also summarizes the probability of seeing extreme values in returns. Since the time the concept of volatility was introduced, researchers have sought alternative measures of financial volatility. One such alternative is the range. Range measures the dispersion of a random variable. Parkinson (1980) argued that volatility measure could be calculated by considering the daily high, daily low, and opening prices of a stock, in addition to the traditional closing prices. The paper also compared traditional measures of volatility that were calculated simply by using closing prices with extreme value methods; Parkinson did this by taking the high and low prices of an asset. The paper concluded that the range-based method was far superior to the standard methods based on returns. Beckers (1983) tested the validity of different volatility estimators. In this paper, Beckers mentioned that the range of a stock price contains important and fresh information. He also mentioned that using the range of a stock price was better than using the close-to-close changes. Kunitomo (1992) improved the Parkinson's original result and proposed a new range-based estimator, which was 10 time more efficient than the standard volatility estimator. In another study, Alizadeh, Brandt, and Diebold (2002) proved that the range-based volatility estimators were efficient compared to the classical volatility proxies, based on log absolute returns or squared returns. It showed that log range was approximately normal. Hence, the range of an asset price for a given period can be used as an informative proxy variable to measure an asset's volatility for a well-defined period, such as a day.

According to the results of Alizadeh et al. (2002), the GARCH family of models and stochastic volatility models (Tylor, 1986) ignored the price fluctuations of the reference period, making them relatively inaccurate and inefficient. Therefore, some researchers focused on an alternative approach to volatility modeling and developed the theoretical framework for range-based models with comprehensive empirical examples. For examples, refer to Chou (2005), Chou (2006), Brandt and Jones (2006), and Chou and Liu (2010). Chou (2005) introduced the Conditional Auto Regressive Range (CARR) model as a special case of the Autoregressive Conditional Duration (ACD) model of Engle (1998). The CARR is employed to model price volatility of an asset by considering range of the log prices for a given fixed time interval. The CARR model is similar to the standard volatility models, such as the GARCH formulation. However, one distinct difference between the two models is that the GARCH model uses rate of return as its volatility measure, while the CARR model uses the range as its volatility measure. The CARR model proposed by Chou is a simple, but an efficient, tool to analyze the volatility clustering property when compared to the GARCH models. This was shown empirically by Chou via out of sample forecasting from S&P 500 data. Chou showed that the effectiveness of volatility estimates produced by the CARR model is higher than the estimates from standard return-based models, such as the GARCH. Brandt and Jones (2006) integrated the properties of the exponential GARCH (Nelson, 1991) with daily log range data, and he proposed the range-based EGARCH model. This model has a simple framework but is an effective tool for capturing the important characteristics in stock return data, such as clustering, negative correlation, and log normality. The range-based EGARCH model is different from the CARR model in many ways. For example, it utilizes the lagged log range, rather than lagged range, as in the CARR model. Moreover, the range-based EGARCH model formulate conditional return volatility, while CARR explain the conditional mean of the range data.

Extensive modifications were made to the original CARR model. Chiang, Chou, and Wang (2016) suggested the lognormal log CARR model in an outlier detection process and showed that the proposed method effectively detected outliers. One major advantage of using the Log CARR model is that these models relax positivity restrictions on the parameters of the conditional expectation function. Xie and Wu (2017) explained the disturbance term in the CARR model by using the gamma distribution (GCARR) and showed through empirical data that the GCARR outperformed the Weibull CARR (WCARR) model in forecasting ability. The multivariate extension to the CARR

(MCARR) model was proposed by Fernandes, Mota and Rocha (2005), and they derived conditions for stationarity, geometric ergodicity, and beta-mixing with exponential decay. Chou and Liu (2009) incorporated the return-based Dynamic Conditional Correlation (DCC) model of Engle (2002) with the CARR model and introduced the new class of range-based DCC models. They concluded that the range-based DCC model outperformed other return-based models (MA100, EWMA, CCC, return-based DCC, and diagonal BEKK) using the RMSE and the MAE, the accepted benchmarks of implied and realized covariance. Different types of range-based volatility models such as Chou and Liu (2010), Miao, Wu and Su (2012), and Xie and Wang (2013) are some of the variations that were found in the published literature. For additional details, refer to Chou, R., Chou, H., and Liu (2015), which provided a comprehensive review of range-based models.

The asymmetric volatility, which is a key phenomenon in financial data, suggested that conditional volatilities show higher fluctuations during downward trends than during upward trends. Traditional methods of modelling return series, such as the ARCH and the GARCH models, use standard deviation, which treat price returns symmetrically. Hence, they are not effective tools for capturing the asymmetric behavior present in the financial data. To model the asymmetry in stock returns, several econometric models were introduced in the literature. The asymmetric ARCH model from Nelson (1991); EGARCH by Nelson and Cao (1992); GJR-GARCH model by Glosten, Jagannathan and Runkle (1993); and QGARCH by Sentana (1995) were developed. These models overcame the drawbacks of the GARCH models. In their paper, Engle and Ng (1993) analyzed how the news effected the conditional volatility and concluded that the EGARCH and the GJR-GARCH capture the asymmetry, but the latter is the better model.

All the above models capture the asymmetry in return data. The CARR model proposed by Chou (2005) use range as the measure of price volatility. The study treated the maximum and the minimum price symmetrically. However, in the same study, he suggested that the CARRX models (CARRX-a, and CARRX-b) by including exogenous variables such as (a) lagged return and (b) lagged absolute returns in the conditional mean equation. The purpose of this incorporation was to model one form of asymmetry, which was the leverage effect of Black and Nelson (1991). Chou (2006) presented an Asymmetric CARR (ACARR) model in which upward and downward price ranges are treated separately. The upward range is defined as the difference between the maximum price and the opening price, and the downward range is defined as the difference between the opening price and the minimum price, all of which are observed within one trading day. These definitions can be extended to periods beyond a day in a similar fashion. Instead of treating the high and the low prices for a given fixed period symmetrically, as in the CARR, the ACARR model incorporate a form of asymmetry by allowing the dynamic structure of the upward price movements to be different from that of the downward price movements. The ACARR model was extended to the ACARRX model by including exogenous variables, such as trading volume (Lamourex and Lastrapes, 1990), lag return to count leverage effect (Black, 1976; Nelson, 1990), or a seasonal factor. It assumed independence between the upward ranges and the downward ranges; therefore, parameters were estimated separately for each movement by using the QMLE method. An empirical study showed that the volatility forecasting ability of the ACARR model was superior to that of the CARR model. Chou and Wang (2014) combined the ACARR model, to capture current asymmetric volatility, with extreme value theory (EVT) to estimate the tail of the residual distribution.

This methodology gave better Value at Risk (VaR) estimates than the GARCH model as used by McNeil and Frey (2000).

Motivated by the independence between upward swing and downward plunge assumptions made by Chou (2006), Xie (2018) proposed the Feedback Asymmetric CARR (FACARR) model. By providing satisfactory evidence, Xie questioned the validity of the independence assumption and found cross-interdependence between upward movement and downward movement. Hence, the FACARR model was proposed as a more practical extension of the ACARR model. Put simply, both upward and downward movements of asset prices were not only modeled asymmetrically, but the conditional mean upward (downward) range was modeled by incorporating lagged downward (upward) ranges into each sub-model. Extensive empirical studies showed that the proposed FACARR performed significantly better than ACARR for both in sample and out of sample forecasting.

It is reasonable to assume that the dynamic movement of the upward (downward) range does not depend only on the lagged downward (upward) price range but also on the conditional mean of downward (upward) ranges. By consolidating on this fact, it was decided to generalize the previous class of asymmetric CARR models and introduce the Generalized Feedback Asymmetric CARR (GFCARR) model. The proposed model attempt to overcome the limitation of previous models by incorporating the cross-feedback term to account for the past conditional means. Since the proposed GFACARR model treat upward and downward price ranges separately, this approach also allow the modeling of the asymmetry found in financial data.

The paper is organized as follows. In Section 2, a brief introduction to the CARR, ACARR and FACARR models are given. The proposed GFACARR model is introduced, in addition to its statistical properties, in section 3. Econometric methodology is presented in Section 4, and the results of a simulation study are presented in Section 5. An empirical study based on three different stock market indices namely, S&P 500, CAC 40 and NIKKEI 225 is discussed in Section 6, and the conclusion is given in Section 7.

2. REVIEW OF CARR, ACARR AND FACARR MODELS

2.1. THE CONDITIONAL AUTOREGRESSIVE RANGE (CARR) MODEL

Chou (2005) proposed the CARR, which is primarily a range-based model. The CARR formulation is used to model the price volatility of an asset by considering range as a measure of this volatility. Let R_t be the price range defined over the fixed time period *t*, where R_t is the difference between the highest (P_t^{high}) and the lowest (P_t^{low}) logarithmic prices of an asset during the time period *t*. That is,

$$R_t = P_t^{high} - P_t^{low}.$$

The CARR model of order (*p*, *q*) is presented as CARR (*p*, *q*) and defined as follows:

$$\begin{aligned}
&R_t = \lambda_t \varepsilon_t, \\
&\mathrm{E}\left(R_t \mid \mathrm{F}_{t-1}\right) = \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i R_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \\
&\varepsilon_t \sim i.i.d.\, f(.), E\left(\varepsilon_t\right) = 1, \text{ and} \\
&0 < \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1, \alpha_i > 0, \beta_j > 0.
\end{aligned}$$

Here λ_t is the conditional expectation of the price range based on all information up to time t-1. The non-negative disturbance term, also known as the standardized range, is denoted by ε_t, which is independent and identically distributed with probability density function $f(.)$ with a non-negative support and a unit mean.

2.2. THE ASYMMETRIC CONDITIONAL AUTOREGRESSIVE RANGE (ACARR) MODEL

The ACARR model presented by Chou (2006) decomposed the range (R_t) series into two components, namely upward range (R_t^u) and downward range (R_t^d). The upward and downward ranges are defined using the differences between the daily high (P_t^{high}), the daily low (P_t^{low}), and the opening (P_t^{open}) logarithmic price of an asset over the time interval associated with t as follows:

$$\begin{aligned}
R_t^u &= P_t^{high} - P_t^{open}, \\
R_t^d &= P_t^{open} - P_t^{low}, \\
R_t &= R_t^u + R_t^d = P_t^{high} - P_t^{open} + P_t^{open} - P_t^{low} = P_t^{high} - P_t^{low}.
\end{aligned} \tag{2.1}$$

Here, the upward range measures the maximum gain or the positive shock to the stock while downward range calculates the minimum gain or the negative impact to the stock price for the time period t.

The CARR model is symmetric because it treats the high and low price in a symmetric way. However, it is possible that the upward and the downward movements are different in their dynamics of stock propagation. To allow the asymmetric behavior in price range data, Chou (2006) proposed and developed the ACARR model. The ACARR model of order (p, q) is as follows:

$$
\begin{aligned}
&R_t = R_t^u + R_t^d, \\
&R_t^u = \lambda_t^u \varepsilon_t^u, \\
&R_t^d = \lambda_t^d \varepsilon_t^d, \\
&\lambda_t^u = \omega^u + \sum_{i=1}^{p} \alpha_i^u R_{t-i}^u + \sum_{j=1}^{q} \beta_j^u \lambda_{t-j}^u, \\
&\lambda_t^d = \omega^d + \sum_{i=1}^{p} \alpha_i^d R_{t-i}^d + \sum_{j=1}^{q} \beta_j^d \lambda_{t-j}^d, \\
&\varepsilon_t^u \sim i.i.d.\, f^u(.), E(\varepsilon_t^u) = 1, \\
&\varepsilon_t^d \sim i.i.d.\, f^d(.), E(\varepsilon_t^d) = 1, \\
&0 < \sum_{i=1}^{p} \alpha_i^u + \sum_{j=1}^{q} \beta_j^u < 1, \alpha_i^u > 0, \beta_j^u > 0, \text{ and} \\
&0 < \sum_{i=1}^{p} \alpha_i^d + \sum_{j=1}^{q} \beta_j^d < 1, \alpha_i^d > 0, \beta_j^d > 0.
\end{aligned} \tag{2.2}
$$

Here $\lambda_t^u \left(= E\left(R_t^u \mid \mathrm{F}_{t-1}\right)\right)$ is the conditional mean of the upward range and $\lambda_t^d \left(= E\left(R_t^d \mid \mathrm{F}_{t-1}\right)\right)$ is the conditional mean of the downward range, both conditional on all information up to time period t-1. The disturbance term of the upward (downward) range model $\varepsilon_t^u \left(\varepsilon_t^d\right)$ is independently and identically distributed with the density function $f^u(.) \left(f^d(.)\right)$ with unit mean. Moreover, the upward and downward disturbance terms are independent such that: $\mathrm{cov}\left(\varepsilon_t^u, \varepsilon_t^d\right) = 0.$ The pairs of parameters, $\left(\omega^u, \omega^d\right), \left(\alpha_i^u, \alpha_i^d\right), \left(\beta_j^u, \beta_j^d\right)$ identify the asymmetric behavior between the upward range and downward range components.

2.3. THE FEEDBACK ASYMMETRIC CONDITIONAL AUTOREGRESSIVE RANGE (FACARR) MODEL

The ACARR model assumes that there is independence between the upward and downward range components and Xie (2018), argued against this assumption and presented the FACARR model. This model includes the cross-interdependence terms on top of the

ACARR setting. Following the same definitions and notations, the FACARR model is defined as follows:

$$
\begin{aligned}
&R_t = R_t^u + R_t^d, \\
&R_t^u = \lambda_t^u \varepsilon_t^u, \\
&R_t^d = \lambda_t^d \varepsilon_t^d, \\
&\lambda_t^u = \omega^u + \sum_{i=1}^{p} \alpha_i^u R_{t-i}^u + \sum_{j=1}^{q} \beta_j^u \lambda_{t-j}^u + \sum_{k=1}^{l} \gamma_k^u R_{t-k}^d, \\
&\lambda_t^d = \omega^d + \sum_{i=1}^{p} \alpha_i^d R_{t-i}^d + \sum_{j=1}^{q} \beta_j^d \lambda_{t-j}^d + \sum_{k=1}^{l} \gamma_k^d R_{t-k}^u, \\
&\varepsilon_t^u \sim i.i.d. f^u(.), E(\varepsilon_t^u) = 1, \\
&\varepsilon_t^d \sim i.i.d. f^d(.), E(\varepsilon_t^d) = 1, \\
&0 < \sum_{i=1}^{p} \alpha_i^u + \sum_{j=1}^{q} \beta_j^u < 1, \alpha_i^u > 0, \beta_j^u > 0, \text{ and} \\
&0 < \sum_{i=1}^{p} \alpha_i^d + \sum_{j=1}^{q} \beta_j^d < 1, \alpha_i^d > 0, \beta_j^d > 0.
\end{aligned}
\tag{2.3}
$$

Similar to the ACARR model the independence between upward and downward disturbance terms are assumed (i.e., $\text{cov}(\varepsilon_t^u, \varepsilon_t^d) = 0$) in the FACARR model. In addition to the previous parameter set discussed in the model (2.2), the FACARR has a new pair of parameters, namely, (γ^u, γ^d), which measures the magnitude and the direction of the lagged upward (downward) range on conditional mean range.

3. GENERALIZED FEEDBACK ASYMMETRIC CONDITIONAL AUTOREGRESSIVE RANGE MODEL (GFACARR) AND STATISTICAL PROPERTIES

Let P_t^{open}, P_t^{high} and P_t^{low} be the opening, high and low logarithmic prices of the speculative asset, respectively, at a given time period *t* (i.e., day). The observed price range for the time period *t* is denoted as R_t, and it is defined as the sum of the upward range (R_t^u) and downward range (R_t^d):

$$R_t = P_t^{high} - P_t^{low} = \left[P_t^{high} - P_t^{open}\right] + \left[P_t^{open} - P_t^{low}\right] = R_t^u + R_t^d .$$

Here, upward and the downward range components are defined the same as in the ACARR model. The proposed GFACARR model is as follows:

$$\begin{aligned}
&R_t = R_t^u + R_t^d,\\
&R_t^u = \lambda_t^u \varepsilon_t^u,\\
&R_t^d = \lambda_t^d \varepsilon_t^d,\\
&E\left(R_t^u \mid \mathrm{F}_{t-1}\right) = \lambda_t^u = \omega^u + \sum_{i=1}^{p^u} \alpha_i^u R_{t-i}^u + \sum_{j=1}^{q^u} \beta_j^u \lambda_{t-j}^u + \sum_{k=1}^{r^u} \gamma_k^u R_{t-k}^d + \sum_{l=1}^{s^u} \delta_l^u \lambda_{t-l}^d,\\
&E\left(R_t^d \mid \mathrm{F}_{t-1}\right) = \lambda_t^d = \omega^d + \sum_{i=1}^{p^d} \alpha_i^d R_{t-i}^d + \sum_{j=1}^{q^d} \beta_j^d \lambda_{t-j}^d + \sum_{k=1}^{r^d} \gamma_k^d R_{t-k}^u + \sum_{l=1}^{s^u} \delta_l^d \lambda_{t-l}^u,\\
&\varepsilon_t^u \sim i.i.d\, f^u(.), E\left(\varepsilon_t^u\right) = 1,\\
&\varepsilon_t^d \sim i.i.d\, f^d(.), E\left(\varepsilon_t^d\right) = 1, \text{ and}\\
&\omega^u > 0, \alpha_i^u > 0, \beta_j^u > 0; \omega^d > 0, \alpha_i^d > 0, \beta_j^d > 0.
\end{aligned} \tag{3.1}$$

Here $\lambda_t^u \left(= E\left(R_t^u \mid \mathrm{F}_{t-1}\right)\right)$ is the conditional mean of the upward range based on all information up to time period *t*-1, and $\lambda_t^d \left(= E\left(R_t^d \mid \mathrm{F}_{t-1}\right)\right)$ is the conditional mean of the downward range on all information up to time period *t*-1. Note that the sigma field generated using information from setup to time period *t*-1, is denoted by F_{t-1}. The upward (downward) range disturbance term is denoted by ε_t^u $\left(\varepsilon_t^d\right)$, and it is independently and identically distributed with unit mean.

In contrast to the FACARR model introduced by Xie (2018), the significance of the proposed formulations that GFACARR model is capable of modeling the conditional expected upward (downward) range at time t based on the lagged downward (upward) ranges along with the previous conditional expectation of downward (upward) ranges.

3.1. THE GFACARR MODEL

Here, the mean conditional upward (downward) range at time period t, is modeled by considering both downward (upward) range and mean conditional downward (upward) range at time t-1, in addition to the existing terms.

The GFACARR model given in the equation (3.1) can be re-written as a bivariate CARR (1, 1) model as follows:

$$\vec{R}_t = \Lambda_t(\Phi)\vec{\varepsilon}_t, \tag{3.2}$$

where $\Lambda_t(\Phi) = diag\left\{\lambda_t^u, \lambda_t^d\right\}$ and $\lambda_t^i\,(i = u, d)$, is the conditional mean of $R_t^i\,(i = u, d)$ given F_{t-1}. Here $\Phi = \left(\omega^u, \alpha^u, \beta^u, \gamma^u, \delta^u, \omega^d, \alpha^d, \beta^d, \gamma^d, \delta^d\right)'$ is the parameter vector and $\vec{\varepsilon}_t = \left(\varepsilon_t^u \quad \varepsilon_t^u\right)'$ has following conditions imposed on it:

Case 1: $\operatorname{cov}\left(\varepsilon_t^u, \varepsilon_t^d\right) = 0$.

A. $\left\{\vec{\varepsilon}_t, t \in \mathbf{Z}^+\right\}$ is a sequence of independent and identically distributed□ ^{2}-valued random variables with $E\left(\vec{\varepsilon}_t\right) = \begin{pmatrix} E\left(\varepsilon_t^u\right) \\ E\left(\varepsilon_t^d\right) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. For the illustrative purpose it is assumed that $\varepsilon_t^i \sim \exp(1), \forall i = u,d.$ Then, the covariance matrix becomes $\Gamma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}_{2\times 2} = I_2.$

B. $\operatorname{cov}\left(\varepsilon_j^i, \varepsilon_k^i\right) = 0, \forall j \neq k, \forall i = u,d.$

C. From conditions1 and 2, conditional covariance matrix of $\vec{R}_t$ as follows and it is given by $H_t(\Phi)$ such that,

$$H_t(\Phi) = \begin{pmatrix} \left(\lambda_t^u\right)^2 & 0 \\ 0 & \left(\lambda_t^d\right)^2 \end{pmatrix}_{2*2} = \Lambda^2_{\ t}(\Phi).$$

Case 2: $\operatorname{cov}\left(\varepsilon_t^u, \varepsilon_t^d\right) = a \neq 0.$

A. $\left\{\vec{\varepsilon}_t, t \in \mathbf{Z}^+\right\}$ is an $\mathbb{R}^2$-valued random variables with $E\left(\vec{\varepsilon}_t\right) = \begin{pmatrix} E\left(\varepsilon_t^u\right) \\ E\left(\varepsilon_t^d\right) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. For the illustrative purpose, itis assumed that $\varepsilon_t^i \sim \exp(1), \forall i = u,d.$ Then the covariance matrix becomes $\Gamma = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}_{2\times 2}$.

B. $\operatorname{cov}\left(\varepsilon_j^i, \varepsilon_k^i\right) = 0, \forall j \neq k, \forall i = u,d.$

C. From conditions 1 and 2, conditional covariance matrix of $\vec{R}_t$ as follows and it is given by $H_t(\Phi)$ such that,

$$H_t(\Phi) = \begin{pmatrix} \left(\lambda_t^u\right)^2 & a\left(\lambda_t^u \lambda_t^d\right) \\ a\left(\lambda_t^u \lambda_t^d\right) & \left(\lambda_t^d\right)^2 \end{pmatrix}_{2*2} = \Lambda^2_{\ t}(\Phi).$$

This representation of model (3.1) coincides with the bivariate GARCH (1, 1) process with constant correlation (see. Bollerslev, 1990; Jeantheau, 1998).

The GFACARR process can be formulated as a bivariate CARR (1, 1) model as follows:

$$\begin{pmatrix} \lambda_t^u \\ \lambda_t^d \end{pmatrix} = \begin{pmatrix} \omega^u \\ \omega^d \end{pmatrix} + \begin{pmatrix} \alpha^u & \gamma^u \\ \gamma^d & \alpha^d \end{pmatrix} \begin{pmatrix} R_{t-1}^u \\ R_{t-1}^d \end{pmatrix} + \begin{pmatrix} \beta^u & \delta^u \\ \delta^d & \beta^d \end{pmatrix} \begin{pmatrix} \lambda_{t-1}^u \\ \lambda_{t-1}^d \end{pmatrix},$$
$$\overrightarrow{\lambda_t} = \vec{\omega} + A\overrightarrow{R_{t-1}} + B\overrightarrow{\lambda_{t-1}}. \tag{3.3}$$

Here, $\overrightarrow{\lambda_t} = \begin{pmatrix} \lambda_t^u \\ \lambda_t^d \end{pmatrix}$, $\overrightarrow{R_{t-1}} = \begin{pmatrix} R_{t-1}^u \\ R_{t-1}^d \end{pmatrix}$, $\vec{\omega} = \begin{pmatrix} \omega^u \\ \omega^d \end{pmatrix}$, $A = \begin{pmatrix} \alpha^u & \gamma^u \\ \gamma^d & \alpha^d \end{pmatrix}$ and $B = \begin{pmatrix} \beta^u & \delta^u \\ \delta^d & \beta^d \end{pmatrix}$.

If the vector $\vec{\omega} \in \mathbb{R}^2_{>0}$, and all the coefficients in $A \in [a_{i,j}], B \in [b_{i,j}] \forall i, j;$ and $i = 1, 2; j = 1, 2.$ are non-negative, then this is sufficient for the non-negativity of the $\overrightarrow{\lambda_t} \in \mathbb{R}^2_{\geq 0}$. However, in this proposed model negative values are allowed for the coefficients of the newly introduced lagged conditional expected upward (downward) term. Since both the range and the conditional mean range are positive variables, it is important to preserve the positivity of the model. We studied closely the conditions for the non-negativity and positivity imposed in the Dynamic Conditional Correlation Multivariate GARCH models (Engle and Sheppard, 2001). Nelson and Cao (1992) introduced non-negativity constraints for the GARCH (*p*, *q*) models by relaxing the above mentioned sufficient condition. Since these conditions were not readily applicable to our model, we modified the conditions to suit our model. Our conditions are that $\vec{\omega} \in \mathbb{R}^2_{>0}$, with positive coefficients, and that $[a_{i,j}] \in \mathbb{R}$ and $[b_{i,j}] \in \mathbb{R}$ such that $[a_{i,j} + b_{i,j}] \in \mathbb{R}, \forall i, j.$ with eigenvalues of $(A + B)_{2\times 2}$, expressed as Δ_1 and Δ_2, follow the restrictions $|\Delta_1| < 1$ and $|\Delta_2| < 1$.

3.2. STATISTICAL PROPERTIES OF THE GFACARR MODEL

3.2.1. Weak Stationarity of GFACARR Model. Since the GFACARR can be presented as the bivariate CARR (1, 1) model, and it can be reparametrized as a bivariate ARMA (1, 1) model. Derivation is given:

$$\begin{aligned}
&\overrightarrow{R_t} - E\left(\overrightarrow{R_t} \mid \mathrm{F}_{t-1}\right) = \overrightarrow{R_t} - \overrightarrow{\lambda_t} = \overrightarrow{\eta_t}, \\
&\overrightarrow{\lambda_t} = \overrightarrow{R_t} - \overrightarrow{\eta_t} = \vec{\omega} + A\overrightarrow{R_{t-1}} + B\left(\overrightarrow{R_{t-1}} - \overrightarrow{\eta_{t-1}}\right), \\
&\overrightarrow{R_t} = \vec{\omega} + A\overrightarrow{R_{t-1}} + B\overrightarrow{R_{t-1}} + \overrightarrow{\eta_t} + (-B)\overrightarrow{\eta_{t-1}}.
\end{aligned} \tag{3.4}$$

Let $\vec{\eta}_t$ be the difference vector and $\overrightarrow{R_t} = \vec{\omega} + (A+B)\overrightarrow{R_{t-1}} + \overrightarrow{\eta_t} - B\overrightarrow{\eta_{t-1}}$ be a Bivariate ARMA (1, 1) model. If all the eigenvalues of the matrix $(A+B)_{2\times 2}$ are positive, but less than one, then the Bivariate ARMA (1, 1) model for $\overrightarrow{R_t}$ is weakly stationary (Tsay, 2002). By following this claim, we proposed the weak stationarity conditions for the GFACARR model.

Theorem 1: Let $\overrightarrow{\lambda_t} = \vec{\omega} + A\overrightarrow{R_{t-1}} + B\overrightarrow{\lambda_{t-1}}$ be the GFACARR process defined in (3.1) - (3.4). If all the eigenvalues of $(A+B)_{2\times 2}$, namely Δ_1 and Δ_2 are such that $|\Delta_i| < 1\ \forall i$, then the GFACARR model for $\overrightarrow{R_t}$ is weak stationary.

The Proof of this theorem will be presented in the Appendix A.

3.2.2. Unconditional Expectation of GFACARR Model. Under the weak stationarity assumption, $E\left(\overrightarrow{R_t}\right) = E\left(\overrightarrow{R_{t-1}}\right)$, and $E\left(\overrightarrow{\eta_t}\right) = E\left(\overrightarrow{R_t} - \overrightarrow{\lambda_t}\right) = 0$ so that:

$$\begin{aligned}
&\overrightarrow{R_t} = \vec{\omega} + (A+B)\overrightarrow{R_{t-1}} + \overrightarrow{\eta_t} + (-B)\overrightarrow{\eta_{t-1}}, \\
&E\left(\overrightarrow{R_t}\right) = \vec{\omega} + (A+B)E\left(\overrightarrow{R_{t-1}}\right) + E\left(\overrightarrow{\eta_t}\right) + E\left[(-B)\overrightarrow{\eta_{t-1}}\right], \\
&E\left(\overrightarrow{R_t}\right) = \vec{\omega} + (A+B)E\left(\overrightarrow{R_{t-1}}\right) = \vec{\omega} + (A+B)E\left(\overrightarrow{R_t}\right), \\
&\left[I - (A+B)\right]E\left(\overrightarrow{R_t}\right) = \vec{\omega}.
\end{aligned}$$

The $\vec{R}_t$ is a weakly stationary and $\det\left[I-(A+B)\right]\neq 0$, hence $E\left(\vec{R}_t\right)$ exists. Thus,

$$E\left(\vec{R}_t\right)=\left[I-(A+B)\right]^{-1}\vec{\omega},$$

$$\left[I-(A+B)\right]^{-1}=\begin{pmatrix}1-\left(\alpha^u+\beta^u\right) & -\left(\gamma^u+\delta^u\right)\\ -\left(\gamma^d+\delta^d\right) & 1-\left(\alpha^d+\beta^d\right)\end{pmatrix}^{-1},$$

$$\left[I-(A+B)\right]^{-1}=\frac{1}{\left\{\left[1-\left(\alpha^u+\beta^u\right)\right]\left[1-\left(\alpha^d+\beta^d\right)\right]-\left[\gamma^u+\delta^u\right]\left[\gamma^d+\delta^d\right]\right\}}\begin{bmatrix}1-\left(\alpha^d+\beta^d\right) & +\left(\gamma^u+\delta^u\right)\\ +\left(\gamma^d+\delta^d\right) & 1-\left(\alpha^u+\beta^u\right)\end{bmatrix},$$

$$E\left(\vec{R}_t\right)=\frac{1}{\left\{\left[1-\left(\alpha^u+\beta^u\right)\right]\left[1-\left(\alpha^d+\beta^d\right)\right]-\left[\gamma^u+\delta^u\right]\left[\gamma^d+\delta^d\right]\right\}}\begin{bmatrix}1-\left(\alpha^d+\beta^d\right) & +\left(\gamma^u+\delta^u\right)\\ +\left(\gamma^d+\delta^d\right) & 1-\left(\alpha^u+\beta^u\right)\end{bmatrix}\begin{bmatrix}\omega^u\\ \omega^d\end{bmatrix},$$

$$E\left(\vec{R}_t\right)=\frac{1}{\left\{\left[1-\left(\alpha^u+\beta^u\right)\right]\left[1-\left(\alpha^d+\beta^d\right)\right]-\left[\gamma^u+\delta^u\right]\left[\gamma^d+\delta^d\right]\right\}}\begin{bmatrix}\left(1-\left(\alpha^d+\beta^d\right)\right)\omega^u+\left(\gamma^u+\delta^u\right)\omega^d\\ \left(1-\left(\alpha^u+\beta^u\right)\right)\omega^d+\left(\gamma^d+\delta^d\right)\omega^u\end{bmatrix}.$$

The unconditional mean of upward range $E\left(R_t^u\right)$ and unconditional mean of downward range $E\left(R_t^d\right)$, can be expressed as follows:

$$E\left(R_t^u\right)=\frac{\left[1-\left(\alpha^d+\beta^d\right)\right]\omega^u+\left(\gamma^u+\delta^u\right)\omega^d}{\left\{\left[1-\left(\alpha^u+\beta^u\right)\right]\left[1-\left(\alpha^d+\beta^d\right)\right]-\left(\gamma^u+\delta^u\right)\left(\gamma^d+\delta^d\right)\right\}},$$

$$E\left(R_t^d\right)=\frac{\left[1-\left(\alpha^u+\beta^u\right)\right]\omega^d+\left(\gamma^d+\delta^d\right)\omega^u}{\left\{\left[1-\left(\alpha^u+\beta^u\right)\right]\left[1-\left(\alpha^d+\beta^d\right)\right]-\left(\gamma^u+\delta^u\right)\left(\gamma^d+\delta^d\right)\right\}}.$$

Finally, the unconditional mean range $E\left(R_t\right)$ is calculated as:

$$E\left(R_t\right)=E\left(R_t^u\right)+E\left(R_t^d\right),$$

$$E\left(R_t\right)=\frac{\left[1-\left(\alpha^d+\beta^d\right)\right]\omega^u+\left(\gamma^u+\delta^u\right)\omega^d+\left[1-\left(\alpha^u+\beta^u\right)\right]\omega^d+\left(\gamma^d+\delta^d\right)\omega^u}{\left[1-\left(\alpha^u+\beta^u\right)\right]\left[1-\left(\alpha^d+\beta^d\right)\right]-\left(\gamma^u+\delta^u\right)\left(\gamma^d+\delta^d\right)}.$$

4. ESTIMATION OF GFACARR MODEL

4.1. BIVARIATE EXPONENTIAL GFACARR TYPE a (BEGFACARR-a) MODEL

Let $\{\varepsilon_t^u\}$ $\left(\{\varepsilon_t^d\}\right)$ be the sequence of independent and identically distributed exponential disturbance term with $E\left(\varepsilon_t^u\right)=1$ $\left(E\left(\varepsilon_t^d\right)=1\right)$, and $\left\{R_t^u\right\}_{t=1}^n=\{R_1^u,R_2^u,\ldots,R_n^u\}$ $\left(\left\{R_t^d\right\}_{t=1}^n=\{R_1^d,R_2^d,R_3^d,\ldots,R_n^d\}\right)$ be the realization of the model $R_t^u=\lambda_t^u\varepsilon_t^u$ $\left(R_t^d=\lambda_t^d\varepsilon_t^d\right)$. The parameter vector $\Phi=\left(\omega^u,\alpha^u,\beta^u,\gamma^u,\delta^u,\omega^d,\alpha^d,\beta^d,\gamma^d,\delta^d\right)'$ can be estimated by using the conditional likelihood method. In this section, we derive the log likelihood function for the proposed GFACARR model.

The conditional distribution of R_t^u and R_t^d given the information up to t-1, can be expressed as follows:

$$f\left(R_t^u \mid \mathrm{F}_{t-1},\Phi\right)\sim\frac{1}{\lambda_t^u}\exp\left(-\frac{R_t^u}{\lambda_t^u}\right),$$
$$f\left(R_t^d \mid \mathrm{F}_{t-1},\Phi\right)\sim\frac{1}{\lambda_t^d}\exp\left(-\frac{R_t^d}{\lambda_t^d}\right).$$

Since $\operatorname{cov}\left(\varepsilon_t^u,\varepsilon_t^d\right)=0,$ conditional distributions of $f\left(R_t^u \mid \mathrm{F}_{t-1},\Phi\right)$ and $f\left(R_t^d \mid \mathrm{F}_{t-1},\Phi\right)$ are conditionally independent, then the conditional joint distribution of the realized range data at time t, given the information set up to time t-1 is given by:

$$f\left(R_t^u,R_t^d \mid \mathrm{F}_{t-1},\Phi\right)=\left[f\left(R_t^u \mid \mathrm{F}_{t-1},\Phi\right)\right]\left[f\left(R_t^d \mid \mathrm{F}_{t-1},\Phi\right)\right],$$
$$f\left(R_t^u,R_t^d \mid \mathrm{F}_{t-1},\Phi\right)=\left[\frac{1}{\lambda_t^u}\exp\left(-\frac{R_t^u}{\lambda_t^u}\right)\right]\left[\frac{1}{\lambda_t^d}\exp\left(-\frac{R_t^d}{\lambda_t^d}\right)\right]. \tag{4.1}$$

Therefore, the conditional likelihood function $L\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right)$ and the log likelihood function of the data $l\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right)$ can be derived as follows:

$$L\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right) = \prod_{t=2}^n f\left(R_t^u, R_t^d \mid \mathrm{F}_{t-1}, \Phi\right),$$

$$l\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right) = \ln\left[L\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right)\right] = \sum_{t=2}^n \ln\left[f\left(R_t^u, R_t^d \mid \mathrm{F}_{t-1}, \Phi\right)\right],$$

$$l\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right) = -\sum_{t=2}^n \left[\left(\ln\left(\lambda_t^u\right) + \frac{R_t^u}{\lambda_t^u}\right) + \left(\ln\left(\lambda_t^d\right) + \frac{R_t^d}{\lambda_t^d}\right)\right]. \tag{4.2}$$

4.2. BIVARIATE EXPONENTIAL GFACARR TYPE b (BEGFACARR-b) MODEL

In this section, we relax the assumption of the independence between the residual durations by introducing the second bivariate distribution with exponential margins. This was proposed by Gumbel (1960), and here the parameter $\upsilon \in [-1,1]$ is used to capture the potential correlation ρ between the upward and downward residual components. However, the coefficient of correlation moves in the narrow interval such that $\rho \in [-0.25, 0.25]$.

Let $\left\{\varepsilon_t^u\right\}$ $\left(\left\{\varepsilon_t^d\right\}\right)$ be the sequence of upward (downward) residual components with $E\left(\varepsilon_t^u\right) = 1$ $\left(E\left(\varepsilon_t^d\right) = 1\right)$ and $\operatorname{cov}\left(\varepsilon_t^u, \varepsilon_t^d\right) = a = \frac{\upsilon}{4}$. Since the marginal distributions are exponentially distributed the conditional distribution of R_t^u and R_t^d given F_{t-1}, it can be expressed as follows:

$$f\left(R_t^u \mid \mathrm{F}_{t-1}, \Phi\right) \propto \frac{1}{\lambda_t^u} \exp\left(-\frac{R_t^u}{\lambda_t^u}\right),$$

$$f\left(R_t^d \mid \mathrm{F}_{t-1}, \Phi\right) \propto \frac{1}{\lambda_t^d} \exp\left(-\frac{R_t^d}{\lambda_t^d}\right).$$

Since $\operatorname{cov}\left(\varepsilon_t^u, \varepsilon_t^d\right) = a,$ then the conditional joint distribution of R_t^u and R_t^d the realized range data at time *t*, given the information set up to time *t*-1 is given by:

$$f\left(R_t^u, R_t^d \mid \mathrm{F}_{t-1}; \Phi\right) = \frac{1}{\lambda_t^u \lambda_t^d} \exp\left[-\left(\frac{R_t^u}{\lambda_t^u} + \frac{R_t^d}{\lambda_t^d}\right)\right]\left[1 + \upsilon\left(2\exp\left\{-\frac{R_t^u}{\lambda_t^u}\right\} - 1\right)\left(2\exp\left\{-\frac{R_t^d}{\lambda_t^d}\right\} - 1\right)\right]. \qquad (4.3)$$

The parameter vector $\Phi = \left(\omega^u, \alpha^u, \beta^u, \gamma^u, \delta^u, \omega^d, \alpha^d, \beta^d, \gamma^d, \delta^d, \upsilon\right)'$ can be estimated by maximizing the conditional likelihood function $l\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right)$. Therefore, the conditional likelihood function $L\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right)$ and the conditional log likelihood function of the data $l\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right)$ can be derived from:

$$L\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right) = \prod_{t=2}^{n} f\left(R_t^u, R_t^d \mid \mathrm{F}_{t-1}, \Phi\right),$$

$$l\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right) = \ln\left[L\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right)\right] = \sum_{t=2}^{n} \ln\left[f\left(R_t^u, R_t^d \mid \mathrm{F}_{t-1}, \Phi\right)\right],$$

$$l\left(\Phi \mid \left\{R_t^u, R_t^d\right\}_{t=1}^n\right) = -\sum_{t=2}^{n}\left\{\left(\ln\left(\lambda_t^u\right) + \frac{R_t^u}{\lambda_t^u}\right) + \left(\ln\left(\lambda_t^d\right) + \frac{R_t^d}{\lambda_t^d}\right) - \ln\left[1 + \upsilon\left(2\exp\left\{-\frac{R_t^u}{\lambda_t^u}\right\} - 1\right)\left(2\exp\left\{-\frac{R_t^d}{\lambda_t^d}\right\} - 1\right)\right]\right\}. \qquad (4.4)$$

5. SIMULATION STUDY

We investigated the finite sample performance of estimators using a simulation study. We used ‘nloptr’, a nonlinear optimization function of R software to generate the relevant data. Length of the time series studied was set to n =1000 and n =3000, and $m = 1000$. Simulation runs were carried out for each parameter sample size combination. We carried out this simulation study for two different cases of the GFACARR model, namely the BEGFACARR-a and the BEGFACARR-b, including or excluding independence between the upward and the downward range disturbance terms. In this simulation study, first, we generated the data for the BEGFACARR-a and the BEGFACARR-b models based on the equations (4.1) and (4.3), respectively. Then the profile likelihood functions (4.2) and (4.4) were maximized by using the constrained nonlinear optimization function 'nloptr' in R. The Mean Absolute Deviation Error (MADE) was utilized as the evaluation criterion. The MADE is defined as, $\frac{1}{m}\sum_{j=1}^{m}\left|\hat{\phi}_j - \phi_j\right|$, where m is the number of replications. Simulation results for the BEGFACARR-a models are reported in Table 1 and Table 2. The BEGFACARR-b model results are summarized in Table 3 and Table 4.

Table 1 and Table 2 show the simulation results related to the upward range component parameters of the BEGFACARR-a model, while Table 2 presents the downward range component of the BEGFACARR-a model. The results show that the estimates are close to the true parameter values in most cases, and that the MADE values are reasonably small, with an improvement seen in the 3000 sample size case. The conclusion based on this simulation study is that the maximum likelihood method provides reliable estimates of the model parameters, in spite of the complex nature of the model when compared to CARR or the FACARR models.

Table 1: Means of MLE estimates and MADE (within parentheses), for upward range component in BEGFACARR-a model

Upward Model	Sample Size	ω^u	α^u	β^u	γ^u	∂^u
True Coefficients		0.01	0.20	0.40	0.10	0.20
UPR_M1	n=1000	0.0113 (0.0103)	0.1968 (0.0320)	0.3871 (0.1111)	0.0973 (0.0256)	0.2079 (0.1203)
	n=3000	0.0101 (0.0068)	0.1992 (0.0183)	0.3937 (0.0620)	0.0989 (0.0139)	0.2063 (0.0775)
True Coefficients		0.01	0.30	0.50	0.10	-0.02
UPR_M2	n=1000	0.0117 (0.0049)	0.2960 (0.0315)	0.5081 (0.0662)	0.0991 (0.0106)	-0.0235 (0.0216)
	n=3000	0.0105 (0.0026)	0.2984 (0.0183)	0.5028 (0.0363)	0.0996 (0.0057)	-0.0211 (0.0116)
True Coefficients		0.15	0.20	0.60	0.10	-0.10
UPR_M3	n=1000	0.1548 (0.0255)	0.2047 (0.0279)	0.5428 (0.0943)	0.0968 (0.0153)	-0.0724 (0.0276)
	n=3000	0.1522 (0.0136)	0.2026 (0.0164)	0.5714 (0.0500)	0.0986 (0.0080)	-0.0863 (0.0137)

Table 2: Means of MLE estimates and MADE (within parentheses), for downward range component in BEGFACARR-a model

Downward Model	Sample Size	ω^d	α^d	β^d	γ^d	∂^d
True Coefficients		0.02	0.10	0.80	0.02	-0.05
DWNR_M1	n=1000	0.0235 (0.0089)	0.1011 (0.0221)	0.7714 (0.1115)	0.0183 (0.0261)	-0.0430 (0.0989)
	n=3000	0.0210 (0.0040)	0.1006 (0.0224)	0.7929 (0.0118)	0.0201 (0.0510)	-0.0500 (0.0143)
True Coefficients		0.04	0.10	0.60	0.03	0.60
DWNR_M2	n=1000	0.0438 (0.0142)	0.0965 (0.0257)	0.5919 (0.0642)	0.0322 (0.0612)	0.6157 (0.1788)
	n=3000	0.0415 (0.0077)	0.0990 (0.0144)	0.5935 (0.0374)	0.0295 (0.0354)	0.6154 (0.1084)
True Coefficients		0.10	0.20	0.40	0.10	0.50
DWNR_M3	n=1000	0.1004 (0.0593)	0.1975 (0.0308)	0.3934 (0.1171)	0.0951 (0.0490)	0.5213 (0.2376)
	n=3000	0.0934 (0.0355)	0.1991 (0.0172)	0.3954 (0.0643)	0.0967 (0.0278)	0.5242 (0.1440)

In the previous simulation study, we assumed the independence between the upward and downward residual components. In the next subsection, we considered the case of dependence between upward and downward residual components. Therefore, we fitted the second bivariate exponential distribution of Gumbel (1960). The BEGFACARR-b model parameters are estimated using MLE method and results are summarized in Table 3 and Table 4. Table 3 exhibits the parameter estimated results for the upward range model parameters and the MADE results, while Table 4 shows that of the downward range component.

Based on the simulation results, both the upward and downward parameters are estimated with reasonable accuracy. Furthermore, the estimated upward and downward components parameters in the BEGFACARR-b model has lower MADE value when compare to the BEGFACARR-a model. However, the estimates for the newly added υ term has a low accuracy when compare to the other model parameters.

6. EMPIRICAL STUDY

6.1. THE DATA SET

In this study, three stock indices from different markets were used to gauge the performance of the proposed GFACARR model. Daily data of the Standard and Poor's 500 (S&P 500) index of United States, CAC 40, which is a benchmark index of the French stock market, and Japan's NIKKEI 225 index were considered. The sample periods for S&P 500, CAC 40, and NIKKEI 225 were January 02, 2002 to December 31, 2019. Daily values for the opening price, closing price, high price, low price, and adjusted price were

Table 3: Means of MLE estimates and MADE (within parentheses), for upward range component in BEGFACARR-b model

Upward Model	Sample Size	ω^u	α^u	β^u	γ^u	∂^u	υ
True Coefficients		0.01	0.20	0.40	0.10	0.20	-0.40
UPR_M1	n=1000	0.0113 (0.0100)	0.1970 (0.0318)	0.3818 (0.1123)	0.0999 (0.0241)	0.2103 (0.1175)	-0.2918 (0.1223)
	n=3000	0.0099 (0.0069)	0.1995 (0.0185)	0.3922 (0.0597)	0.1003 (0.0139)	0.2076 (0.0728)	-0.2879 (0.1126)
True Coefficients		0.01	0.30	0.50	0.10	-0.02	-0.20
UPR_M2	n=1000	0.0115 (0.0051)	0.2973 (0.0316)	0.5002 (0.0681)	0.0998 (0.0099)	-0.0212 (0.0218)	-0.1463 (0.0916)
	n=3000	0.0105 (0.0026)	0.2986 (0.0185)	0.5010 (0.0384)	0.1002 (0.0056)	-0.0209 (0.0122)	-0.1439 (0.0650)
True Coefficients		0.15	0.20	0.60	0.10	-0.10	-0.40
UPR_M3	n=1000	0.1573 (0.0279)	0.1990 (0.0293)	0.6043 (0.1338)	0.1004 (0.0137)	-0.1082 (0.0586)	-0.2912 (0.1231)
	n=3000	0.1524 (0.0147)	0.1992 (0.0179)	0.6011 (0.0715)	0.1006 (0.0077)	-0.1029 (0.0297)	-0.2875 (0.1131)

Table 4: Means of MLE estimates and MADE (within parentheses), for downward range component in BEGFACARR-b model

Downward Model	Sample Size	ω^d	α^d	β^d	γ^d	∂^d
True Coefficients		0.02	0.10	0.80	0.02	-0.05
DWNR_M1	n=1000	0.0225 (0.0082)	0.0989 (0.0210)	0.7907 (0.0989)	0.0228 (0.0189)	-0.0599 (0.0786)
	n=3000	0.0207 (0.0038)	0.1000 (0.0119)	0.7964 (0.0482)	0.0194 (0.0124)	-0.0505 (0.0423)
True Coefficients		0.04	0.10	0.60	0.03	0.60
DWNR_M2	n=1000	0.0435 (0.0146)	0.0952 (0.0260)	0.5977 (0.0651)	0.0461 (0.0432)	0.5921 (0.1635)
	n=3000	0.0409 (0.0077)	0.0990 (0.0150)	0.6008 (0.0358)	0.0337 (0.0285)	0.5934 (0.0978)
True Coefficients		0.10	0.20	0.40	0.10	0.50
DWNR_M3	n=1000	0.1100 (0.0637)	0.1944 (0.0320)	0.3735 (0.1217)	0.0973 (0.0472)	0.5454 (0.2355)
	n=3000	0.1022 (0.0065)	0.1986 (0.0162)	0.3888 (0.0691)	0.0960 (0.0279)	0.5226 (0.1442)

Finance web page (https://finance.yahoo.com/) by using a 'quantmod' package in R software. The data set was divided in to two sub-samples where the first sub-sample, also known as in-sample period, was used to estimate the model parameters and in-sample predictions. In-sample periods for S&P 500, CAC 40, and NIKKEI 225 spanned from January 02, 2002 to December 31, 2018. The second-sub sample, which is also called the out-of-sample period, was used for out-of-sample forecasting. The out-of-sample periods for S&P 500, CAC 40, and NIKKEI 225 were from January 1, 2019 to December 31, 2019. In general, Table 5 presents the summarization of the three stock indices; more specifically Table 5A, Table 5B and Table 5C present the summary statistics of the S&P 500, CAC 40, and NIKKEI 225 daily stock indices, respectively. Daily price range, daily upward range, and daily downward range values were calculated as discussed in equation (2.1).

According to the summary statistic results, both upward and downward price range series for all three stock indices have large positive skewness, and these values are suggested that a positively skewed density functions should be used to model the disturbance terms. After 22 lags of sample autocorrelations were examined, the large test statistic values and very small p-values (<0.0001) conclude that the data exhibit a strong persistence in daily price range data. Downward range components have higher Ljung-Box statistic than that for the upward range components, which means that downward range component is more persistent than the upward range component. Furthermore, higher values for the mean and standard deviation of the downward range component when compare to the upward range component, is a primary indication of the difference between the upward and downward range components' volatility structures. The correlation coefficient between upward range and downward range components for all the three stock indices are significant at 0.001 significance level, and these negative correlation values suggest that periods of higher downward range volatility are related to lower upward range volatility periods.

Table 5: Summary statistics of the daily range, upward range and downward range of S&P 500, CAC 40 and NIKKEI 225 indices

Table 3A: Summary statistics of S&P 500: 01/02/2002 – 12/31/2019			
Summary Statistics	Upward Range Component	Downward Range Component	Range
Minimum	0.0000	0.0000	0.1456
Mean	0.6019	0.6466	1.2484
Maximum	10.2457	9.5522	10.9041
Standard Deviation	0.7189	0.8340	0.9931
Skewness	3.5231	3.3977	3.3472
Q (22)	3552.20***	6906.20***	32270.00***
Correlation (UPR, WNR)	-0.1885***		
Table 3B: Summary statistics of CAC 40: 01/02/2002 – 12/31/2019			
Summary Statistics	Upward Range Component	Downward Range Component	Range
Minimum	0.0000	0.0000	0.1388
Mean	0.7029	0.7969	1.4998
Maximum	8.4229	7.7503	9.2607
Standard Deviation	0.7430	0.8635	1.0222
Skewness	2.7823	2.4814	2.3132
Q (22)	2339.00***	7267.30***	27538.00***
Correlation (UPR, WNR)	-0.1969***		
Table 3C: Summary statistics of NIKKEI 225: 01/02/2002 – 12/31/2019			
Summary Statistics	Upward Range Component	Downward Range Component	Range
Minimum	0.0000	0.0000	0.0000
Mean	0.6233	0.6794	1.3027
Maximum	11.7433	13.7634	13.7634
Standard Deviation	0.6897	0.8199	0.9123
Skewness	3.3471	4.3674	3.8723
Q (22)	2029.80***	1548.60***	14469.00***
Correlation (UPR, WNR)	-0.2791***		

Note: *** indicate significance at 1% level. Q (22) is the Ljung-Box statistics of lag 22.

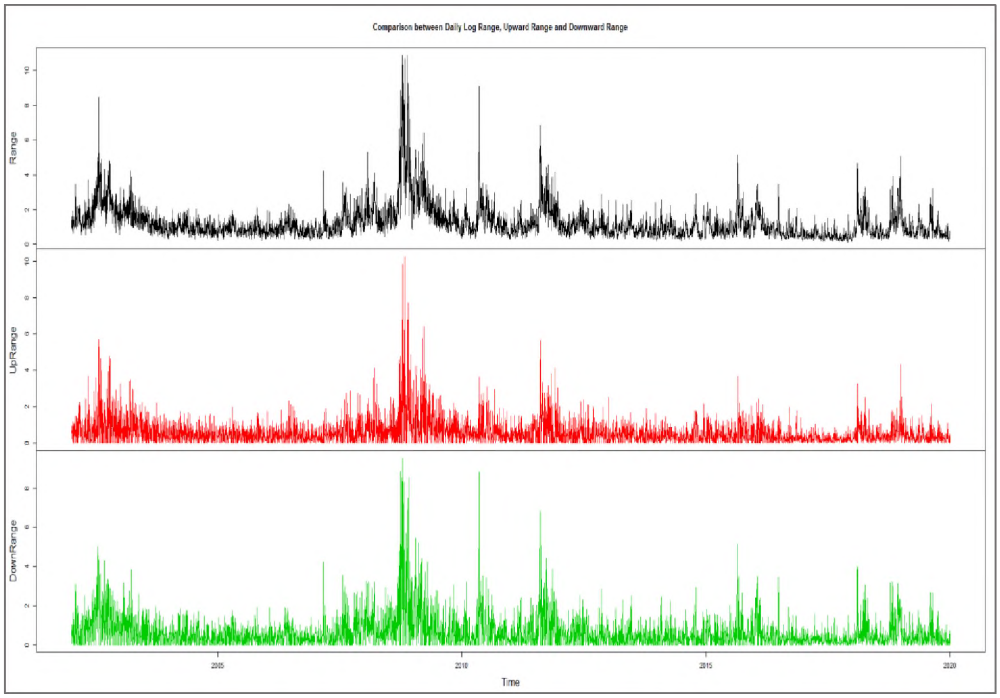

Figure 1: S&P 500 daily price range (black), daily upward range (red) and daily downward range (green) for the period of 01/02/2002: 12/31/2019

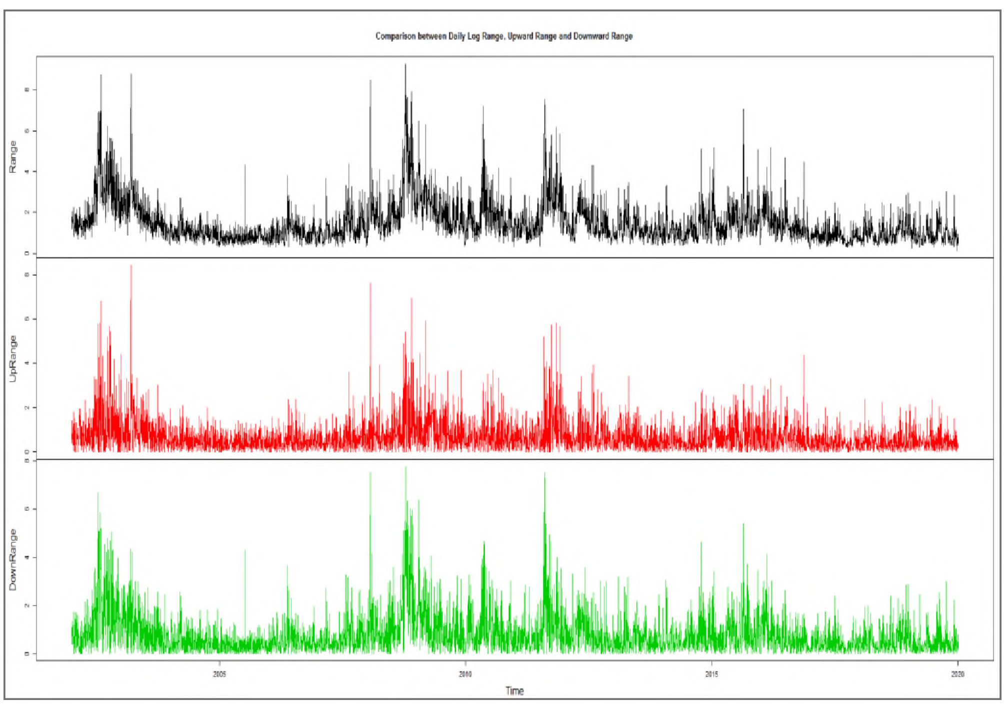

Figure 2: CAC 40 daily price range (black), daily upward range (red) and daily downward range (green) for the period of 01/02/2002: 12/31/2019

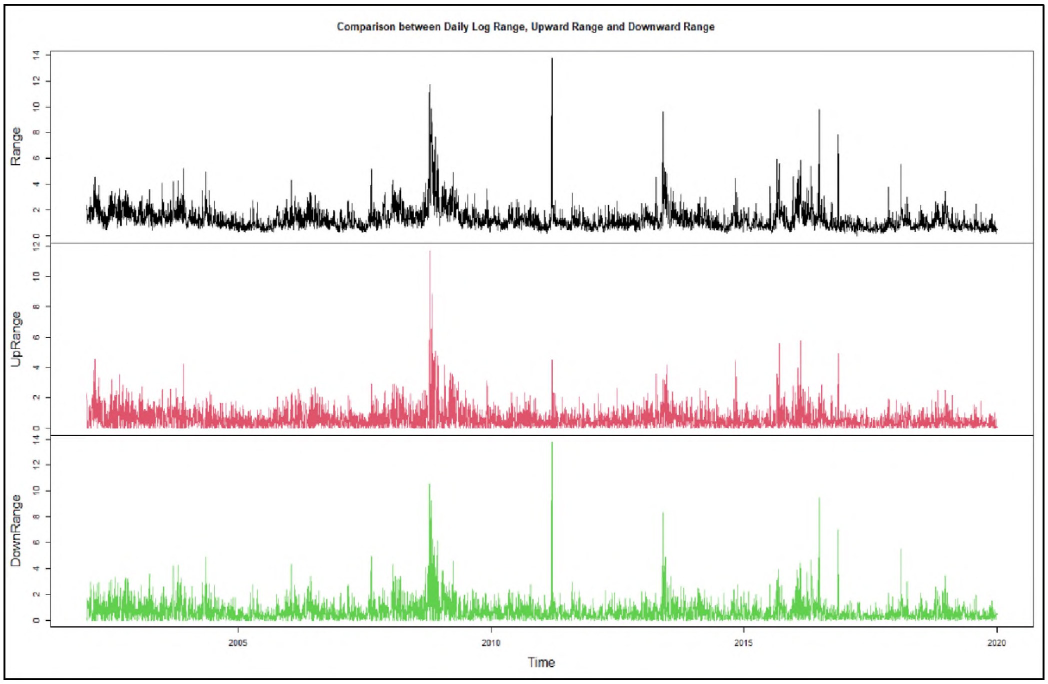

Figure 3: NIKKEI 225 daily price range (black), daily upward range (red) and daily downward range (green) for the period of 01/01/2002-12/31/2019

Time series plots for the daily price range, daily upward range, and daily downward range series of the S&P 500, CAC 40, and NIKKEI 225 are presented in Figures1 through 3, respectively. According the figures, both upward and downward price range data have zeros. This is an important factor that needs to be considered when selecting the appropriate distributions to model the price series separately. In this study, we used exponential distribution to model both the upward and downward price ranges because the support of this distribution includes zero.

6.2. IN-SAMPLE ESTIMATION RESULTS

In this section, we discussed the parameter estimation for the FACARR and compared it with the BEGFACARR-a and BEGFACARR-b models. Model parameters were estimated using the MLE method, as discussed in Section 4, and results are presented in Tables 6 through 8. The Ljung-Box Q test based on 22 lags was considered to check whether the residual series of upward and downward range components over time are random and independent. The Pearson correlation between upward disturbance term and downward disturbance term was calculated and significance was tested. In addition to that, the AIC, AICC and BIC values were also employed as the model selection criteria.

Based on the Ljung Box Q test results for upward (downward) range disturbance term, for the FACARR, BEGFACARR-a and BEGFACARR-b are random and independent for all three stock indices namely S&P 500, CAC 40 and NIKKEI 225. The FACARR and BEGFACARR-a models assume that the upward and downward disturbance terms are independent, while BEGFACARR-b assumes that they are correlated. According to the Pearson correlation and their corresponding p-values (<0.0001) all three models exhibit significant correlation between upward and downward disturbance terms for all the stock indices. Therefore, among the three fitted models BEGFACARR-b satisfied all the model assumptions for all the three stock indices.

Based on the AIC, AICC, and BIC values of the FACARR, BEGFACARR-a, and BEGFACARR-b models for all three stock indices indicate that BEGFACARR-a model performs slightly better than the other two models during upward and downward range components. However, the BEGFACARR-b model fit better to the full range period data than the FACARR and BEGFACARR-a models. Moreover, when compared to the FACARR model, with BEGFACARR-a and BEGFACARR-b processes, the latter models captured the negative relationship between the current conditional mean of upward (downward) range and previous price range data or conditional mean of downward (upward) range. Since we selected the BEGFACARR-b model over the FACARR and

BEGFACARR-a process by considering ability to satisfying all the model assumptions and overall model selection criteria, we compared the BEGFACARR-b model performance with the FACARR model. Tables 9, 10, 11, and 12 summarize the comparison of results between the FACARR and the BEGFACARR-b models, including their performance during the 2007-2009 recession period. Table 9 presents the in-sample comparison between the FACARR model and the BEGFACARR-b model. For the overall range component, the BEGFACARR-b model have higher prediction accuracy when compared to the FACARR model. In certain situations, the FACARR model had lower RMSE, MAE, or both RMSE and MAE values than the BEGFACARR-b process for the upward range and downward range components. Based on the results in Table 8, the proposed BEGFACARR-b model has lower RMSE and MAE when compared to the FACARR model during the recession period. This suggests that the BEGFACARR-b model fits the data from the recession periods better than the FACARR model.

In-sample prediction by the EGFACARR model for the S&P 500, CAC 40 and NIKKEI 225 are given in the Figure 4, Figure 5 and Figure 6.

6.3. OUT-OF-SAMPLE FORECAST

The out of Sample performance from the proposed BEGFACARR-a and BEGFACARR-b models were compared with the FACARR model, and the MAE and the RMSE were used as the forecasting performance evaluation indicators. The model with the smaller forecasting error values indicates that it is relatively better than the other models.

Table 6: Parameter estimated values, residual diagnostic results and model selection criteria of FACARR and BEGFACARR-a and BEGFACARR-b models for S&P 500 index (p-value)

	S&P 500		
	FACARR	BEGFACARR-a	BEGFACARR-b
ω^u	0.0152	0.0061	0.0025
α^u	0.0262	0.0045	0.0079
β^u	0.7810	0.6297	0.4922
γ^u	0.1576	0.1450	0.1409
δ^u		0.1890	0.3158
ω^d	0.0124	0.0058	0.0034
α^d	0.1004	0.1193	0.1252
β^d	0.8499	1.1004	1.2202
γ^d	0.0325	0.0160	0.0163
δ^d		-0.2596	-0.3973
υ			-1.0000
$\varepsilon_t^u \sim i.i.d.$	32.8260 (0.0644)	23.1810 (0.3916)	22.1000 (0.4540)
$\varepsilon_t^d \sim i.i.d.$	26.6620 (0.2244)	28.592 (0.1569)	29.3370 (0.1355)
$cor\left(\varepsilon_t^u, \varepsilon_t^d\right)$	-0.5721 (<0.0001)	-0.5759 (<0.0001)	-0.5778 (<0.0001)
AIC-UPR	3187.6410	**3177.6420**	3179.6690
AIC-DWNR	3906.1750	**3898.9990**	3900.5730
AIC-RANGE	7093.8160	7076.6410	**5211.4510**
AICC-UPR	3187.6510	**3177.6570**	3179.6830
AICC-DWNR	3906.1840	**3899.0130**	3900.5880
AICC-RANGE	7093.8500	7076.6930	**5211.5130**
BIC-UPR	3213.0870	**3209.4500**	3211.4760
BIC-DWNR	3931.6200	**3930.8060**	3932.3810
BIC-RANGE	7144.7080	7140.2560	**5281.4280**

Table 7: Parameter estimated values residual diagnostic results and model selection criteria of FACARR and BEGFACARR-a and BEGFACARR-b models for CAC 40 index (p-value)

	CAC 40		
	FACARR	BEGFACARR-a	BEGFACARR-b
ω^u	0.0269	0.0325	0.0324
α^u	0.0371	0.0341	0.0618
β^u	0.7680	0.2485	0.4631
γ^u	0.1381	0.1305	0.1689
δ^u		0.4602	0.2136
ω^d	0.0133	0.0326	0.0122
α^d	0.1116	0.1359	0.1021
β^d	0.8213	1.1448	0.8903
γ^d	0.0573	0.0360	0.0373
δ^d		-0.7421	-0.0450
υ			-1.0000
$\varepsilon_t^u \sim i.i.d.$	30.6070 (0.1044)	32.6820 (0.0665)	33.6310 (0.0535)
$\varepsilon_t^d \sim i.i.d.$	27.5340 (0.1916)	30.7230 (0.1019)	29.8980 (0.1210)
$cor\left(\varepsilon_t^u, \varepsilon_t^d\right)$	-0.5239 (<0.0001)	-0.5297 (<0.0001)	-0.5298 (<0.0000)
AIC-UPR	4951.2570	**4938.7700**	4946.6420
AIC-DWNR	6032.7320	**6030.7500**	6037.0630
AIC-RANGE	10983.9900	10969.5200	**9514.4300**
AICC-UPR	4951.2670	**4938.7840**	4946.6560
AICC-DWNR	6032.7420	**6030.7640**	6037.0770
AICC-RANGE	10984.0200	10969.5700	**9516.4900**
BIC-UPR	4976.7660	**4970.6570**	4978.5280
BIC-DWNR	**6058.2410**	6062.6360	6068.9490
BIC-RANGE	11035.0100	11033.2900	**9586.8600**

Table 8: Parameter estimated values, residual diagnostic results, and model selection criteria of FACARR and BEGFACARR-a and BEGFACARR-b models for NIKKEI 225 index (p-value)

	NIKKEI 225		
	FACARR	BEGFACARR-a	BEGFACARR-b
ω^u	0.0319	0.0133	0.0313
α^u	0.0758	0.0649	0.0579
β^u	0.6973	0.2297	1.7412
γ^u	0.1613	0.1690	0.1269
δ^u		0.4552	-0.8795
ω^d	0.0196	0.0122	0.0394
α^d	0.1012	0.1322	0.1549
β^d	0.8198	1.2345	-0.1289
γ^d	0.0561	0.0593	0.0067
δ^d		-0.4808	0.9686
υ			-1.0000
$\varepsilon_t^u \sim i.i.d.$	18.1580 (0.6966)	15.2400 (0.8518)	18.7420 (0.6612)
$\varepsilon_t^d \sim i.i.d.$	16.4550 (0.7927)	13.2460 (0.9261)	16.8580 (0.7712)
$cor\left(\varepsilon_t^u, \varepsilon_t^d\right)$	-0.4826 (<0.0001)	-0.4858 (<0.0001)	-0.4917 (<0.0001)
AIC-UPR	3917.9450	**3902.9700**	3912.8760
AIC-DWNR	4768.6140	**4767.1500**	4768.1290
AIC-RANGE	8686.5590	8670.1180	**7047.8510**
AICC-UPR	3917.9540	**3902.9850**	3912.8910
AICC-DWNR	4768.6230	**4767.1630**	4768.1430
AICC-RANGE	8686.5930	8670.1710	**7047.9140**
BIC-UPR	3943.2840	**3934.6450**	3944.5510
BIC-DWNR	4793.9540	**4798.8230**	4799.8040
BIC-RANGE	8737.2380	8733.4680	**7117.5350**

Table 9: In-sample comparison between FACARR, and EGFACARR-b models for S&P 500, CAC 40 and NIKKEI 225

Index	Model	Upward Range		Downward Range		Range	
		MAE	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
S&P 500	FACARR	0.4156	0.6145	0.4953	**0.7405**	0.4011	0.6038
	BEGFACARR-b	**0.4128**	**0.6142**	**0.4931**	0.7417	**0.3947**	**0.6037**
CAC40	FACARR	**0.4691**	0.6758	**0.5340**	**0.7584**	0.4680	0.6861
	BEGFACARR-b	0.4694	**0.6733**	0.5352	0.7609	**0.4661**	**0.6823**
NIKKEI225	FACARR	0.4413	**0.6236**	**0.5117**	**0.7866**	0.4572	0.7051
	BEGFACARR-b	**0.4393**	0.6267	0.5135	0.7884	**0.4497**	**0.7040**

Table 10: In-sample recession period comparison between FACARR, and BEGFACARR-b models for S&P 500, CAC 40 and NIKKEI 225

Index	Model	Recession Period	
		MAE	RMSE
S&P500	FACARR	0.5582	0.7210
	BEGFACARR-b	**0.5477**	**0.7195**
CAC40	FACARR	1.0340	0.6958
	BEGFACARR-b	**1.0235**	**0.6825**
NIKKEI225	FACARR	0.4977	0.6170
	BEGFACARR-b	**0.4862**	**0.6082**

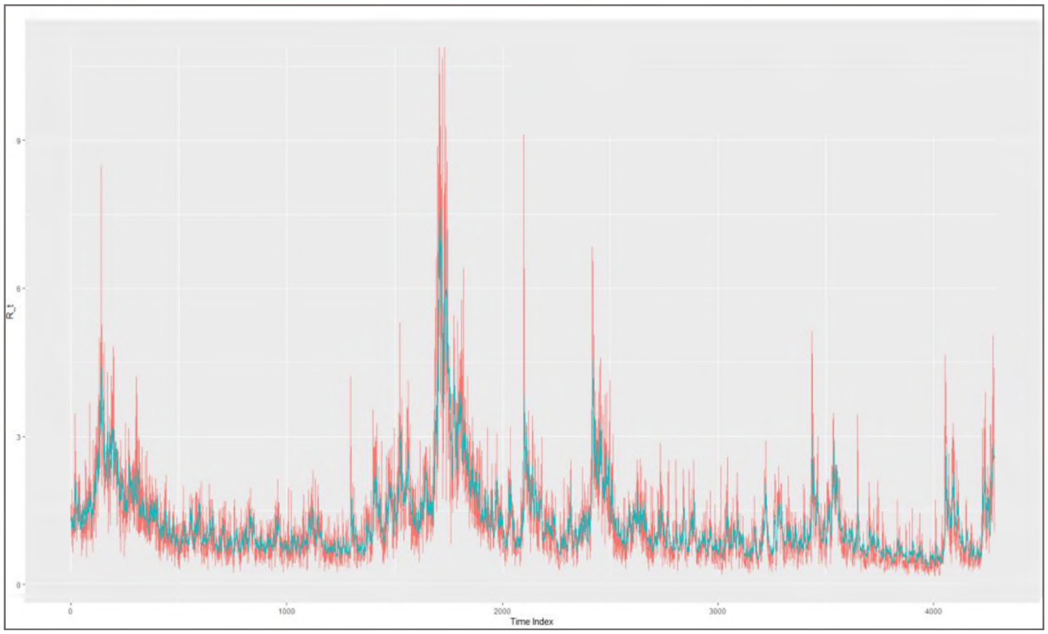

Figure 4: In-sample perdition of fitted BEGFACARR-b model (green) for the S&P 500 price range (red) index

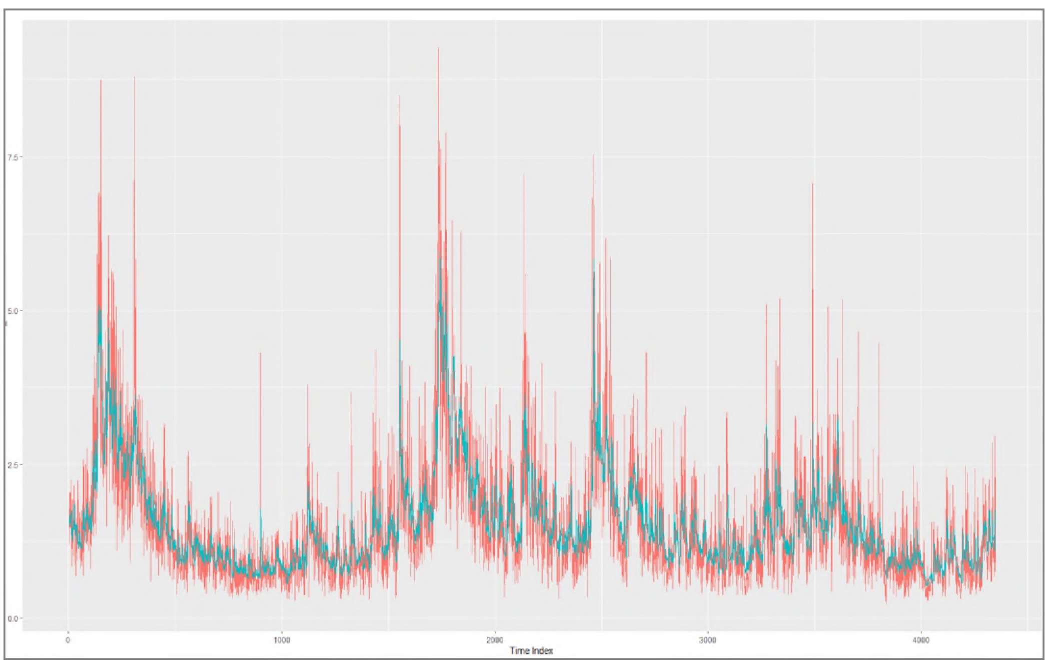

Figure 5: In-sample perdition of fitted BEGFACARR-b model (green) for the CAC 40 (red) index

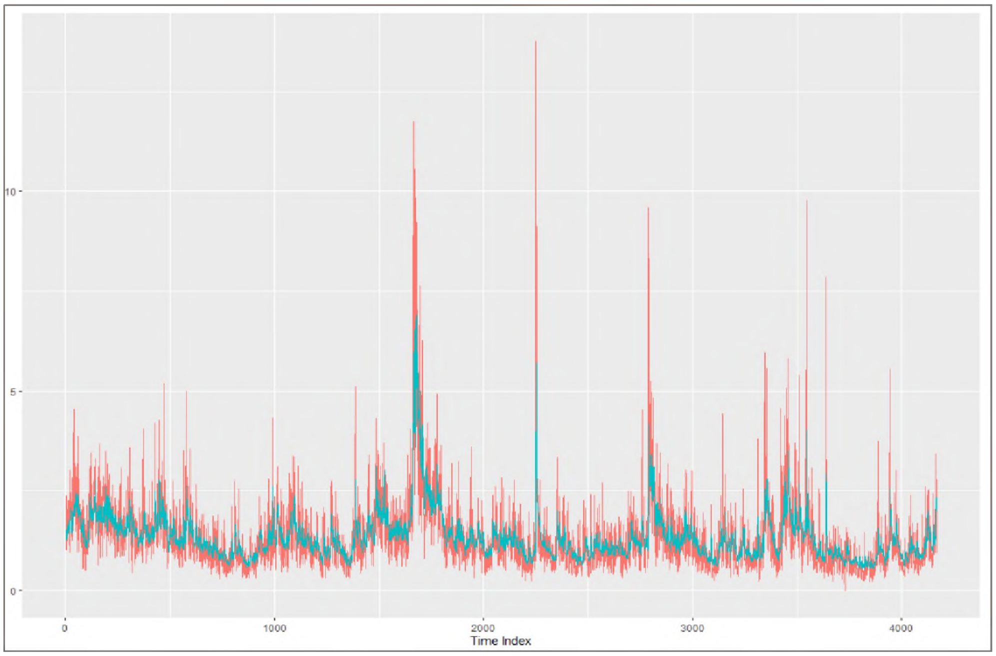

Figure 6: In-sample perdition of fitted BEGFACARR-b model (green) for the NIKKEI 225 (red) index

For out of sample prediction, a recursive window estimation method was performed. Table 9 presents the out of sample forecasting results. Based on the forecasting errors, the BEGFACARR-b model have lower MAE and RMSE values for all the three stock indices when compared to those from the FACARR model. Moreover, Diebold & Marino's (1995) test is used to check for a significant difference between the BEGFACARR-b model forecasting accuracy and that of the FACARR model. If a significant difference exists, then we checked the BEGFACARR-b model against the FACARR model for accuracy in forecasting future price range data. The DM test statistics and corresponding p-values suggested with 90% confidence that the proposed BEGFACARR-b model is more accurate than FACARR model in forecasting future values for CAC 40 and NIKKEEI 225 indices. However, for the S&P 500 stock index, there is no significant difference between the BEGFACARR-b and the FACARR forecasting accuracy. The out of sample data and the forecasted values from the BEGFACARR-b model are presented in the Figures 7 through 10.

Table 11: Out of sample comparison between FACARR, BEGFACARR-a and BEGFACARR-b models for S&P 500, CAC 40 and NIKKEI 225

Stock Index	Model	Accuracy	
		MAE	RMSE
S&P500	FACARR	0.3136	0.4071
	BEGFACARR-a	0.3102	0.4048
	BEGFACARR-b	**0.3050**	**0.4018**
CAC40	FACARR	0.3358	0.4589
	BEGFACARR-a	0.3305	0.4545
	BEGFACARR-b	**0.3295**	**0.4540**
NIKKEI225	FACARR	0.2943	0.3719
	BEGFACARR-a	0.2858	0.3615
	BEGFACARR-b	**0.2786**	**0.3566**

Table 12: DM test statistics results

Stock Index	Alternative Hypothesis	Test Statistics (p-value)
S&P 500	BEGFACARR-b model and FACARR model have different forecast accuray	-0.5694 (0.5691)
CAC40	Forecast BEGFACARR-b model is more accurate than that of the FACARR model	-1.3112 (0.0949)
NIKKEI 225	Forecast BEGFACARR-b model is more accurate than that of the FACARR model	-3.5957 (0.0002)

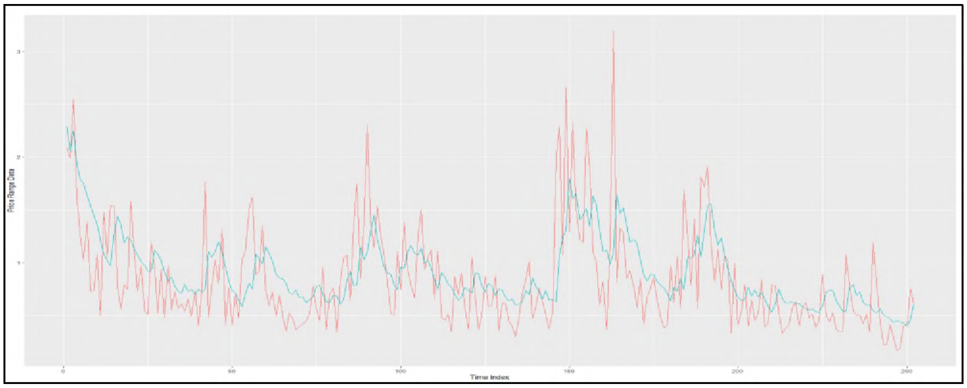

Figure 7: Out-of-sample forecasted values by BEGFACARR-b (green) for the S&P 500 (red) index

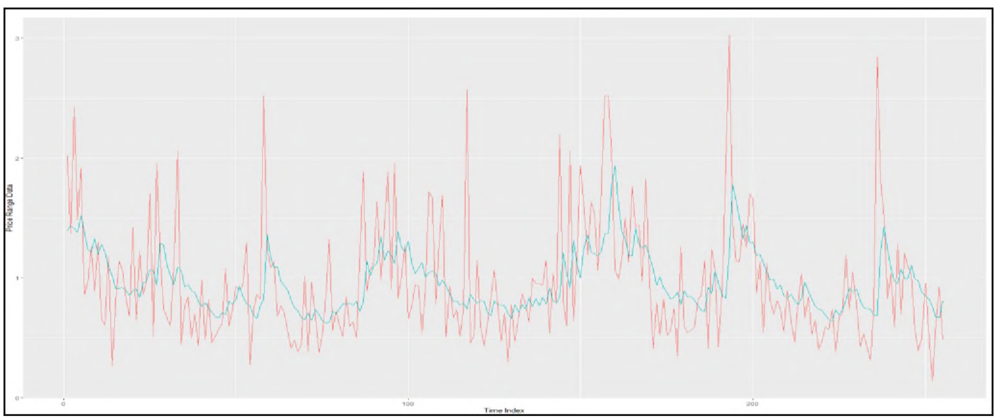

Figure 8: Out-of-sample forecasted values by BEGFACARR-b (green) for the CAC 40 (red) index

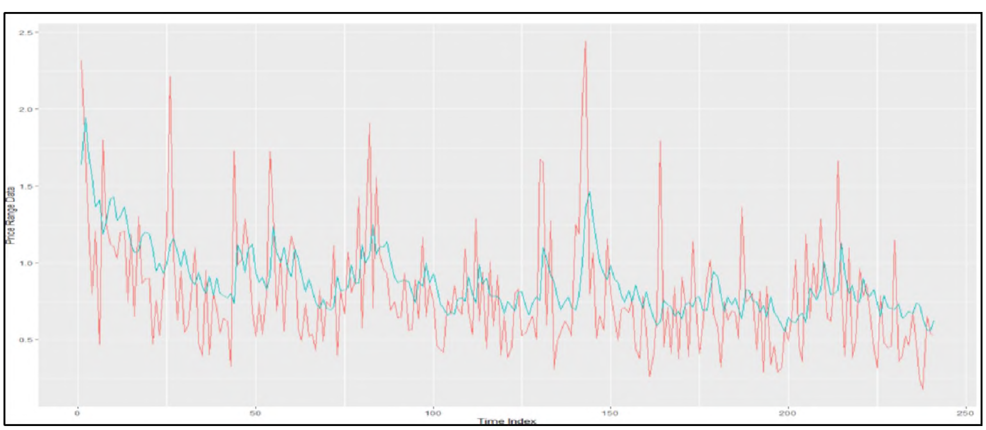

Figure 9: Out-of-sample forecasted values by BEGFACARR-b (green) for the NIKKEI 225 (red) index

7. CONCLUSIONS

In this paper, we proposed the GFACARR model, which is a bivariate CARR type model, to accommodate asymmetric propagation of upward and downward ranges, while accommodating a complete dynamic feedback mechanism between these two components and their conditional means. The GFACARR process uses previous downward (upward) price ranges and conditional mean downward (upward) range values to model the conditional mean upward (downward) range. Furthermore, the GFACARR model is capable of modeling the negative relationship between upward and downward range data, which could not be achieved using the FACARR model. In this study, we proved the weak stationarity conditions for the proposed GFACARR model. In addition to that, we considered the two different scenarios based on the upward and downward range component disturbance terms. If the two terms were independent, we used the Bivariate Exponential GFACARR type a (BEGFACARR-a) model, and if the two components correlated, we employed the Bivariate Exponential GFACARR type b (BEGFACARR-b) model to perform the analysis. According to the simulation study, the MLE method can be employed to estimate model parameters in the BEGFACARR-a and the BEGFACARR-b models with greater accuracy. The performance of the proposed model was gauged through an empirical study by using three stocks indices, namely S&P 500, CAC 40, and NIKKEI 225. Since the BEGFACARR-b model was satisfied all the model assumptions and provided the lowest AIC, AICC and BIC values for price range period when compared to BEGFACARR-a, it was selected to model the in-sample data. Then the BEGFACARR-b model performance was compared to the FACARR model. According to the performance evaluation indicators we employed, the BEGFACARR-b model has relatively lower errors when predicting the overall range, and it performs better at predicting ranges during recession periods. However, in some non-recession situations, the FACARR has slightly better performance in in-sample predictions, than the BEGFACARR-b model, with respect to predicting upward or downward ranges. Smaller forecasting error values were obtained for the out of sample price range data from the BEGFACARR-b model for two indices studied namely CAC 40 and NIKKEI 225, and the performance of the FACARR and BRGFACARR-b were not statistically significant for the S&P 500 index. Overall, the results indicated that BEGFACARR-b model beats the FACARR model for out of sample forecasting.

REFERENCES

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. Journal of Finance, 57(3), 1047–1092.

Beckers, Stan (1983). Variance of Security Price Return Based on High, Low and Closing Prices. Journal of Business 56, 97–112.

Black, F., (1976). The pricing of commodity contracts, Journal of Financial Economics, 3, 167-179.

Bollerslev, Tim (1986). Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics 31, 307–327.

Brandt, M.W. and C.S. Jones (2006), “Volatility Forecasting with Range-Based EGARCH Models”. Journal of Business and Economic Statistics, 24, 470-486.

Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model. Journal of Money, Credit and Banking, Vol. 37, pp. 561-582.

Chou, R. (2006). Modeling the asymmetry of stock movements using price ranges. Advances in Econometrics, 20, 231–257.

Chou, R., Chou, H., & Liu, N. (2015). Range volatility: A review of models and empirical studies. In C. -F. Lee, & J. C. Lee (Eds.), Handbook of Financial Econometrics and Statistics. Berlin, Germany: Springer, pp. 2029–2050.

Chou, R., & Liu, N. (2010). The economic value of volatility timing using a range-based volatility model. Journal of Economic Dynamics and Control, 34(11), 2288–2301.

Chou, H., & Wang, D. (2014). Estimation of tail-related value-at-risk measures: Range-based extreme value approach. Quantitative Finance, 14(2), 293–304.

Engle, R.F., (2002). Dynamic conditional correlation-A simple class of multivariate GARCH models. Journal of Business & Economic Statistics, 20, 339-350.

Engle, R.F., (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation. Econometrica 50, 987–1008.

Engle, R.F., & Russell, J.R., (1998). Autoregressive Conditional Duration: A New Model for Irregular Spaced Transaction Data. Econometrica 66, 1127–1162.

Engle, R.F., & Ng, V.K., (1993) Measuring and Testing the Impact of News on volatility. The Journal of Finance Vol. 48, No. 5 (Dec. 1993), pp. 1749-1778.

Engle, R.F. and K. Sheppard (2001), "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH", NBER Working Paper 8554, National Bureau of Economic Research, Cambridge, MA.

Fernandes, M., Mota, B., Rocha, G., (2005). A multivariate conditional autoregressive range model. Economics Letters, 86, 435-440.

Glosten, L., Jagannathan, R., and Runkle, D. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779-1801.

Gumbel, E. (1960). Bivariate Exponential Distributions. *Journal of the American Statistical Association, 55*(292), 698-707.

Jeantheau, T. (1998). Strong Consistency of Estimators for Multivariate Arch Models. *Econometric Theory, 14*(1), 70-86.

Kunitomo, N. (1992). Improving the Parkinson method of estimating security price volatilities. Journal of Business, 65(2), 295–302.

Lamoureux, C. G., & Lastrapes, W. D., (1990). Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. The Journal of Finance, 45, 221-229.

McNeil, A. J., & Fery R., (2000). Estimation of tail-risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance, 7, 271-300.

Miao, D. W., Wu, C., & Su, Y. (2013). Regime-switching in volatility and correlation structure using range-based methods with Markov switching. Economic Modelling, 3, 87–93.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59, 347–370.

Nelson, D.B. and Cao, C.Q. (1992) 'Inequality Constraints in the Univariate GARCH Model', Journal of Business and Economic Statistics, Vol. 10, No. 3, pp. 229-235.

Parkinson, Michael (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. Journal of Business 53, 61–65.

Sentana, E., (1995). Quadratic ARCH models. Review of Economics Studies 62: 639-661.

Taylor, S. (1994) "Modelling stochastic volatility: A review and comparative study" *Mathematical Finance 4,* 183–204.

Tsay, R. S. (2005). *Analysis of financial time series*. Hoboken, NJ: Wiley-Inter science. ISBN: 978-0-471-69074-0

Xie H. (2018). Financial volatility modeling: The feedback asymmetric conditional autoregressive range model. *Journal of Forecasting*. 38:11–28.

Xie, H., & Wang, S. Y. (2013). A new approach to model financial markets. *Journal of Systems Science and Complexity, 26(3),* 432–440.

Xie, H., & Wu, X. (2017). A conditional autoregressive range model with gamma distribution for financial volatility modelling. *Economic Modelling*, *64*, 349–356.

APPENDIX

Proposition 1: Let Δ_1 and Δ_2 be the eigenvalue of the matrix $(A+B)_{2\times2}$. If Δ_1 and Δ_2 satisfy $|\Delta_i| < 1 \forall i$, then $\det[I-(A+B)z]$ has roots z_1 and z_2 such that $|z_i| > 1, \forall i$.

Proof of Proposition 1:

$$\vec{\lambda_t} = \vec{\omega} + A\overrightarrow{R_{t-1}} + B\overrightarrow{\lambda_{t-1}},$$

$$R_t - E(R_t \mid \mathrm{F}_{t-1}) = R_t - \lambda_t = \eta_t,$$

$$\vec{R_t} = \vec{\omega} + (A+B)\overrightarrow{R_{t-1}} + \vec{\eta_t} - B\overrightarrow{\eta_{t-1}},$$

The Eigenvalues of $(A+B)$ are:

$$\det[(A+B)-\Delta I] = \det\left[\begin{pmatrix} \alpha^u+\beta^u & \gamma^u+\delta^u \\ \gamma^d+\delta^d & \alpha^d+\beta^d \end{pmatrix} - \begin{pmatrix} \Delta & 0 \\ 0 & \Delta \end{pmatrix}\right],$$

$$\det[(A+B)-\Delta I] = \left[(\alpha^u+\beta^u-\Delta)(\alpha^d+\beta^d-\Delta)-(\gamma^u+\delta^u)(\gamma^d+\delta^d)\right],$$

$$f(\Delta) = \Delta^2 - (\alpha^u+\beta^u+\alpha^d+\beta^d)\Delta + (\alpha^u+\beta^u)(\alpha^d+\beta^d) - (\gamma^u+\delta^u)(\gamma^d+\delta^d) = 0,$$

The solutions for the $f(\Delta)=0$, are Δ_1 and Δ_2, the eigenvalues of $(A+B)_{2\times2}$ and:

$$\Delta_1 = \frac{(\alpha^u+\beta^u+\alpha^d+\beta^d) + \sqrt{\left[(\alpha^u+\beta^u+\alpha^d+\beta^d)^2 - 4\left[(\alpha^u+\beta^u)(\alpha^d+\beta^d)-(\gamma^u+\delta^u)(\gamma^d+\delta^d)\right]\right]}}{2},$$

$$\Delta_1 = \frac{(\alpha^u+\beta^u+\alpha^d+\beta^d) - \sqrt{\left[(\alpha^u+\beta^u+\alpha^d+\beta^d)^2 - 4\left[(\alpha^u+\beta^u)(\alpha^d+\beta^d)-(\gamma^u+\delta^u)(\gamma^d+\delta^d)\right]\right]}}{2},$$

Let $Q(L)\vec{R_t} = \vec{\omega} + M(L)\vec{\eta_t}$, be a bivariate ARMA (1, 1) process similar to the model (3.4) where L be the lag operator and $Q(z) = [I-(A+B)z]$.

Let z_1 and z_2 be the roots of the equation $\det\left[Q(z)\right]=0,$ such that $\det\left[Q(z_1)\right]=\det\left[Q(z_2)\right]=0.$

Then,

$$z_1=\frac{\left(\alpha^u+\beta^u+\alpha^d+\beta^d\right)+\sqrt{\left[\left(\alpha^u+\beta^u+\alpha^d+\beta^d\right)^2-4\left[\left(\alpha^u+\beta^u\right)\left(\alpha^d+\beta^d\right)-\left(\gamma^u+\delta^u\right)\left(\gamma^d+\delta^d\right)\right]\right]}}{2\left[\left(\alpha^u+\beta^u\right)\left(\alpha^d+\beta^d\right)-\left(\gamma^u+\delta^u\right)\left(\gamma^d+\delta^d\right)\right]}=\frac{\Delta_1}{\Delta_1\Delta_2}=1/\Delta_2,$$

$$z_2=\frac{\left(\alpha^u+\beta^u+\alpha^d+\beta^d\right)-\sqrt{\left[\left(\alpha^u+\beta^u+\alpha^d+\beta^d\right)^2-4\left[\left(\alpha^u+\beta^u\right)\left(\alpha^d+\beta^d\right)-\left(\gamma^u+\delta^u\right)\left(\gamma^d+\delta^d\right)\right]\right]}}{2\left[\left(\alpha^u+\beta^u\right)\left(\alpha^d+\beta^d\right)-\left(\gamma^u+\delta^u\right)\left(\gamma^d+\delta^d\right)\right]}=\frac{\Delta_2}{\Delta_1\Delta_2}=1/\Delta_1.$$

Here $|\Delta_i|<1\ \forall i$. Hence roots of $\det\left[Q(z)\right]=0,$ are z_1 and z_2 such that $|z_1|=\dfrac{1}{|\Delta_2|}>1$ and $|z_2|=\dfrac{1}{|\Delta_1|}>1.$

Theorem 1: Let $\overrightarrow{\lambda_t}=\vec{\omega}+A\overrightarrow{R_{t-1}}+B\overrightarrow{\lambda_{t-1}}$ be a GFACARR process defined in (3.1) - (3.4). If all the eigenvalues of $(A+B)_{2\times 2}$, namely Δ_1 and Δ_2 are such that $|\Delta_i|<1\ \forall i$, then the GFACARR model for $\overrightarrow{R_t}$ is weakly stationary.

Proof of Theorem 1:

Let consider the model in equation (3.3),

$$\begin{pmatrix}\lambda_t^u\\ \lambda_t^d\end{pmatrix}=\begin{pmatrix}\omega^u\\ \omega^d\end{pmatrix}+\begin{pmatrix}\alpha^u & \gamma^u\\ \gamma^d & \alpha^d\end{pmatrix}\begin{pmatrix}R_{t-1}^u\\ R_{t-1}^d\end{pmatrix}+\begin{pmatrix}\beta^u & \delta^u\\ \delta^d & \beta^d\end{pmatrix}\begin{pmatrix}\lambda_{t-1}^u\\ \lambda_{t-1}^d\end{pmatrix},$$

$$\begin{pmatrix}\lambda_t^u\\ \lambda_t^d\end{pmatrix}=\begin{pmatrix}\omega^u\\ \omega^d\end{pmatrix}+\begin{pmatrix}\alpha^u & \gamma^u\\ \gamma^d & \alpha^d\end{pmatrix}\begin{pmatrix}\lambda_{t-1}^u\varepsilon_{t-1}^u\\ \lambda_{t-1}^d\varepsilon_{t-1}^d\end{pmatrix}+\begin{pmatrix}\beta^u & \delta^u\\ \delta^d & \beta^d\end{pmatrix}\begin{pmatrix}\lambda_{t-1}^u\\ \lambda_{t-1}^d\end{pmatrix}. \tag{A.1}$$

The expression in (A.1) can be reparametrized in a manner similar to Jeantheau (1998) and can be presented in the following matrix form:

$$V_{t,\Phi}=\left(A(\varepsilon_{t-1})+B\right)V_{t-1,\Phi}+W. \tag{A.2}$$

Here,

$$V'_{t,\Phi} = \left(\lambda_t^u(\Phi) \quad \lambda_t^d(\Phi)\right),\ W' = \left(\omega^u \quad \omega^d\right),$$

$$A(\varepsilon_{t-1}) + B = \begin{pmatrix} \alpha^u \varepsilon_{t-1}^u + \beta^u & \gamma^u \varepsilon_{t-1}^d + \delta^u \\ \gamma^d \varepsilon_{t-1}^u + \delta^d & \alpha^d \varepsilon_{t-1}^d + \beta^d \end{pmatrix}.$$

Finally, we can derive above $V_{t,\Phi}$ with $\xi_t' = \varepsilon'_{t-1} = \left(\varepsilon_{t-1}^u \quad \varepsilon_{t-1}^d\right)$ and

$$V_{t,\Phi} = F(\xi_t)V_{t-1,\Phi} + W,$$
$$V_{t,\Phi} = F(\xi_t)\left(F(\xi_{t-1})V_{t-2,\Phi} + W\right) + W = F(\xi_t)F(\xi_{t-1})V_{t-2,\Phi} + W + F(\xi_t)W,$$
$$\vdots$$
$$V_{t,\Phi} = F(\xi_t)F(\xi_{t-1})F(\xi_{t-2})...F(\xi_{t-k+1})V_{t-k,\Phi} + W + \sum_{i=1}^{k-1} F(\xi_t)F(\xi_{t-1})....F(\xi_{t-i+1})W.$$

First, we have to prove that $\sum_{i=1}^{k-1} F(\xi_t)F(\xi_{t-1})....F(\xi_{t-i+1})W \xrightarrow{a.s} L^1$, when $k \to \infty$.

Let consider $E\left(F(\xi_t)F(\xi_{t-1})....F(\xi_{t-i+1})\right)W = F^i W$, here $F = A + B = UDU^{-1}$ based on the spectral decomposition of F with $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ where λ_1 and λ_2 are the eigenvalues such that $|\lambda_i| < 1$ and $F^i = U\begin{pmatrix} \lambda_1^i & 0 \\ 0 & \lambda_2^i \end{pmatrix}U^{-1} \to 0$.

Therefore, $\sum_{i=1}^{k-1} F(\xi_t)F(\xi_{t-1})....F(\xi_{t-i+1})W \xrightarrow{a.s} L^1$.

Now we have to prove that $F(\xi_t)F(\xi_{t-1})F(\xi_{t-2})...F(\xi_{t-k+1})V_{t-k,\Phi} \xrightarrow{a.s} L^1$ when $k \to \infty$. Let assume that $k \to \infty$ and $V_{t-k,\Phi} \to C$.

Then, $E\left(F(\xi_t)F(\xi_{t-1})F(\xi_{t-2})...F(\xi_{t-k+1})V_{t-k,\Phi}\right) = F^k E\left(V_{t-k,\Phi}\right) = F^k C \xrightarrow{k\to\infty} 0.$

Finally, we can conclude that $V_{t,\Phi} \xrightarrow{a.s} L^1$.

In another independent study Bollerslev and Engle (1993) proved that vector GARCH (*p*, *q*) process is covariance stationary if and only if all the roots of the characteristic polynomial $\det\left[I - A(z) - B(z)\right] = 0$ lie outside the unit circle. By applying the proposition 1 to the GFACARR process defined in (3.1) -(3.4) we can show that the $\det\left[I - A(z) - B(z)\right] = 0$ has roots outside the unit circle. Combining the above result with results Bollerslev and Engle (1993) we can conclude that GFACARR is weekly stationary.

Therefore, we can conclude that if the eigenvalues of $(A+B)_{2\times 2}$, Δ_1 and Δ_2 such that $|\Delta_i| < 1 \ \forall i$, then the GFACARR is weak stationary.

III. THRESHOLD ASYMMETRIC CONATIONAL AUTOREGRESSIVE RANGE (TACARR) MODEL

ABSTRACT

This paper proposes a Threshold Asymmetric Conditional Autoregressive Range (TACARR) formulation for modeling the daily price ranges of financial assets. The disturbance term of the range process is assumed to follow a threshold distribution with positive support. The study assumes that the conditional expected range process switches between two market regimes. The two market regimes are named as the upward market and the downward market. A self-adjusting threshold component that is driven by the past financial information determines the current market regime. The proposed model is able to capture aspects such as asymmetry and heteroscedastic behavior in the financial markets. This model addresses several inefficiencies found in existing price range models including the Conditional Autoregressive Range (CARR), Asymmetric CARR (ACARR), Feedback ACARR (FACARR) and Threshold Autoregressive Range (TARR) models. Parameters of the model are estimated using the Maximum Likelihood Estimation (MLE). The simulation studies show that the MLE method performs well and it estimate the TACARR model parameters with high accuracy. We assessed the performance of the TACARR model using IBM index data and results show that the proposed TACARR model was useful for in-sample prediction and out-of-sample forecasting volatility.

Key Words: Volatility Modeling, Asymmetric Volatility, CARR Models, Threshold Variables.

1. INTRODUCTION

Modelling economic volatility is indispensable to better understanding the dynamics of financial markets. Financial volatility of asset prices has been discussed extensively in the financial and econometric literature over past few decades. Engle (1982) proposed the Autoregressive Conditional Heteroscedasticity (ARCH) model to address the complexities of time-varying volatility and volatility clustering in financial time series. In the ARCH formulation, the conditional volatility is modeled as a function of past returns. Bollerslev (1986) proposed the Generalized Autoregressive Conditional Heteroscedasticity (GARCH), which remains one of the most popular volatility models up to date. The GARCH model is an extension of ARCH formulation, and it models the conditional volatility as a function of lagged squared returns, as well as past conditional variances. Since both models aforementioned focus on modeling price returns, they can be identified as examples of return-based volatility models.

In many financial time series applications, standard deviation is the most common measure of stock return volatility since it not only calculates the dispersion of returns, but also summarizes the probability of seeing extreme values in returns. Since the origination of the concept of volatility, researchers have sought alternative measures of measuring it. Parkinson (1980) argued that volatility measures could be calculated using the daily high, daily low, and opening prices of a stock in addition to the traditional closing prices. Parkinson concluded that the range-based method was far superior to the standard methods based on returns. Beckers (1983) tested the validity of different volatility estimators. The study showed that using the range of a stock price was better than using the close-to-close changes. Kunitomo (1992), improved the Parkinson’s original result and proposed a new range-based estimator which is ten times more efficient than the standard volatility estimator. In another study, Alizadeh, Brandt and Diebold (2002) proved that the range-based volatility estimators are highly efficient compared to the classical volatility proxies based on log absolute returns or squared returns.

Some scholars focused on the alternative approach to modeling volatility and developed theoretical frameworks for range-based models, along with comprehensive empirical examples. For example, the works of Chou (2005), Chou (2006), Brandt and Jones (2006), and Chou and Liu (2010). Chou (2005) introduced the Conditional Auto Regressive Range (CARR) model as a special case of Autoregressive Conditional Duration (ACD) model of Engle (1998). The CARR is employed to model price volatility of an asset by considering range of the log prices for a given fixed time interval. Formulation of the CARR model is similar to that of the standard GARCH volatility model. One distinction between the two models is that the GARCH model uses rate of return as its volatility measure, while the CARR model uses the range as its volatility measure. The CARR model proposed by Chou is a simple, but an efficient, tool to analyze the volatility clustering property compared to the GARCH model. Chou showed this empirically via an out-of-sample forecasting of S&P 500 data. Brandt and Jones (2006) integrated the properties of exponential GARCH (Nelson, 1991) with daily log range data and proposed a ranged-based Exponential GARCH model. This model has a simple framework, but it is an effective tool for capturing the important characteristics that are present in stock return data such as clustering, negative correlation, and log normality.

Extensive modifications to the CARR model include works of Chiang, Chou and Wang (2016), who suggested the application of the Lognormal Log CARR model in the outlier detection process. They showed that the proposed method could effectively detect outliers. One major advantage of using a Log CARR model is that it allows the relaxing of positivity restrictions on the parameters when calculating conditional expectation. Xie and Wu (2017), explained the disturbance term in the CARR model using the gamma distribution (GCARR) and showed that the GCARR outperformed Weibull CARR (WCARR) model in its forecasting ability through an empirical study.

The asymmetric volatility, which is a key phenomenon in financial data, suggested that conditional volatilities show higher fluctuations during downward trends than during upward trends. The CARR model proposed by Chou (2005), used range as the measure of price volatility. The study treated maximum price and minimum price symmetrically. However, in the same study, he suggested the CARRX models (CARRX-a, and CARRX-b) including exogenous variables such as (a) lagged return and (b) lagged absolute returns in the conditional mean equation. The purpose of this incorporation was to model one form of asymmetry, the leverage effect of Black and Nelson (1991). Chou (2006) presented the Asymmetric CARR (ACARR) model in which both upward and downward price ranges were treated separately. The upward range is defined as the difference between the maximum price and the opening price, whereas the downward range is defined as the difference between the opening price and the minimum price, all observed within a trading day. These definitions can be extended to periods beyond a day in a similar manner. The ACARR model was extended to the ACARRX model by including exogenous variables such as trading volume (Lamourex and Lastrapes, 1990), lag returns (Black, 1976; Nelson, 1990), or a seasonal factor to count leverage effect. The FACARR model was proposed by Xie (2018), which is a more practical extension of the ACARR model. In addition to the asymmetric nature price ranges, Xie allowed the conditional mean upward (downward) range to be modeled by incorporating lagged downward (upward) ranges into each sub-model. Extensive empirical studies showed that the proposed FACARR performed significantly better than the ACARR in both in-sample and out-of-sample forecasting.

All the above models capture the asymmetry in price range data either by introducing a leverage variable or treating upward and downward price range series separately. Another popular approach to examine the asymmetric behavior in volatility is the use of threshold component. In general, a threshold is introduced to time series process to examine the behavioral changes according to the different cut off points. The initial idea of threshold models in time series analysis was introduced by Tong (1978). Tong pointed out the limitations in the linear Gaussian time series models and emphasized the advantage of using nonlinear time series models. He also proposed a nonlinear threshold autoregressive (TAR) model. Tsay (1989), introduced a simple but effective method for testing and modeling procedures for TAR models. Threshold ARCH (TARCH) proposed by Zakoian (1991) model the standard deviation conditional on the sign of the previous time periods' returns. Zakoian (1994) improved the existing TARCH model by incorporating the lagged conditional standard deviation and named it as the Threshold GARCH (TGARCH). The GJR-GARCH model developed by Glosten, Jagannathan, and Runkle (1993) and the TGARCH model share notable conceptual similarities. Li and Lam (1995) modeled the asymmetric behavior in the stock returns using a threshold-type ARCH model. Using the Hang Seng index, they showed that the conditional mean of the return series fluctuated according to the ups and downs of the financial market on the previous day. Zhang, Russel, and Tsay (2001) proposed a nonlinear durational model and named it as Threshold Autoregressive Conditional Duration (TACD) to analyze the transaction data. The TACD permits the expected duration to behave nonlinearly based on past durations. Men, Kolkiewicz and Wirjanto (2019) proposed the Threshold Stochastic Conditional Duration (TSCD) model which is an extension to the SCD models originally proposed by Bauwens, Luc and David (2004). In the TSCD model, innovations follow the threshold distributions with positive support, and employ Gamma and Weibull distributions to model innovations. The latent random variable in the TSCD follows a TAR (1) process, and it switches between two regimes. Chen, Gerlach and Lin (2008) proposed the threshold heteroscedastic models in a range-based setting to analyze the intraday price range. In this paper they introduced a nonlinear volatility model for range and named it as the Threshold Conditional Autoregressive Range (TARR) model. This model is able to capture the asymmetry in range volatility by using a fixed threshold. They also introduced the TARRX model in which an exogenous variable is used, which is compared against a preset threshold value to determine the regime switching behavior. All the above threshold models switch regimens based on a fixed, predetermined threshold. Therefore, there is need to develop models where regime switching occurs dynamically, with past data alone determining the switch without a predetermined threshold. The main goal of this paper is to fill this need.

This paper proposes a parsimonious nonlinear time series model that can capture the heteroscedastic and asymmetric behaviors existing in the financial markets. We name this process the Threshold Asymmetric Conditional Autoregressive Range (TACARR) model. The TACARR model permits conditional mean range to depend nonlinearly on the past range series values. In this model, disturbance term of the price range model is expected to follow a threshold distribution with a positive support. Additionally, we assumed that the conditional mean switches between one of the two market regimes (upward market and downward market). In our formulation, the threshold values are self-adjusting, and every time new information arrives it dynamically determines the status of the market regime. The dynamic threshold component is driven by past upward and downward price returns. Moreover, the proposed TACARR model models the heteroscedastic volatility structure and capture the asymmetric behavior with fewer parameters compared to competing models such as the FACARR. Another objective of this study was to demonstrate the usefulness of the TACARR model in estimating and forecasting volatilities. In this paper, we estimate the model parameters for both exponential and lognormal TACARR models using the maximum likelihood method. We also derive expressions for the one step ahead out-of-sample volatility forecasting. Finally, we compare the performance of TACARR model with other range-based and asymmetric models by using IBM data.

This paper is organized as follows. Section 2 reviews the conditional heteroscedastic ranged-based models such as CARR, ACARR and TARR models. Section 3 introduces the TACARR model and its statistical properties. Following that, in Section 3, we develop maximum likelihood estimation methods to estimate the model parameters for Exponential TACARR (ETACARR) model and Lognormal TACARR (LNTACARR) model. Section 4 presents the out-of-sample forecasting method and performance evaluation techniques. Section 5 discusses the simulation study for both ETACARR and LNTACARR models. In Section 6, results of an empirical study of the proposed TACARR models based on IBM data is presented and the results are compared with those for other range-based models. Finally, Section 7 presents the concluding remarks.

2. REVIEW OF CARR, ACARR, FACARR AND TARR MODELS

2.1. THE CONDITIONAL AUTOREGRESSIVE RANGE MODEL

Chou (2005) proposed the CARR, which is primarily a range-based model. The CARR model is used to fit the price volatility of an asset by considering range as a measure of price volatility. Let P_s be the logarithmic price of an asset at time point $s \in (t-1, t]$, and the highest and lowest logarithmic prices of an asset during the interval $(t-1, t]$ are P_t^{high} and P_t^{low} respectively. Let R_t be the price range defined over the fixed time period $(t-1, t]$ is formulated as follows:

$$R_t = P_t^{high} - P_t^{low}.$$

The CARR model of order (*p*, *q*) is presented as CARR (*p*, *q*) and defined as follows:

$$\begin{aligned}
&R_t = \lambda_t \varepsilon_t, \\
&\mathrm{E}(R_t \mid \mathrm{F}_{t-1}) = \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i R_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \\
&\varepsilon_t \sim f(.), i.i.d., E(\varepsilon_t) = 1, \\
&0 < \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1, \alpha_i > 0, \beta_j > 0.
\end{aligned}$$

Here, λ_t is the conditional expectation of the price range based on all information up to time *t*. The non-negative disturbance term, also known as the standardized range, is denoted by ε_t which is independent and identically distributed with probability density function $f(.)$ with a non-negative support and a unit mean.

2.2. THE ASYMMETRIC CONDITIONAL AUTOREGRESSIVE RANGE (ACARR) MODEL

The ACARR model presented by Chou (2006), decomposed the range (R_t) series into two components, namely upward range (R_t^u) and downward range (R_t^d). Upward and downward ranges are expresses as the differences between the daily high (P_t^{high}), daily low (P_t^{low}), and the opening (P_t^{open}) logarithmic price of an asset respectively, over the time interval t.

$$\begin{aligned} R_t^u &= P_t^{high} - P_t^{open}, \\ R_t^d &= P_t^{open} - P_t^{low}, \\ R_t &= R_t^u + R_t^d = P_t^{high} - P_t^{open} + P_t^{open} - P_t^{low} = P_t^{high} - P_t^{low}. \end{aligned} \tag{2.1}$$

Here, the upward range measures the maximum gain or the positive shock to the stock, while downward range calculates the minimum gain or the negative impact to the stock price for the time period t.

The CARR model is symmetric because it treats the high and low price symmetrically. However, it is possible to assume that the upward and downward movements exhibit different in their dynamics of the volatility shocks. To allow the asymmetric behavior in price range data, Chou (2006) proposed and developed the ACARR model. ACARR model of order (p, q) is presented as follows:

$$
\begin{aligned}
&R_t = R_t^u + R_t^d, \\
&R_t^u = \lambda_t^u \varepsilon_t^u, \\
&R_t^d = \lambda_t^d \varepsilon_t^d, \\
&E\left(R_t^u \mid \mathrm{F}_{t-1}\right) = \lambda_t^u = \omega^u + \sum_{i=1}^{p} \alpha_i^u R_{t-i}^u + \sum_{j=1}^{q} \beta_j^u \lambda_{t-j}^u, \\
&E\left(R_t^d \mid \mathrm{F}_{t-1}\right) = \lambda_t^d = \omega^d + \sum_{i=1}^{p} \alpha_i^d R_{t-i}^d + \sum_{j=1}^{q} \beta_j^d \lambda_{t-j}^d, \\
&\varepsilon_t^u \sim f^u(.), i.i.d., E\left(\varepsilon_t^u\right) = 1, \\
&\varepsilon_t^d \sim f^d(.), i.i.d., E\left(\varepsilon_t^d\right) = 1, \\
&0 < \sum_{i=1}^{p} \alpha_i^u + \sum_{j=1}^{q} \beta_j^u < 1, \alpha_i^u > 0, \beta_j^u > 0, \\
&0 < \sum_{i=1}^{p} \alpha_i^d + \sum_{j=1}^{q} \beta_j^d < 1, \alpha_i^d > 0, \beta_j^d > 0.
\end{aligned}
\tag{2.2}
$$

Here, $\lambda_t^u \left(= E\left(R_t^u \mid \mathrm{F}_{t-1}\right)\right)$ is the conditional mean of the upward range on all information up to time period t-1, and $\lambda_t^d \left(= E\left(R_t^d \mid \mathrm{F}_{t-1}\right)\right)$ is the conditional mean of the downward range on all information up to time period t-1. The disturbance term of the upward (downward) range model is $\varepsilon_t^u \left(\varepsilon_t^d\right)$ independently and identically distributed as the density function $f^u(.)$ $\left(f^d(.)\right)$ with unit mean. Moreover, the pairs of parameters, $\left(\omega^u, \omega^d\right), \left(\alpha_i^u, \alpha_i^d\right), \left(\beta_j^u, \beta_j^d\right)$ are identified the asymmetric behavior between the upward range and downward range.

2.3. THE FEEDBACK ASYMMETRIC CONDITIONAL AUTOREGRESSIVE RANGE (FACARR) MODEL

The ACARR model assumes that there is independence between the upward and downward shocks, and Xie (2018) argued against this assumption and presented the FACARR model. This model includes the cross-interdependence terms on top of the

ACARR setting. Following the same definitions and notations, the FACARR model is defined as follows:

$$
\begin{aligned}
&R_t = R_t^u + R_t^d, \\
&R_t^u = \lambda_t^u \varepsilon_t^u, \\
&R_t^d = \lambda_t^d \varepsilon_t^d, \\
&\lambda_t^u = \omega^u + \sum_{i=1}^{p} \alpha_i^u R_{t-i}^u + \sum_{j=1}^{q} \beta_j^u \lambda_{t-j}^u + \sum_{k=1}^{l} \gamma_k^u R_{t-k}^d, \\
&\lambda_t^d = \omega^d + \sum_{i=1}^{p} \alpha_i^d R_{t-i}^d + \sum_{j=1}^{q} \beta_j^d \lambda_{t-j}^d + \sum_{k=1}^{l} \gamma_k^d R_{t-k}^u, \\
&\varepsilon_t^u \sim f^u(.),\ i.i.d.,\ E(\varepsilon_t^u) = 1, \\
&\varepsilon_t^d \sim f^d(.), i.i.d.,\ E(\varepsilon_t^d) = 1, \\
&0 < \sum_{i=1}^{p} \alpha_i^u + \sum_{j=1}^{q} \beta_j^u < 1, \alpha_i^u > 0, \beta_j^u > 0, \text{ and} \\
&0 < \sum_{i=1}^{p} \alpha_i^d + \sum_{j=1}^{q} \beta_j^d < 1, \alpha_i^d > 0, \beta_j^d > 0.
\end{aligned} \tag{2.3}
$$

In addition to the previous parameter set discussed in the ACARR model in (2.2), FACARR has a new pair of parameters, namely (γ^u, γ^d), which measures the magnitude and the direction of the lagged upward (downward) range on conditional mean range.

2.4. THE THRESHOLD AUTOREGRESSIVE RANGE (TARR) MODEL

Chen el.at (2008) proposed the Threshold Autoregressive Range (TARR) model which is a range-based threshold heteroskedastic model to analyses the price range data. Let P_s be the logarithmic price of an asset at time $s \in (t-1, t]$, and the price range of fix time interval $(t-1, t]$ is defined as $R_t = P_t^{high} - P_t^{low}$, then the TARR model with order (1, 1) is defined as given in the equation (2.4). Moreover, the market regimes were defined based on the previous range data R_{t-i} such that if R_{t-i} is greater (lower) than the predefined fixed threshold value (i.e., mean, median or a quartile). In this study in-sample mean price range $\overline{R}$ was considered as the threshold value.

$$
\begin{aligned}
&R_t = \lambda_t \varepsilon_t, \\
&E(R_t \mid \mathrm{F}_{t-1}) = \lambda_t = \begin{cases} \lambda_t^{(r1)} = \omega^{(r1)} + \sum_{i=1}^{p} \alpha_i^{(r1)} R_{t-i} + \sum_{j=1}^{q} \beta_j^{(r1)} \lambda_{t-j} : R_{t-i} \geq \overline{R} \\ \lambda_t^{(r2)} = \omega^{(r2)} + \sum_{i=1}^{p} \alpha_i^{(r2)} R_{t-i} + \sum_{j=1}^{q} \beta_j^{(r2)} \lambda_{t-j} : R_{t-i} < \overline{R} \end{cases}, \\
&\varepsilon_t = \begin{cases} \varepsilon_t^{(r1)} \sim f^{(r1)}(.)\, i.i.d., E\left(\varepsilon_t^{(r1)}\right) = 1 \ : R_{t-i} \geq \overline{R} \\ \varepsilon_t^{(r2)} \sim f^{(r2)}(.)\, i.i.d., E\left(\varepsilon_t^{(r2)}\right) = 1 : R_{t-i} < \overline{R} \end{cases}, \\
&cov\left(\varepsilon_t^{(r1)}, \varepsilon_t^{(r2)}\right) = 0. \\
&\omega^{(r1)} > 0, \alpha_i^{(r1)} \geq 0, \beta_j^{(r1)} \geq 0, \\
&\omega^{(r2)} > 0, \alpha_i^{(r2)} \geq 0, \beta_j^{(r2)} \geq 0.
\end{aligned} \tag{2.4}
$$

Here, *r1* represents the market regime 1, while *r2* denotes the market regime 2. In this model market regimes are determined by a constant threshold $\overline{R}$, which is estimated using the sample data. To be more specific, if the stock performs above the average stock price, then market belongs to *r1* regime and if it below the average stock then it belongs to *r2* market regime.

3. THRESHOLD ASYMMETRIC CONDITIONAL AUTOREGRESSIVE RANGE (TACARR) MODEL AND STATISTICAL PROPERTIES

Let $\{R_t\}$ be a sequence of price range values for the speculative asset defined over N time intervals such that $t = 1, 2, 3, ..., N$. Here R_t is calculated by taking difference between the highest $\left(P_t^{high}\right)$ and lowest $\left(P_t^{low}\right)$ logarithmic price of an asset during the time period *t*. In this study, we divided the market into two regimes namely upward market and downward market which were determined based on past information about upward price range and downward price range components. In contrast to TARR, where the threshold remained static, in this paper, we introduced a novelty method whereby the threshold values kept self-adjusting as and when the new information arrived.

The proposed TACARR model of order (l, p, q) is presented as follows:

$$
\begin{aligned}
&R_t = \lambda_t \varepsilon_t, \\
&E(R_t \mid \mathrm{F}_{t-1}) = \lambda_t = \begin{cases} \lambda_t^{(U)} = \omega^{(U)} + \sum_{i=1}^{p} \alpha_i^{(U)} R_{t-i} + \sum_{j=1}^{q} \beta_j^{(U)} \lambda_{t-j} : C_{l,t}^{(U)} \geq C_{l,t}^{(D)} \\ \lambda_t^{(D)} = \omega^{(D)} + \sum_{i=1}^{p} \alpha_i^{(D)} R_{t-i} + \sum_{j=1}^{q} \beta_j^{(D)} \lambda_{t-j} : C_{l,t}^{(U)} < C_{l,t}^{(D)} \end{cases}, \\
&\varepsilon_t = \begin{cases} \varepsilon_t^{(U)} \sim f^{(U)}(.) i.i.d., E\left(\varepsilon_t^{(U)}\right) = 1 : C_{l,t}^{(U)} \geq C_{l,t}^{(D)} \\ \varepsilon_t^{(D)} \sim f^{(D)}(.) i.i.d., E\left(\varepsilon_t^{(D)}\right) = 1 : C_{l,t}^{(U)} < C_{l,t}^{(D)} \end{cases}, \\
&C_{l,t}^{(U)} = \sum_{i=1}^{l} I_i \left[R_{t-i}^{u} \geq R_{t-i}^{d} \right], \\
&C_{l,t}^{(D)} = \sum_{i=1}^{l} I_i \left[R_{t-i}^{u} < R_{t-i}^{d} \right], \\
&cov\left(\varepsilon_t^{(U)}, \varepsilon_t^{(D)}\right) = 0, \\
&\omega^{(U)} > 0, \alpha_i^{(U)} \geq 0, \beta_j^{(U)} \geq 0, \\
&\omega^{(D)} > 0, \alpha_i^{(D)} \geq 0, \beta_j^{(D)} \geq 0.
\end{aligned} \tag{3.1}
$$

Here, l be the length of the time span which is used to determine the market regime. The error term for a given market regime m=(U=*upward market* , D=*downward market*), is denoted by $\left\{\varepsilon_t^{(m)}\right\}$ and is an independent and identically distributed sequence with non-negative support $[0, \infty)$ density function $f^{(m)}(.)$ such $E\left[\varepsilon_t^{(m)}\right] = 1$ and $cov\left(\varepsilon_t^{(U)}, \varepsilon_t^{(D)}\right) = 0$. Here, F_{t-1} be the sigma field generated from all the information set including range, upward range and downward range up to time *t*-1, which is expressed as $\left\{\{R_s\},\{R_s^u\},\{R_s^d\}\right\}_{s=1}^{t-1}$. Let $M_t = m$ be the current market condition at time t, that is determined based on the lagged values of upward $\left(R_{t-i}^u\right)$and downward $\left(R_{t-i}^d\right)$price range series which is expressed as $\left\{\{R_s^u\},\{R_s^d\}\right\}_{s=1}^{t-1} = \mathcal{M}_{t-1} \subset \mathcal{F}_{t-1}$ where, $\mathcal{M}_t$ is the sigma field generated from upward and downward range information up to time *t*-1 . To be more specific M_t is defined as:

$$M_t = m = \begin{cases} U : C_{l,t}^{(U)} \geq C_{l,t}^{(D)} \\ D : C_{l,t}^{(U)} < C_{l,t}^{(D)} \end{cases}. \tag{3.2}$$

Here,

$C_{l,t}^{(m)} = g\left(\left\{R_s^u, R_s^d\right\} : s \leq t-1\right)$ where $C_{l,t}^{(U)} = \sum_{i=1}^{l} I_i\left[R_{t-i}^u \geq R_{t-i}^d\right]$, $C_{l,t}^{(D)} = \sum_{i=1}^{l} I_i\left[R_{t-i}^u < R_{t-i}^d\right]$.

Then λ_t be the conditional expectation of R_t given $\mathcal{F}_{t-1}$and this term is formulated by considering different the market regimes *m*. Moreover, upward market regime means that the stock price is rising over the time, while downward market regimes implies that the stock price is dropping over the time.

We proposed the TACARR model as an alternative to the regular CARR model in which conditional expectation is expressed by assuming symmetric behavior of the price range series. The proposed TACARR model could overcome this major drawback in the regular CARR model by permitting the conditional mean to depend nonlinearly on past price series information and using threshold to capture the asymmetry. The proposed the TACARR model considered the most recent dynamic structure to segregate the market which was the main advantage over the TARR model in which fixed value is used to determine the market regimes. Moreover, the TACARR model can also be viewed as an asymmetric alternative to the ACARR and FACARR models. Both models treated price ranges asymmetrically however the proposed TACARR model had the freedom to model price volatility using a less (or equal) number of variables with compare to the FACARR (ACARR) model. Therefore, the TACARR model was an effective parsimonious model. Moreover, empirical studies showed that upward and downward price range data contained excess number of zero thus, some distributions such as lognormal distribution had to be discarded. However, in our proposed model, we analyzed the range data, which was positive in nature, hence we were able to consider positive support distributions to analyze the price range data.

3.1. THE ARMA REPRESENTATION OF TACARR MODEL

This section derives the ARMA representation of the proposed TACARR model. Let define zero mean martingale difference process $\{\eta_t\}$ such that:

$$\eta_t = R_t - E\left(R_t \mid \mathrm{F}_{t-1}\right) = R_t - \lambda_t,$$
$$E\left(\eta_t\right) = E\left(R_t - \lambda_t\right) = 0,$$
$$\operatorname{cov}\left(\eta_t, \eta_{t-h}\right) = E\left(\eta_t \eta_{t-h}\right) = 0 \text{ for } h \geq 1.$$

The proposed TACARR (*l, p, q*) given in (3.1) can be rearranged as an ARMA process with order *k* and *q*, where $k = \max(p,q)$,

$$R_t - \eta_t = \omega^{(m)} + \sum_{i=1}^{p} \alpha_i^{(m)} R_{t-i} + \sum_{j=1}^{q} \beta_j^{(m)} \left(R_{t-j} - \eta_{t-j}\right),$$
$$R_t = \omega^{(m)} + \sum_{i=1}^{p} \alpha_i^{(m)} R_{t-i} + \sum_{j=1}^{q} \beta_j^{(m)} R_{t-j} - \sum_{j=1}^{q} \beta_j^{(m)} \eta_{t-j} + \eta_t,$$
$$R_t = \omega^{(m)} + \sum_{i=1}^{k=\max(p,q)} \left(\alpha_i^{(m)} + \beta_i^{(m)}\right) R_{t-i} - \sum_{j=1}^{q} \beta_j^{(m)} \eta_{t-j} + \eta_t.$$

Here $\alpha_i^{(m)} = 0$, for $i > p$ and $\beta_j^{(m)} = 0$, for $j > q$, where *m*= (*U*, *D*).

3.2. THE UNCONDITIONAL EXPECTATION OF TACARR MODEL FOR EACH MARKET REGIME

Under the stationary assumption $E(R_{t-i}) = E(R_t) = E(\lambda_i)$, then the unconditional expectation of price range data can be derived as follows:

Let $\mathrm{F}_{t-1} = \left\{ \{R_s\}_{s=1}^{t-1}, \mathrm{M}_{t-1} \right\}$ be the sigma field generated from all the information set up to time t-1,

$$E(R_t \mid \mathrm{F}_{t-1}) = E\left(R_t \mid \{R_s\}_{s=1}^{t-1}, \mathrm{M}_{t-1}\right) = \lambda_t = \omega^{(m)} + \sum_{i=1}^{p} \alpha_i^{(m)} R_{t-i} + \sum_{j=1}^{q} \beta_i^{(m)} \lambda_{t-j},$$

$$E\left[E\left(R_t \mid \{R_s\}_{s=1}^{t-1}, \mathrm{M}_{t-1}\right)\right] = E(\lambda_t) = E\left(\omega^{(m)} + \sum_{i=1}^{p} \alpha_i^{(m)} R_{t-i} + \sum_{j=1}^{q} \beta_i^{(m)} \lambda_{t-j}\right),$$

$$E(R_t) = E(\lambda_t) = \omega^{(m)} + \sum_{i=1}^{p} \alpha_i^{(m)} E(R_{t-1}) + \sum_{j=1}^{q} \beta_j^{(m)} E(\lambda_{t-1}),$$

$$E(R_t) = \begin{cases} \dfrac{\omega^{(u)}}{1 - \sum_{i=1}^{p} \alpha_i^{(u)} - \sum_{j=1}^{q} \beta_j^{(u)}}, \forall R_t \in U \\ \dfrac{\omega^{(d)}}{1 - \sum_{i=1}^{p} \alpha_i^{(d)} - \sum_{j=1}^{q} \beta_j^{(d)}}, \forall R_t \in D \end{cases}.$$

4. PARAMETER ESTIMATION METHOD

In this section the Maximum Likelihood Estimation (MLE) method was developed to estimate the proposed TACARR model parameters. Here, we considered two versions of the TACARR model based on the distribution of the residual terms $\{\varepsilon_t\}$. In the Exponential TACARR (ETACARR) model, the residuals of the price range process follow a threshold exponential distribution expressed as (4.1). The Lognormal TACARR

(LNTACARR) process has residuals which follows a threshold lognormal distribution in which the parameters of the distribution are driven by market behavior presented in (4.4). Latter parts of this section, out-of-sample forecasting and performance evaluation methods were presented and discussed.

4.1. PARAMETER ESTIMATION METHOD FOR EXPONENTIAL TACARR (ETACARR) MODEL

Let $\left\{\varepsilon_t^{(m)}\right\}$ be the sequence of independent and identically distributed exponential disturbance term for a given market behavior *m*= (*U* =*upward market, D* =*downward market*) such that:

$$\varepsilon_t = \begin{cases} \varepsilon_t^{(U)} \sim \exp(1)\ i.i.d., E\left(\varepsilon_t^{(U)}\right) = 1 : C_{l,t}^{(U)} \geq C_{l,t}^{(D)} \\ \varepsilon_t^{(D)} \sim \exp(1) i.i.d., E\left(\varepsilon_t^{(D)}\right) = 1 : C_{l,t}^{(U)} < C_{l,t}^{(D)} \end{cases}, \tag{4.1}$$

Where $cov\left(\varepsilon_t^{(U)}, \varepsilon_t^{(D)}\right) = 0$. The parameter vector $\Phi = \left(\omega^{(U)}, \alpha^{(U)}, \beta^{(U)}, \omega^{(D)}, \alpha^{(D)}, \beta^{(D)}\right)'$ can be estimated by using the conditional likelihood function applying maximum likelihood estimation procedure.

4.1.1. The Log Likelihood Function for the ETACARR Model.

The conditional distribution of R_t given the information up to *t*-1 can be expressed as follows:

$$f\left(R_t \mid \mathrm{F}_{t-1}, \Phi\right) \square \frac{1}{\lambda_t} \exp\left(-\frac{R_t}{\lambda_t}\right),$$

$$\lambda_t = \begin{cases} \lambda_t^{(U)} = \omega^{(U)} + \sum_{i=1}^{p} \alpha_i^{(U)} R_{t-i} + \sum_{j=1}^{q} \beta_j^{(U)} \lambda_{t-j} : C_{l,t}^{(U)} \geq C_{l,t}^{(D)} \\ \lambda_t^{(D)} = \omega^{(D)} + \sum_{i=1}^{p} \alpha_i^{(D)} R_{t-i} + \sum_{j=1}^{q} \beta_j^{(D)} \lambda_{t-j} : C_{l,t}^{(U)} < C_{l,t}^{(D)} \end{cases}. \tag{4.2}$$

Therefore, the conditional likelihood function $L(\Phi | \mathrm{F}_{t-1})$ and the log likelihood function of the data $l(\Phi | \mathrm{F}_{t-1})$ can be derived as follows:

$$L(\Phi | \mathrm{F}_{t-1}) = \prod_{t=2}^{n} f(R_t | \mathrm{F}_{t-1}, \Phi),$$
$$l(\Phi | \mathrm{F}_{t-1}) = \ln\left[L(\Phi | \mathrm{F}_{t-1})\right] = \sum_{t=2}^{n} \ln\left[f(R_t | \mathrm{F}_{t-1}, \Phi)\right],$$
$$l(\Phi | \mathrm{F}_{t-1}) = -\sum_{t=2}^{n}\left(\ln(\lambda_t) + \frac{R_t}{\lambda_t}\right). \tag{4.3}$$

4.2. PARAMETER ESTIMATION METHOD FOR LOGNORMAL TACARR (LNTACARR) MODEL

In this section, we employed the lognormal distribution to model the error term in the model. Let $\left\{\varepsilon_t^{(m)}\right\}$ be the sequence of independent and identically distributed lognormal disturbance term for a given market regimes *m*= (*U=upward market, D=downward market*) such that:

$$\varepsilon_t = \begin{cases} \varepsilon_t^{(U)} \sim LN\left(-\frac{\theta_{(U)}^2}{2};\theta_{(U)}^2\right) .i.i.d; E\left(\varepsilon_t^{(U)}\right) = 1 : C_{l,t}^{(U)} \geq C_{l,t}^{(D)} \\ \varepsilon_t^{(D)} \sim LN\left(-\frac{\theta_{(D)}^2}{2};\theta_{(D)}^2\right) .i.i.d; E\left(\varepsilon_t^{(D)}\right) = 1 : C_{l,t}^{(U)} < C_{l,t}^{(D)} \end{cases}, \tag{4.4}$$

where $cov\left(\varepsilon_t^{(U)}, \varepsilon_t^{(D)}\right) = 0$. The maximum likelihood estimation procedure is used to estimate the parameter vector $\Phi = \left(\omega^{(U)}, \alpha^{(U)}, \beta^{(U)}, \omega^{(D)}, \alpha^{(D)}, \beta^{(D)}, \theta_{(U)}, \theta_{(D)}\right)'$.

4.2.1. The Log Likelihood Function for the LNTACARR Model. The conditional distribution of R_t given the information up to *t*-1 can be expressed as follows:

$$f\left(R_t \mid \mathrm{F}_{t-1},\Phi\right)=\frac{1}{\sqrt{2\pi\theta^2_{(M_t)}}R_t}\exp\left[-\frac{1}{2\theta^2_{(M_t)}}\left(\ln\left(R_t\right)-\ln\left(\lambda_t\right)+\frac{\theta^2_{(M_t)}}{2}\right)^2\right],$$

$$\lambda_t=\begin{cases}\lambda_t^{(U)}=\omega^{(U)}+\sum_{i=1}^{p}\alpha_i^{(U)}R_{t-i}+\sum_{j=1}^{q}\beta_j^{(U)}\lambda_{t-j}:C_{l,t}^{(U)}\geq C_{l,t}^{(D)}\\ \lambda_t^{(D)}=\omega^{(D)}+\sum_{i=1}^{p}\alpha_i^{(D)}R_{t-i}+\sum_{j=1}^{q}\beta_j^{(D)}\lambda_{t-j}:C_{l,t}^{(U)}<C_{l,t}^{(D)}\end{cases}. \quad (4.5)$$

Here, M_t is the current market condition, as formulated in (3.2), which is driven by the information of upward and downward price range series up to time t-1. Therefore, the conditional likelihood function $L\left(\Phi \mid \mathrm{F}_{t-1}\right)$ and the log likelihood function of the data $l\left(\Phi \mid \mathrm{F}_{t-1}\right)$ can be derived as follows:

$$L\left(\Phi \mid \mathrm{F}_{t-1}\right)=f\left(R_t \mid \mathrm{F}_{t-1},\Phi\right),$$

$$l\left(\Phi \mid \mathrm{F}_{t-1}\right)=\ln\left[L\left(\Phi \mid \mathrm{F}_{t-1}\right)\right]=\sum_{t=2}^{n}\ln\left[f\left(R_t \mid \mathrm{F}_{t-1},\Phi\right)\right],$$

$$l\left(\Phi \mid \mathrm{F}_{t-1}\right)=-\frac{1}{2}\sum_{t=2}^{n}\left[\ln(2\pi\theta^2_{(M_t)})+2\ln(R_t)+\frac{1}{\theta^2_{(M_t)}}\left(\ln\left(R_t\right)-\ln\left(\lambda_t\right)+\frac{\theta^2_{(M_t)}}{2}\right)^2\right]. \quad (4.6)$$

The parameters for the proposed model were estimated by using the MLE method as discussed in the above section using in-sample data. Next, we evaluate the in-sample performance of the proposed TACARR model with other conditional heteroscedastic range-based model by comparing the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) values.

$$RMSE=\sqrt{\frac{\sum_{t=1}^{N}\left(R_t-\hat{R}_t\right)^2}{N}};\ MAE=\frac{\sum_{t=1}^{N}\left|R_t-\hat{R}_t\right|}{N}.$$

Here, R_t is the price range at time t and $\hat{R}_t$ be the predicted price range at time t.

4.3. OUT-OF-SAMPLE FORECASTING

Under out of sample forecasting, we used the rolling window approach to forecast out-of-sample values. In the rolling window approach, first we divided the entire sample period (sample size =*T*) into two periods namely in-sample period (in-sample size=$N<T$) and out-of-sample period. The first one-step-ahead out-of-sample forecasting is carried out using the all the *N* in-sample data. The method is given bellow:

Let define, $R_{1:N}(1)$ be the one step ahead forecast of R_{N+1} where $R_{N+1} = \lambda_{N+1}\varepsilon_{N+1}$.

Then:

$$R_{1:N}(1) = E(R_{N+1} \mid \mathrm{F}_N) = \lambda_{N+1},$$

$$\lambda_{N+1} = \begin{cases} \lambda_{N+1}^{(U)} = \omega^{(U)} + \sum_{i=1}^{p} \alpha_i^{(U)} R_{N+1-i} + \sum_{j=1}^{q} \beta_j^{(U)} \lambda_{N+1-j} : C_{l,t}^{(U)} \geq C_{l,t}^{(D)} \\ \lambda_{N+1}^{(D)} = \omega^{(D)} + \sum_{i=1}^{p} \alpha_i^{(D)} R_{N+1-i} + \sum_{j=1}^{q} \beta_j^{(D)} \lambda_{N+1-j} : C_{l,t}^{(U)} < C_{l,t}^{(D)} \end{cases}. \tag{5.1}$$

Therefore, one step ahead forecast value $R_{1:N}(1)$ is calculated by using the conditional expectation of range given information up to time *N*.

After calculating the forecasted value for the $(N+1)^{\text{th}}$ observation ($f_{N+1} = R_{1:N}(1)$), then the sample window is moved to(2: *N*+1) to forecast $(N+2)^{\text{th}}$ observation (f_{N+2}). Next, we considered the window of (2: *N*+1) as the new in-sample data and recalculated the model parameters based on this new data. After the estimation, the estimated parameters were applied to the one step ahead forecasting method in equation (5.1) to calculate $R_{2:N+1}(1)$ which the forecasted value is for f_{N+2}. This process was repeated until all the future values were estimated in the out-of-sample data. Moreover, to check the forecasting accuracy of the proposed model with other competitive range models DM test was used (see Diebold & Marino, 1995).

5. SIMULATION STUDY

We investigated the finite sample performance of estimators using a simulation study. We used 'nloptr', which is a nonlinear optimization function of R software to generate the relevant data. Length of the time series studies was set to $n = 1000$ and $n = 3000$, and $s = 1000$ simulations runs were carried out for each parameter sample size combination. We carried out this simulation study for the two different error distributions such as exponential and lognormal. The idea of the proposed model was motivated by the TARR model, and in both TARR and TACARR models we had different regimes. In this study, we proposed two TACARR type models namely exponential TACARR and lognormal TACARR model. In this simulation study, we considered the different parameter combinations for ETACARR (1, 1, 1) model and LNTACARR (1, 1, 1). Simulation study consisted of two parts. First, we generated the price range data for the proposed ETACARR model and the LNTACARR model based on the equation (4.1) – (4.2) and (4.4) – (4.5), respectively. Then, we maximized the profile likelihood functions (4.3) and (4.6), for the ETACARR model, and LNTACARR model respectively using the constrained nonlinear optimization function 'nloptr' in R. The Mean Absolute Deviation Error (MADE) is utilized as the evaluation criterion. The MADE is defined as, $\frac{1}{s}\sum_{j=1}^{s}\left|\hat{\phi}_j - \phi_j\right|$ where s is the number of replications. Simulation results are reported in Table 1 and Table 2.

Table 1 presents the simulation study results for the ETACARR model, and according to the results, we can see that MLE method did a good job in estimating model parameters. Accuracy increased with the size of the sample size.

According to the simulation results presented in Table 2, we can see that LNTACARR model parameters were estimated with higher accuracy by using the MLE method mentioned in the above equation (4.6). It can also be seen that MADE value was decreasing when the sample size was increasing.

Table 1: Means of MLE estimates and MADE (within parentheses), for ETACARR model with order (1, 1, 1)

	Upward Market			**Downward Market**		
	$\omega^{(U)}$	$\alpha^{(U)}$	$\beta^{(U)}$	$\omega^{(D)}$	$\alpha^{(D)}$	$\beta^{(D)}$
True Parameter	0.01	0.10	0.80	0.10	0.20	0.70
n=1000	0.0170 (0.0152)	0.0977 (0.0253)	0.7864 (0.0449)	0.0996 (0.0248)	0.1982 (0.0340)	0.7045 (0.0599)
n=3000	0.0130 (0.0101)	0.0995 (0.0142)	0.7943 (0.0283)	0.1003 (0.0153)	0.2004 (0.0196)	0.6995 (0.0367)
True Parameter	0.01	0.30	0.60	0.10	0.20	0.50
n=1000	0.0132 (0.0096)	0.3011 (0.0379)	0.5867 (0.0583)	0.1001 (0.0170)	0.1964 (0.0407)	0.5033 (0.0769)
n=3000	0.0110 (0.0058)	0.3007 (0.0223)	0.5955 (0.0349)	0.1008 (0.0098)	0.2006 (0.0239)	0.4965 (0.0443)
True Parameter	0.05	0.15	0.50	0.10	0.20	0.30
n=1000	0.0551 (0.0247)	0.1542 (0.0395)	0.4702 (0.1581)	0.0985 (0.0261)	0.1959 (0.0462)	0.3122 (0.1582)
n=3000	0.0517 (0.0126)	0.1505 (0.0221)	0.4897 (0.0809)	0.1008 (0.0154)	0.2015 (0.0264)	0.2944 (0.0933)

Table 2: Means of MLE estimates and MADE (within parentheses), for LNTACARR model with order (1, 1, 1)

n=3000	n=1000	True Parameter	n=3000	n=1000	True Parameter	n=3000	n=1000	True Parameter		
0.0490 (0.0077)	0.0502 (0.0122)	0.05	0.0106 (0.0054)	0.0125 (0.0085)	0.01	0.0110 (0.0075)	0.0131 (0.0109)	0.01	$\omega^{(U)}$	Upward Market
0.1525 (0.0209)	0.2080 (0.0597)	0.15	0.3002 (0.0235)	0.2974 (0.0394)	0.30	0.0998 (0.0102)	0.0983 (0.0178)	0.10	$\alpha^{(U)}$	
0.5037 (0.0514)	0.4430 (0.0928)	0.50	0.5974 (0.0338)	0.5906 (0.0551)	0.60	0.7980 (0.0204)	0.7950 (0.0318)	0.80	$\beta^{(U)}$	
0.0898 (0.0026)	0.0911 (0.0046)	0.09	0.9968 (0.0279)	0.9924 (0.0489)	1.00	0.2491 (0.0071)	0.2483 (0.0125)	0.25	$\theta^2_{(U)}$	
0.1008 (0.0071)	0.0997 (0.0111)	0.10	0.1002 (0.0092)	0.1005 (0.0157)	0.10	0.1002 (0.0126)	0.1000 (0.0195)	0.10	$\omega^{(D)}$	Downward Market
0.2136 (0.0200)	0.2009 (0.0298)	0.20	0.1995 (0.0233)	0.1991 (0.0410)	0.20	0.2000 (0.0166)	0.1997 (0.0289)	0.20	$\alpha^{(D)}$	
0.2824 (0.0468)	0.3007 (0.0727)	0.30	0.5000 (0.0428)	0.5007 (0.0756)	0.50	0.7002 (0.0302)	0.7016 (0.0482)	0.70	$\beta^{(D)}$	
0.0401 (0.0011)	0.0398 (0.0021)	0.04	0.9993 (0.0274)	0.9947 (0.0516)	1.00	0.6401 (0.0171)	0.6379 (0.0317)	0.64	$\theta^2_{(D)}$	

6. EMPIRICAL RESULTS

6.1. THE DATA SET

In this study, IBM stock indices were used to gauge the performance of the proposed TACARR model and compare it with the competitive models. The sample periods for the IBM data spanned from January 01, 2002 to March13, 2020. Daily values for the opening price, closing price, high price, low price, and adjusted price were reported over the span of the study period. The data set was obtained from the Yahoo Finance (https://finance.yahoo.com/) by using the ‘quantmod’ package in R software. The data set was divided in to two sub samples: the first sub sample which was also known as in-sample period and this sample was used to estimate the model parameter and in-sample predictions. In-sample periods for IBM spanning from January 01, 2002 to December 31, 2019. The second sub sample, which was also called as out-of-sample period, and this sample was used for out-of-sample forecasting. Out-of-sample periods for IBM, elapsed from January 1, 2020 to March 13, 2020. Table 3 presents the summary statistics of the IBM range data, which was calculated as given in (2.1).

Summary statistics for the IBM stock index is presented in the Table 3. According to the table we detected the high persistence on the IBM stock. For an example, the Ljung-Box statistics results for lags 1, 5 and 22 show that all form of range data exhibit highly significant correlations. Since the upward component had higher Ljung-Box test statistic values for all lags, than the downward price range component specified that the high persistence exist in the downward price range series. The price range data did not have zero range, but upward range and downward range components had a considerable number of zero data points. Positive skewness and absence of zeros in range data suggested positive support pdf, such as a lognormal can be used. By contrast, upward range and downward range components were positively skewed with notable numbers of zeros that implies a positively skewed nonnegative support pdf, such as an exponential distribution, must be used to model these components.

Table 3: Summary statistics of the IBM price range data

Statistics	Price Range	Upward Price Range	Downward Price Range
Number of Days	4581	4581	4581
Minimum	0.2928	0.0000	0.0000
Mean	1.6834	0.8717	0.8118
Maximum	11.2642	8.0510	8.4991
Standard Deviation	1.0774	0.8279	0.8741
Skewness	2.7976	2.4293	2.7510
Number of Zeros	0	107	160
Q (1)	1891.7***	235.45***	467.34***
Q (5)	8013***	923.36***	2954***
Q (22)	25498***	3271.1***	5915***

Note: *** indicates 1% significance level.

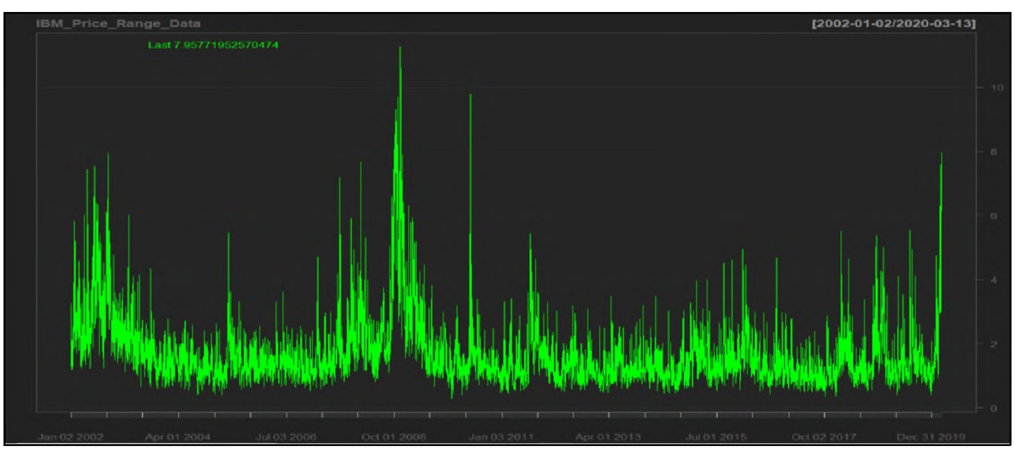

Figure 1: IBM price range data for the period of 01/01/2002 to 03/13/2020

According to the IBM price range data, as shown in Figure 1, the price volatility was high at the beginning of the sample period and it decreased. Then the daily price volatility fluctuated rapidly during 2007-2009 recession due to the financial crisis. After the economic bubble ended volatility dropped down and again at end of the year 2019 to the beginning of the year 2020, volatility was increased. This is due to the Covid-19 pandemic and its influence on the financial market.

6.2. IN-SAMPLE ESTIMATION RESULTS

In this paper, we introduced two versions of the TACARR models based on the pdf, which were used to model the residuals distribution. For an example the ETACARR model as mentioned in (4.1) - (4.2), used the exponential density to model the disturbance term while the LNTACARR model (4.4) -(4.5), used the lognormal densities to model the residual terms, which were representing different market segments. It was vital to identify which model better fit the dynamic structure of the price range data. To do that Kolmogorov-Simonov (KS) test was employed. The KS test was used to compare whether the standardized residual series followed the reference distribution. Moreover, we calculated the LLF (Log Likelihood Function), AIC (Akaike Information Criteria), and BIC (Bayesian Information Criteria) for each model and compared the results. The model with smaller AIC, BIC values and larger LLF value were considered to be a significantly better model than the other. Furthermore, diagnostic tests, such as the Ljung-Box test, for residuals were considered to check whether residuals were independent and identically distributed.

Parameter estimation results for the IBM price range data for the ETCARR and LNTACARR are summarized in the Table 4, and it contains two panels where upper panel (A) presents the parameter estimation results and model selection statistics such as LLF, AIC and BIC values. The lower panel (B) summarizes the diagnostic test results for the standardized residuals. The KS tests for the exponential and lognormal cases were separately considered to identify whether the residual series followed the hypothesized distribution with the estimated parameters. Then Ljung-Box test statistics for 1, 5 and 22, lags with the corresponding p-values were used to check whether the residual series are exhibited any serial correlations.

According to Table 4, Lognormal TACARR (l,1,1) models for all the different cases of l values, had smaller AIC, smaller BIC and larger LLF values when compared to its exponential alternative. Hence, it suggests that the LNTACRR models fit the data better than its exponential alternative. The LNTACARR models indicate that upward and downward markets have different variance parameters ($\theta^2_{(U)} \neq \theta^2_{(D)}$). The persistence estimates for the downward market is higher than that of the upward market ($\alpha^{(U)} + \beta^{(U)} < \alpha^{(D)} + \beta^{(D)}$). This shows that the downward market was more volatile than the upward market. Moreover, KS tests for the LNTACRR models suggest that the standardized residuals followed the hypothesized threshold lognormal distribution. However, for the ETACARR models the standardized residuals did not follow the proposed threshold exponential distribution and this proven by the KS test statistics. Based on the p-values for the Ljung-Box test for lags 1, 5, and 22, it can be concluded that we failed to reject the null hypothesis at 0.01 significance level, and it suggest that residual series were not serially correlated. Therefore, it can be concluded that the LNTACARR model fits better than the ETACARR model for all three cases of l values.

Table 4: Estimation and diagnostic test results of the TACARR models with exponential and lognormal disturbance term for IBM data

Table 4A: Estimation of the TACARR model (standard errors)						
	TACARR (22,1,1)		TACARR (5,1,1)		TACARR (1,1,1)	
	Exponential	Lognormal	Exponential	Lognormal	Exponential	Lognormal
$\omega^{(U)}$	0.0569 (0.0324)	0.0535 (0.0119)	0.0749 (0.0359)	0.0684 (0.0129)	0.0855 (0.0493)	0.0798 (0.0181)
$\alpha^{(U)}$	0.1680 (0.0432)	0.1683 (0.0164)	0.1776 (0.0428)	0.1674 (0.0155)	0.1615 (0.0395)	0.1573 (0.0147)
$\beta^{(U)}$	0.7944 (0.0544)	0.7960 (0.0207)	0.7682 (0.0543)	0.7831 (0.0197)	0.7672 (0.0553)	0.7777 (0.0206)
$\theta^2_{(U)}$		0.1316 (0.0039)		0.1361 (0.0039)		0.1386 (0.0040)
$\omega^{(D)}$	0.0554 (0.0316)	0.0488 (0.0114)	0.0364 (0.0366)	0.0321 (0.0135)	0.0313 (0.0547)	0.0248 (0.0198)
$\alpha^{(D)}$	0.2301 (0.0455)	0.2175 (0.0168)	0.2169 (0.0438)	0.2138 (0.0163)	0.2373 (0.0456)	0.2291 (0.0166)
$\beta^{(D)}$	0.7374 (0.0562)	0.7535 (0.0207)	0.7680 (0.0559)	0.7720 (0.0209)	0.7627 (0.0631)	0.7709 (0.0230)
$\theta^2_{(D)}$		0.1411 (0.0042)		0.1359 (0.0040)		0.1313 (0.0040)
LLF	-6547.26	-3609.47	-6546.68	-3605.79	-6544.64	-3592.76
AIC	13106.52	7234.95	13105.36	7227.57	13101.27	7201.51
BIC	13145.03	7286.30	13143.87	7278.92	13139.79	7252.86
Table 4B: Diagnostic test results of the TACARR model (p-values)						
	TACARR (22,1,1)		TACARR (5,1,1)		TACARR (1,1,1)	
	Exponential	Lognormal	Exponential	Lognormal	Exponential	Lognormal
KS	0.3578 (0.0000)	0.0254 (0.1080)	0.3593 (0.0000)	0.0274 (0.0672)	0.3609 (0.0000)	0.0252 (0.1136)
Q (1)	4.615 (0.3169)	6.2995 (0.0121)	5.2874 (0.0215)	7.014 (0.0081)	3.7791 (0.0519)	5.2215 (0.0223)
Q (5)	10.672 (0.0583)	12.26 (0.0315)	11.805 (0.0376)	13.271 (0.0210)	9.8418 (0.0799)	11.281 (0.0461)
Q (22)	30.55 (0.1102)	33.61 (0.0538)	32.457 (0.0699)	35.575 (0.0337)	28.95 (0.1463)	31.66 (0.0834)

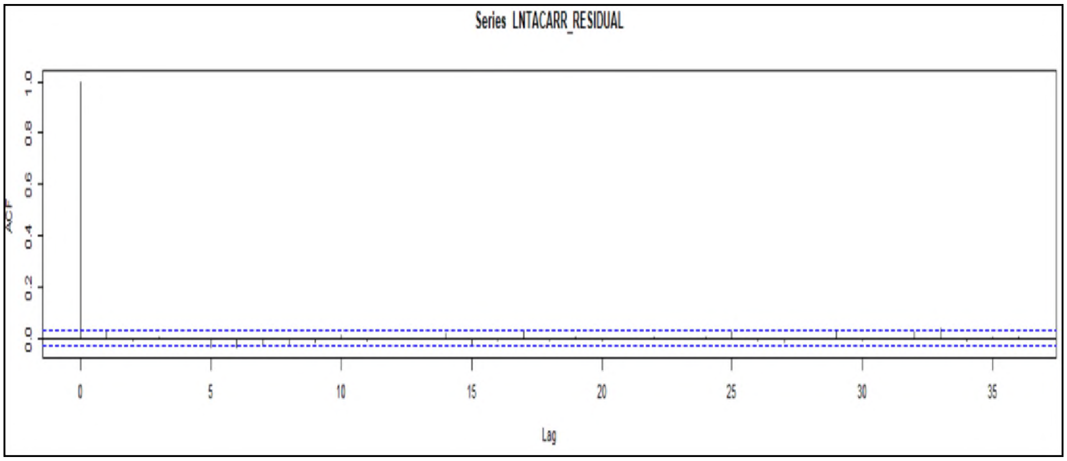

Figure 2: ACF plot of the LNTACARR (1, 1, 1) residuals

The ACF plot indicat that all residual of the LNTACARR (1, 1, 1) were within the 95% confidence interval. Therefore, graphically, it is seen that residuals are independent and identically distributed.

In the next sub section, we discuss how to select the optimal lag (l) in the proposed LNTACARR model. The optimal lag selection is an important task because it decides how many previous periods (i.e., day, week, or months) that we need to consider for categorizing the market. Here we considered three different lag values: 1, 5 and 22. In the case where l=1determined whether the status of the market regime was an upward market or downward market, based on the previous days upward and downward range data. Put simply, l=1, considered the most recent financial information to decide the market regime. Similar to that, when l=5 market is segregated based on the volatility information of the past business week, while the case of l=22 divided the market into two regimes, based on the market information flow throughout last business month. (Please note econometrics and finance literature 5 days is considered as one business week, and 22 days equals one business month). Since the number of lags became the deciding factor of market conditions, it is important to select the optimal lags. To select the optimal lags, we compared the AIC, BIC and LLF values of LNTACARR (*l*, 1, 1) model and picked the suitable *l*. To be more specific, we selected the LNTACARR (*l*, 1, 1) model with lowest AIC, BIC and largest LLF values. According to Table 4, the LNTACARR (1, 1, 1) model had the lowest AIC and BIC values and also largest LLF value; therefore, for the IBM data *l*=1, is the optimal number of previous lags to decide the market structure. This concludes that most recent market information, such as previous day financial news, is vital to deciding the status of the market regime or the current market structure rather than the financial knowledge gathered from the previous business month. This can also be viewed as, LNTACARR model with *l*=1 is more sensitive to the market information through the self-adjusting threshold component. This threshold component decide the market status by comparing upward and downward lagged price range data. Since price range data is not treated symmetrically, as was done in CARR models, we can say that the proposed model addresses the asymmetric behavior in the financial market.

Among all the TACARR (l, 1, 1) models, we accounted from the previous section, we finally selected LNTACARR (1, 1, 1) model as our candidate model for the IBM stock data. We also used this model to compare and contrast the adequacy with other conditional heteroscedastic range-based models.

$$\lambda_t = \begin{cases} \lambda_t^{(U)} = 0.0798 + 0.1573 R_{t-1} + 0.7777 \lambda_{t-1} : R_{t-1}^u \geq R_{t-1}^d \\ \lambda_t^{(D)} = 0.0248 + 0.2291 R_{t-1} + 0.7709 \lambda_{t-1} : R_{t-1}^d < R_{t-1}^u \end{cases},$$

$$\varepsilon_t = \begin{cases} \varepsilon_t^{(U)} \sim LN\left(-0.0693; 0.1386\right) : R_{t-1}^u \geq R_{t-1}^d \\ \varepsilon_t^{(D)} \sim LN\left(-0.0656; 0.1313\right) : R_{t-1}^u < R_{t-1}^d \end{cases}.$$

To gauge the in-sample performance of the chosen LNTACARR (1, 1, 1) model we considered RMSE and MAE values. Then, we compared the accuracy measurements with other competitive alternative range models, such as the LNCARR (1, 1), ACARR (1, 1), FACARR (1, 1) and LNARR (1, 1). However, the empirical study showed that upward and downward ranges had large numbers of zeroes for the IBM price range data. Due to this, lognormal distribution cannot be considered to model upward and downward price range components. Therefore, exponential upward and downward disturbance terms were used in the ACARR and the FACARR models. Moreover, we extended the comparison between models for the 2007-2009 economic recession period.

Based on the comparison results presented in the Table 5, the proposed LNTACARR model had the lowest RMSE and MAE values for the in-sample period when compared to that of the other candidate models. This result suggested that with respect to the model accuracy measurements, the proposed model performed slightly better than the others. Moreover, during the economic recession period, the lowest RMSE value was recorded in LNTACARR model. However, MAE value was lower in the FACARR model. This implied that the above model maintains higher prediction accuracy and more suitable to analyze the high volatile data with compared to the LNCARR (1, 1), ACARR (1, 1), FACARR (1, 1) and LNTARR (1, 1) models.

In-sample prediction for the LNTACARR model with order (1, 1, 1) for IBM data is presented in the Figure 3. According to the figure the proposed model had the same structural pattern that can be seen in the observed data. When there was a period with high volatilities, the LNTACARR model also estimated the high values for these periods.

Table 5: In-sample comparison between LNCARR (1, 1), ACARR (1, 1), FACARR (1, 1), LNTARR (1, 1) and LNTACARR (1, 1) for IBM data

Table 5A: Model performance comparison during full in-sample period					
Statistic	LNCARR (1,1)	ACARR (1,1)	FACARR (1,1)	LNTARR (1,1)	LNTACARR (1,1,1)
RMSE	0.7289	0.7568	0.7241	0.7276	**0.7224**
MAE	0.4989	0.5139	0.4969	0.4967	**0.4946**
Table 5B: Model performance comparison during economic recession period					
Statistic	LNCARR (1,1)	ACARR (1,1)	FACARR (1,1)	LNTARR (1,1)	LNTACARR (1,1,1)
RMSE	0.9347	0.9396	0.9233	0.9370	**0.9224**
MAE	0.6970	**0.6815**	0.6890	0.6983	0.6890

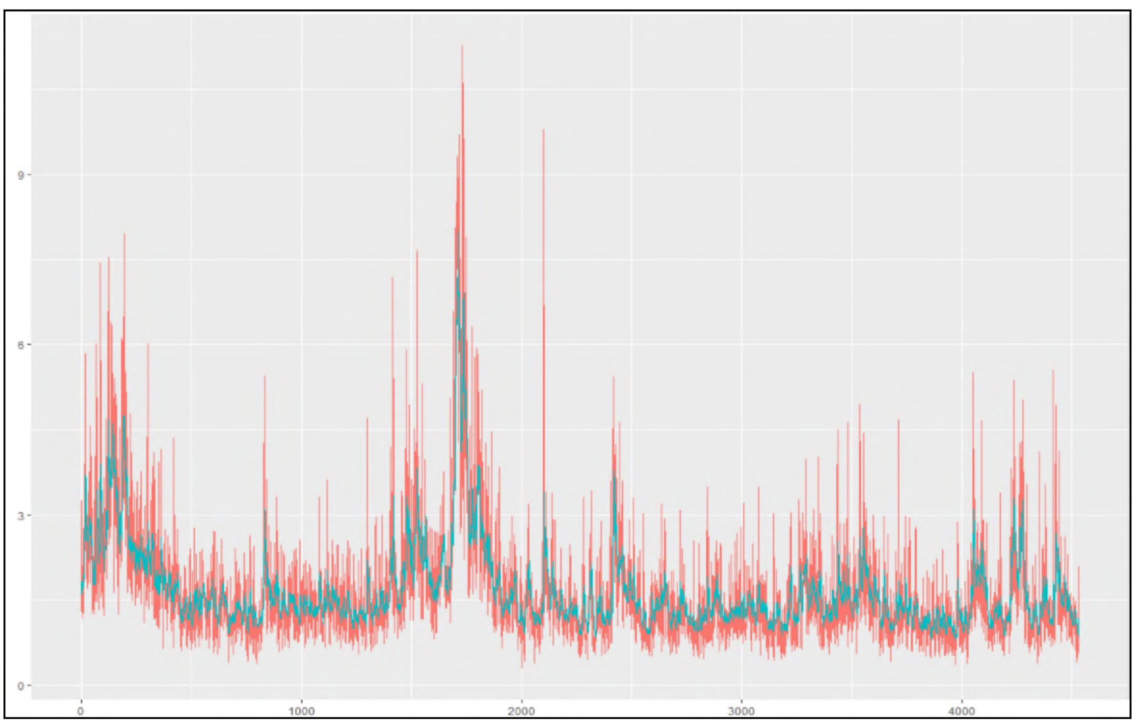

Figure 3: In-sample prediction (green) of the LNTACARR (1, 1, 1) model for the IBM price rage data (red)

6.3. OUT-OF-SAMPLE FORECASTING

Out-of-sample performance of the proposed Lognormal TACARR (1, 1, 1) model was compared with four other models namely the LNCARR (1, 1), ACARR (1, 1), FACARR (1, 1), and the LNTARR (1, 1). In this study, the out-of-sample period started on January 01, 2020 and ended on March 13, 2020. Length of the out-of-sample period equals to 50 days. The out-of-sample period showed high volatility due to the impact of the Covid-19 Pandemic on the financial market. First, the RMSE and MAE values were calculated and used these values as the performance indicator to gauge the performance of the proposed model. These results are presented in the Table 6. Then, we considered the Diebold & Marino (DM) test to check whether the proposed LNTACARR model with order (1, 1, 1) had a better forecasting accuracy than the other competitive models. In this test, the null hypothesis was that the LNTACARR (1, 1, 1) model had lower forecasting accuracy. The alternative hypothesis was stated that the one step ahead forecasted value of the LNTACARR (1, 1, 1) model was more accurate than the forecast values of its competitive model. The DM test result is summarized in Table 7.

Table 6**:** Out-of-sample comparison between LNCARR (1, 1), ACARR (1, 1), LNTARR (1, 1) and LNTACARR (1, 1, 1) for IBM data

Statistic	LNCARR (1,1)	ACARR (1,1)	FACARR (1,1)	LNTARR (1,1)	LNTACARR (1,1,1)
RMSE	1.2720	1.5203	1.2205	1.2828	**1.1858**
MAE	0.8371	0.9414	0.8024	0.8440	**0.7752**

According to Table 6, the LNTACARR model with order (1, 1, 1) had the lowest RMSE and MAE when compared to the other four models. Therefore, based on these accuracy measurements, it was concluded that proposed model performs better than LNCARR, ACARR, FACARR and LNARR models.

Figure 4 exhibits the graphical comparison LNTACARR model for out-of-sample forecasting values with LNCARR, ACARR and LNTARR models. According to the comparison, the proposed model has the ability to pick the high volatility values when compared to the other three models. In general, all the models in this study had larger forecasting errors during the out-of-sample period because this time spans represented the early COVID -19 days, and this was a high volatile period. However, the proposed model quickly adapted to the situation based on the past market volatility, hence it performed slightly better than the other four models.

Table 7**:** Diebold & Marino (DM) test results on IBM out-of-sample data

Null Hypothesis	**DM test statistics (p value)**
Forecast LNCARR (1,1) model is more accurate than that of the LNTACARR (1,1)	-2.4721 (0.0067)
Forecast ACARR (1,1) model is more accurate than that of the LNTACARR (1,1)	-2.7660 (0.0028)
Forecast FACARR (1,1) model is more accurate than that of the LNTACARR (1,1)	-2.2861 (0.0111)
Forecast LNTARR (1,1) model is more accurate than that of the LNTACARR (1,1,1)	-2.7505 (0.0032)

According to Table 7, it was concluded with 95% confidence that the propose LNTACARR model had higher forecasting accuracy than the other asymmetric range-based heteroscedastic models, such as LNCARR, ACARR and LNTARR.

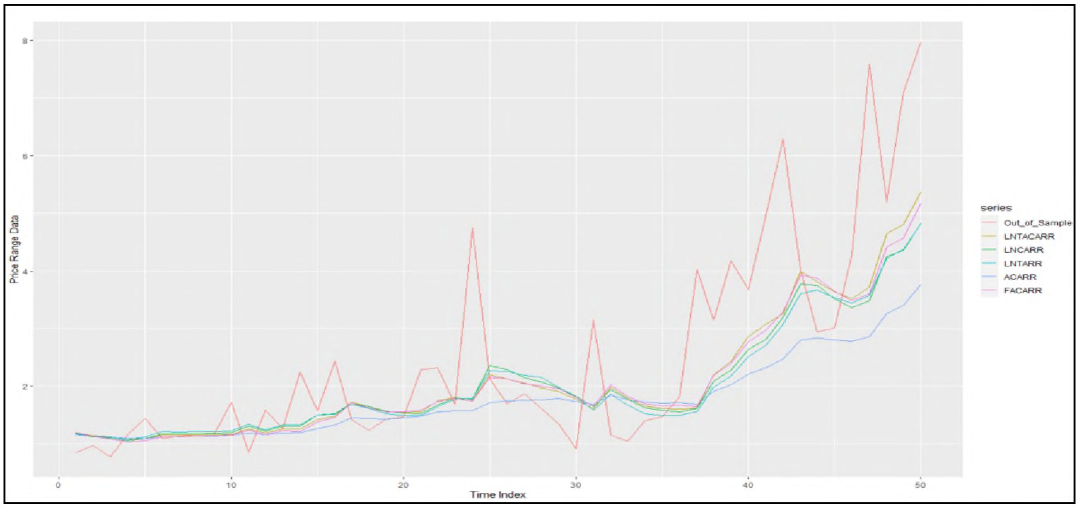

Figure 4: The out-of-sample forecast value comparison for IBM price range data

7. CONCLUSIONS

In this paper, we proposed Threshold Asymmetric Conditional Autoregressive Range (TACARR) model, which is a threshold heteroscedastic range-based model for modeling and forecasting financial price range data. We introduced a novel method whereby the threshold values self-adjust as new information arrives. Therefore, the proposed model is an effective approach for capturing the financial asymmetric behavior in the market by adjusting the threshold value according to the market behavior. In this study two market behaviors (regimes), namely upward market and downward markets were considered. Since the market regime is decided based on the past values of upward and downward price range data, it can be seen as a tool for dynamically capturing the regime switching while capturing the asymmetric behavior in price range volatility. Also, we allow the disturbance term in the model to behave differently in each market regime. We also investigated the best time frame over which switching decision should be made, based on historical data. We compared previous one day, previous one week, and previous one month as periods over which a switching decision can be made. In addition, we developed maximum likelihood estimation methods to estimate the model parameters for both ETACARR and LNTACARR models, which a simulation study showed as providing accurate parameters estimates. IBM price range data was used for to illustrate the model fit to empirical data. This study results show that the LNTACARR model performed better than its exponential alternative across all the different periods considered for making switching decisions. Moreover, we found that switching based on previous day's upward and downward ranges provided the best fit. For the in-sample and recession data, the predicted values of LNTACARR (1, 1, 1) model had higher accuracy when compared to the LNCARR, ACARR, FACARR and LNTARR models. Diagnostic test results for the model suggested that the residuals were independent and identically distributed, and it followed a lognormal distribution. Finally, out-of-sample forecasting evaluation was considered and according to the RMSE and MAE values the LNTACARR performed slightly better than the other four models. Furthermore, the DM test for the forecasting accuracy indicated that the proposed model had more accurate forecast than LNCARR, ACARR, FACARR and LNTARR models for the IBM price range data.

REFERENCES

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. Journal of Finance, 57(3), 1047–1092.

Bauwens, Luc, and David Veredas. (2004). The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations. Journal of Econometrics 119: 381–482.

Beckers, Stan (1983). Variance of Security Price Return Based on High, Low and Closing Prices. Journal of Business 56, 97–112.

Black, F., (1976). The pricing of commodity contracts, Journal of Financial Economics, 3, 167-179.

Bollerslev, Tim (1986). Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics 31, 307–327.

Brandt, M.W. and C.S. Jones (2006), "Volatility Forecasting with Range-Based EGARCH Models". Journal of Business and Economic Statistics, 24, 470-486.

Chen C.W.S, Gerlach, R. and Lin, E.M.H. (2008) "Volatility Forecasting Using Threshold Heteroskedastic Models of the Intra-day Range" Computational Statistics and Data Analysis, Special Issue on Statistical and Computational Methods in Finance,52, 2990-3010.

Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model. Journal of Money, Credit and Banking, Vol. 37, pp. 561-582.

Chou, R. (2006). Modeling the asymmetry of stock movements using price ranges. Advances in Econometrics, 20, 231–257.

Chou, R., & Liu, N. (2010). The economic value of volatility timing using a range-based volatility model. Journal of Economic Dynamics and Control, 34(11), 2288–2301.

Chou, H., & Wang, D. (2014). Estimation of tail-related value-at-risk measures: Range-based extreme value approach. Quantitative Finance, 14(2), 293–304.

Engle, R.F., (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation. Econometrica 50, 987–1008.

Engle, R.F., & Russell, J.R., (1998). Autoregressive Conditional Duration: A New Model for Irregular Spaced Transaction Data. Econometrica 66, 1127–1162.

Glosten, L., Jagannathan, R. and Runkle, D. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance 48: 1779-1801.

Kunitomo, N. (1992). Improving the Parkinson method of estimating security price volatilities. Journal of Business, 65(2), 295–302.

Lamoureux, C. G., & Lastrapes, W. D., (1990). Heteroscedasticity in Stock Return Data: Volume versus GARCH Effects. The Journal of Finance, 45, 221-229.

Li, W., & Lam, K. (1995). Modelling Asymmetry in Stock Returns by a Threshold Autoregressive Conditional Heteroscedastic Model. Journal of the Royal Statistical Society. Series D (The Statistician), 44(3), 333-341.

Men, Z., Kolkiewicz, A. W., & Wirjanto, T. S. (2019). Threshold stochastic conditional duration model for financial transaction data. Journal of Risk and Financial Management. 12(2), 1-21. Retrieved from www.scopus.com

Nelson, D. (1991). Conditional heteroscedasticity in asset returns: A new approach. Econometrica, 59, 347–370.

Nelson, D.B. and Cao, C.Q. (1992) 'Inequality Constraints in the Univariate GARCH Model', Journal of Business and Economic Statistics, Vol. 10, No. 3, pp. 229-235.

Parkinson, Michael (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. Journal of Business 53, 61–65.

Tong, H., (1978). On a threshold model. In Pattern recognition and Signal Processing (ed. C.H. Chen), pp. 575-586. Amsterdam: Sijthoff and Noordhoff.

Tsay, R. (1989). Testing and Modeling Threshold Autoregressive Processes. Journal of the American Statistical Association, 84(405), 231-240.

Xie H. (2018). Financial volatility modeling: The feedback asymmetric conditional autoregressive range model. Journal of Forecasting. 2019; 38:11–28.

Xie, H. B., & Wu, X. Y. (2017). A conditional autoregressive range model with gamma distribution for financial volatility modeling. Economic Modelling, 64, 349–356.

Zakoian, J. M. (1991). Threshold heteroskedastic models, Technical report, INSEE.

Zakoïan, J.M. (1994). Threshold Heteroskedastic Models. Journal of Economic Dynamics and Control, 18, 931-955.

Zhang, M. Y., Russell, J., & Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. Journal of Econometrics, 104,179–207.

IV. AN INTEGER GARCH MODEL FOR POISSON PROCESSES WITH TIME VARYING ZERO INFLATION

ABSTRACT

A time-varying zero inflated Poisson process is proposed to model a time series of count data with serial dependence. The model assumes that the intensity of the underlying Poisson Process evolves according to a generalized conditional heteroskedastic (GARCH) type model. The proposed model is a generalization of the zero-inflated Poisson Integer GARCH model, as proposed by Fukang Zhu in 2012, which in return can be considered a generalization of the Integer GARCH (INGARCH) model proposed by Ferland, Latour, and Oraichi in 2006. The proposed model builds on these previous formulations by incorporating flexibility to allow the zero inflation parameter to vary over time, according to a deterministic function or to be driven by an exogenous variable. Both the Expectation Maximization (EM) and the Maximum Likelihood Estimation (MLE) approaches were presented as possible estimation methods. A simulation study showed that both parameter estimation methods provided good estimates. Application to two real-life data sets showed that the proposed INGARCH model with the time varying zero inflated component provided a better fit than the traditional zero-inflated INGARCH model.

Key Words: Poisson Process, Integer-valued Time Series, Count Data, GARCH models, Periodicity, Zero Inflation.

1. INTRODUCTION

The standard Poisson point process, which assumes statistical independence between observations, is not suitable for modeling the time series of counts that display serial dependence. One way to address this deficiency is to define a Poisson process with its conditional mean at a given time dependent on the past conditional means and or past counts. Rydberg and Shepard (1999) proposed such a model, where the current conditional mean is a linear function of both the observed count and the conditional mean at the pervious time point. Similar models were also proposed by other authors, and these were discussed in Chapter 4 of the book by Kedem and Fokianos (2002). Andreas Heinen (2003) generalized the lag one model of Rydberg and Shepard to include an arbitrary number of lags for both the past counts and past means and named it the Autoregressive Conditional Poisson model with lags p and q (ACP (p, q)). The formulation of this model resembles that of a generalized conditional heteroscedastic (GARCH) model, but unlike the GARCH formulation that models the conditional variance of the process, the ACP models the conditional mean. Heinnen, however, derived the properties of his model only for the ACP (1, 1) case and properties for the general case was investigated by Gharamani and Thavaneswaran (2009), who referred to the Heinen paper as the origin of the ACP model. Independently, Ferland, Latour, and Oraichi (2006) proposed to the authors what he termed the Integer GARCH (INGARCH) process, which is essentially the same as the ACP model. The INGARCH (or ACP) model, however, did not accommodate zero inflation, and Zhu (2012a) proposed a zero inflated INGARCH formulation to incorporate the ability to model count data with zero counts that cannot be fitted well by the regular INGARCH model. The zero inflation probability in Zhu's model is constant over time, which is a drawback in situations where the relative number of zero counts tends to vary seasonally or with time. The proposed time-varying zero inflated Poisson INGARCH model (TVZIP-INGARCH) was developed to address this shortcoming.

Modelling integer-valued time series was discussed extensively in a broad range of disciplines, such as biostatistics and finance. Integer-valued Autoregressive (INAR) Poisson models were introduced by McKenzie (1985) as well as by Al-Osh and Alzaid (1987) for analyzing equidispersed count data with serial correlation. Quddus (2008) conducted an empirical study using an INAR model to analyze traffic accidents in Great Brittan and compared the performance with the results from fitting a real-valued Autoregressive Moving Average (ARMA) model; he found that the INAR Poisson model performed well when the counts are relatively low. Another formulation developed for analyzing count data time series was the Generalized Linear Autoregressive Moving Average (GLARMA) model (Davis et al., 2003). In this model, the conditional mean of the Poisson process depends on the past count information as well as current and past values of explanatory variables. In their approach, the logarithm of the conditional mean of the Poisson process expressed a linear function of covariates, with the noise process rewritten as an ARMA model. As mentioned before, the Autoregressive Conditional Poisson (ACP) model, which is similar to the observation driven GLARMA model, as introduced by Andreas Heinen (2003), addresses the common issues in time series count data such as discreetness, over dispersion and serial correlation. In ACP models, counts are assumed to be generated via a Poisson distribution with its conditional mean (intensity) obeying an autoregressive process expressed either by using the classical Generalized

Autoregressive Conditional Heteroskedastic (GARCH) model (Bollerslev, 1986) or the Autoregressive Conditional Duration (ACD) model (Engle and Russell, 1998). In contrast to the GLARMA model, the ACP formulation models the conditional mean of the Poisson process directly, rather than its logarithm. The unconditional variance of the count variable in the ACP model is higher than the unconditional expectation. This class of models accommodates both over dispersion and autocorrelation. As mentioned previously, the ACP model is exactly the same as the model proposed by Ferland et al. (2006). When discussing published literature on this topic, these models are referred to as ACP or INGARCH, interchangeably, based on what the authors of the cited works use in referring the model. When not discussing a particular publication, itis referred to as the INGARCH model.

The ACP formulation utilizes the classical GARCH (p, q) model to describe how the conditional mean of a Poisson process propagates over time. The ACP model (as well as the INGARCH Model) of order (p, q) is defined as follows:

$$
\begin{aligned}
&X_t \mid \mathrm{F}_{t-1} \sim P(\lambda_t); \forall t \in \mathbf{Z} \\
&\lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j},
\end{aligned}
\tag{1.1}
$$

where $\{X_t : t \in \square\}$ is the count process, λ_t defines the conditional mean of X_t given the past information, $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, i = 1,\ldots,p, j = 1,\ldots,q, p \geq 1, q \geq 0$ with the added condition $0 \leq \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. Note that some of the results presented in this paper are available in Heinen (2003) but are reported herein for completeness. In addition to presenting the model, Heinen also derived the stationarity conditions, covariance functions, and addressed the problem of maximum likelihood estimation (MLE) of the parameters.

Testing the parameters of a Poisson autoregressive model was considered by Zhu and Wang (2011). To address the scarcity of literature available in the general case of INGARCH (p, q) models, Weib (2008) extended the previous results and derived the set of Yule-Walker type equations for the autocorrelation function for the general INGARCH case. As pointed out by several authors (Weib, 2008 and Zhu 2012b), the INGARCH formulations area popular set of tools utilized in modeling the over-dispersion and serial dependency inherent to count data. Negative Binomial (NB), Generalized Poisson (GP), and Double Poisson (DP) are well known discrete distributions that can also be used as an alternative to the Poisson process (Zhu, 2012c). Combining such models with a zero inflation component is a natural step.

Zhu (2011) discussed the modelling of integer valued time series with over dispersion and handling potential extreme observations. Zhu (2012b) generalized the Poisson INGARCH process to handle both over dispersion and under dispersion cases. Further, he provided real examples for the proposed model. In this paper, Zhu used a maximum likelihood method to estimate the parameters. A negative binomial INGARCH model (NB-INGARCH), which is an alternative to the Poisson INGARCH model, was proposed and the stationary conditions and the autocorrelation function of the process were obtained by Ye, Garcia, Pourahmadi, and Lord (2012). These authors also allowed the negative binomial INGRACH model to incorporate covariates, so that the relationship between a time series of counts and correlated external factors could be properly modeled. Zhu (2012a) extended his previous work in Zhu (2011) and introduced the zero-inflated

Poisson and zero-inflated negative Binomial integer-valued GARCH models, and he showed how the EM algorithm can be used to estimate the parameters of the model. The underlying processes in Zhu's models are based on either zero-inflated Poisson or zero-inflated negative binomial, but they do not allow such zero inflation to be varied by time or influenced by any external factor.

Some empirical time series count data with large number of zero counts display strong cyclical behavior or seasonality with respect to the observed zero values. Ignoring this time varying property of the zero inflation parameter decreases the performance of the model. Recognizing this, Yang (2012), discussed the importance of modeling zero inflation as a time varying function. In his article, he assumed that both the zero inflation and the intensity parameter is driven by the linear combination of past observations of exogenous variables and connects them to the mean of the count data via a log link function.

We proposed a somewhat different approach. In our approach, it is the zero inflation probability, rather than the mean of the Poisson process that is allowed to be governed by exogenous variables. We also allow the zero inflation probability to be driven by a deterministic function, such as a sinusoidal wave. In addition, the intensity of the Poisson process is assumed to vary dynamically through a GARCH type model. Thus, the INGARCH part of the proposed model can be viewed as observation driven, in the sense that recursive substitutions can be employed to show that the current intensity of the process conditional on the past is a linear function of past observations.

The remainder of this paper is organized as follows: In Section 2, the Time Varying Zero Inflated Poisson INGARCH model (TVZIP-INGARCH) with a deterministic cyclically varying zero-inflation component is introduced; thereafter, the TVZIP-

INGARCH model in which the zero-inflation parameter will be driven by an exogenous set of stochastic variables is discussed. Parameter estimation procedures are presented in Section 3, which is followed by the simulation study presented in Section 4. Results and interpretations of the results based on fitting the proposed model to empirical data are presented in Section 5. A discussion and conclusions are presented in Section 6.

2. THE TIME VARYING ZERO INFLATED INGARCH MODEL

As Zhu (2012a) supplied, probability mass function (pmf) of a zero inflated Poisson model with parameter vector (λ, ω), with representing the count, can be written in the following form:

$$P(X=k)=\omega\delta_{k,0}+(1-\omega)\frac{\lambda^k e^{-\lambda}}{k!},\ k=0,1,2,\ldots,\text{ where } 0<\omega<1 \text{ and}$$

$$\delta_{k,0}=\begin{cases}1; k=0\\ 0; k\neq 0.\end{cases}$$

Further, Zhu (2012a) presented the mean and the variance of the distribution as follows:

$$E(X)=\lambda(1-\omega) \text{ and } Var(X)=\lambda(1-\omega)(1+\lambda\omega)>E(X) \text{ for } 0<\omega<1.$$

Moving on to define the time-varying zero inflated INGARCH model, $\{X_t : t\in\mathbb{Z}\}$ is a discrete time series of count data, and also F_{t-1} is the sigma field generated by $\{X_l : l\leq t-1\}$.The conditional distribution of X_t given F_{t-1} is described by a zero inflated Poisson distribution (*ZIP*) with parameter vector (λ_t, ω_t). Then, $X_t\,|\,\mathrm{F}_{t-1}\sim ZIP(\lambda_t, \omega_t)$ where,

$$P(X_t = k \mid \mathrm{F}_{t-1}) = \omega_t \delta_{k,0} + (1-\omega_t)\frac{\lambda_t^k e^{-\lambda_t}}{k!},$$

$$P(X_t = k \mid \mathrm{F}_{t-1}) = \begin{cases} \omega_t + (1-\omega_t)e^{-\lambda_t}; k = 0, \\ (1-\omega_t)\dfrac{\lambda_t^k e^{-\lambda_t}}{k!}; k \neq 0. \end{cases} \tag{2.1}$$

The dynamic propagation of the conditional mean of the Poisson process is defined by $\lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}$, where $\beta_j \geq 0; i = 1,2,3,...p; j = 1,2,3,...,q; p \geq 1, q \geq 1,$ and $t \in \square$. Furthermore, $\omega_t = g(\mathbf{V}_t, \underline{\Gamma}) \in (0,1)\ \forall\ t \in \square$, is a function of variables, propagating over time, which are later used to model the time varying zero inflation function. Note that elements of the vector $\mathbf{V}_t$ may consist of stochastic exogenous variables that vary with time, or it may be a scaler equal to time *t*. In addition, $\underline{\Gamma}$ denotes vector of parameters. It is assumed that $0 < \omega_t < 1$ for all $t \in \square$.

The above model is denoted by TVZIP-INGARCH (*p*, *q*). If $p > 0$ and $q = 0$, then the model becomes a TVZIP-INARCH model with order p or in abbreviated form TVZIP-INARCH (*p*). The conditional mean and conditional variance of X_t are given by the following equations:

$$E(X_t \mid \mathrm{F}_{t-1}) = (1-\omega_t)\lambda_t,\ Var(X_t \mid \mathrm{F}_{t-1}) = (1-\omega_t)\lambda_t(1+\omega_t\lambda_t). \tag{2.2}$$

See Appendix A., for the derivation of the conditional mean and conditional variance.

The conditional variance to conditional mean ratio or the dispersion ratio of ZIP distribution is:

$$\frac{Var(X_t \mid \mathrm{F}_{t-1})}{E(X_t \mid \mathrm{F}_{t-1})} = \frac{(1-\omega_t)\lambda_t(1+\omega_t\lambda_t)}{(1-\omega_t)\lambda_t} > (1+\lambda_t\omega_t). \tag{2.3}$$

The result in (2.3) indicates that TVZIP-INGARCH(p,q) can be used to model integer valued time series with over dispersion, if the values of ω_t bounded below by a positive constant.

2.1. CASE 1: ZERO INFLATION DRIVEN BY A DETERMINISTIC FUNCTION OF TIME

It is assumed that the zero inflation function $\omega_t = g(\mathbf{V}_t, \underline{\Gamma})$ is such $\mathbf{V}_t$ is a scaler equal to t. For illustrative purposed we will assume the function g is defined as follows:

$$\omega_t = g(\mathbf{V}_t, \underline{\Gamma}) = A\sin\left(\frac{2\pi}{S}t\right) + B\cos\left(\frac{2\pi}{S}t\right) + C \tag{2.4}$$

where S is the seasonal length, and $\underline{\Gamma} = \begin{pmatrix} A \\ B \\ C \end{pmatrix}$.

As mentioned above, the time-varying zero inflation function $\omega_t = g(\mathbf{V}_t, \underline{\Gamma})$ should always be bounded between zero and one. The values of A, B, and C needed to satisfy the above criterion are derived in Appendix B. Note that a simple example is used where the function g consists of a sine function and a cosine functions of equal period, but g could be any other function that, with proper selection of parameters, can be bounded between zero and one.

2.2. CASE 2: ZERO INFLATED FUNCTION DRIVEN BY AN EXOGENOUS VARIABLES

The above model also accommodates the case where the zero inflation probability is determined by one or more exogenous variables. In this case $g(\mathbf{V}_t, \underline{\Gamma})$ is considered a logistic regression function of the vector of exogenous variables $\mathbf{V}_t$, which is a scaler seasonal autoregressive time series, a vector seasonal time series, or a scaler or vector time series that varies non-seasonally. For illustrative purposes, consider the case where $\mathbf{V}_t$ is a scaler purely seasonal autoregressive time series, denoted by V_t, with period s. Then we can write,

$$V_t = \eta V_{t-s} + \varepsilon_t, \text{where } \varepsilon_t \text{ are } i.i.d. \sim N(0,1);$$

$$\omega_t = g(V_t, \underline{\Gamma}) = \frac{1}{1+e^{-(\delta_0+\delta_1 V_t)}}, \tag{2.5}$$

with $\underline{\Gamma} = \begin{pmatrix} \delta_0 \\ \delta_1 \end{pmatrix}$. Here δ_0 and $\delta_1 \in \mathbb{R}$.

3. ESTIMATION PROCEDURE

The use of both the Expectation Maximization (EM) algorithm and the Maximum Likelihood (ML) method to estimate the model parameters was discussed in the following. For brevity, only the TVZIP-INGARCH (1, 1) process was considered, but the procedure for the general case followed a similar manner, even though the computations would be more complex.

3.1. EXPECTATION MAXIMIZATION ESTIMATION FOR TVZIP-INGARCH (1, 1)

Let $X_1, X_2, \ldots, X_n$ be generated from the model (2.1). Following the formulation in Section 2, there are two types of zeros generated by this model. They are the zeroes coming from the Poisson distribution with the intensity parameter λ_t and the zeroes generated by a Bernoulli process with the probability of obtaining a zero given by the zero inflation parameter. Therefore, a given observation is hypothetically categorized as arising out of a Bernoulli process or is an observation from the Poisson distribution. Let us define $\{Z_t\}$ to be a Bernoulli random variable such that $Z_t = 1$ if X_t is a generated from the Bernoulli process and $Z_t = 0$ if it is generated by the Poisson distribution. Then: $Z_t \sim Bernoulli(\omega_t)$ with $\mathrm{P}(Z_t = 1) = \omega_t$ and $\mathrm{P}(Z_t = 0) = (1-\omega_t)$.

Also, let $Z = (Z_1, Z_2, \ldots, Z_n)$, $\Theta = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \beta_2, \ldots, \beta_q)^T = (\theta_0, \theta_1, \ldots, \theta_{p+q})$ and $\omega_t = g(\mathbf{V}_{t,}\underline{\Gamma})$.Note that $\Gamma = (\gamma_0, \gamma_1, \ldots, \gamma_r)$, where r is the dimension of the vector $\mathbf{V}_t$. For notational simplicity, we define the composite parameter vector $\Phi = (\Gamma^T, \Theta^T)^T = (\phi_1, \phi_2, \ldots, \phi_k) \subseteq R^{r+p+q+2}$, with the original parameters renamed as ϕ_j, j= 1, 2…k. This simplified notation is used in situations where generic statements are made without reference to a specific portion of (2.1).

Paralleling the derivations in Zhu (2012), the conditional log likelihood can be written as (see Appendix B for details),

$$l(\Phi) = \sum_{t=p+1}^{n} \left\{ Z_t \log(\omega_t) + (1-Z_t)\left[\log(1-\omega_t) + X_t \log(\lambda_t) - \lambda_t - \log(X_t!)\right]\right\}. \quad (3.1)$$

The first derivatives of the conditional log likelihood function (3.1) with respect to $\Gamma = (\gamma_0, \gamma_1, \ldots, \gamma_r)$ and $\Theta = (\theta_0, \theta_1, \ldots, \theta_{p+q})$ are as follows:

$$\frac{dl(\Phi)}{d\gamma_i}=\frac{dl(\Phi)}{d\omega_t}\frac{d\omega_t}{d\gamma_i}=\sum_{t=p+1}^{n}\left\{\frac{Z_t}{\omega_t}-\frac{(1-Z_t)}{(1-\omega_t)}\right\}\frac{d\omega_t}{d\gamma_i},\ i=0,1,\ldots,r, \tag{3.2}$$

$$\frac{dl(\Phi)}{d\theta_j}=\frac{dl(\Phi)}{d\lambda_t}\frac{d\lambda_t}{d\theta_j}=\sum_{t=p+1}^{n}(1-Z_t)\left\{\frac{X_t}{\lambda_t}-1\right\}\frac{d\lambda_t}{d\theta_j},\ j=0,1,\ldots,p+q. \tag{3.3}$$

Finally, by combining (3.2) and (3.3) the first derivative of the conditional log likelihood function with respect to Φ is given by:

$$\frac{dl(\Phi)}{d\phi_k}=\frac{dl(\Phi)}{d\omega_t}\frac{d\omega_t}{d\phi_k}+\frac{dl(\Phi)}{d\lambda_t}\frac{d\lambda_t}{d\phi_k};\phi_k\in\Phi=(\Gamma,\Theta),$$

$$\frac{d\omega_t}{d\phi_k}=0\text{, if }\phi_k\notin\Gamma\ \text{ and }\ \frac{d\lambda_t}{d\phi_k}=0\ \text{ if }\phi_k\notin\Theta. \tag{3.4}$$

The two step (E step and M step) Expectation Maximization algorithm is used to estimate the parameter vector $\Phi=\left(\Gamma^T,\Theta^T\right)^T$. Let $\tau_t=E(Z_t\mid X_t,\Phi)$ and we replace Z_t by $\hat{Z}_t=\tau_t$, and define $Z=(Z_1,Z_2,\ \ldots\ ,Z_n)^T$. Following this replacement of Z in the log likelihood function, $l\left(\Phi,\hat{Z}\right)$ will be maximized.

E Step: Find τ_t using the equation:

$$\tau_t=\begin{cases}\dfrac{\omega_t}{\omega_t+(1-\omega_t)e^{-\lambda_t}} & : X_t=0\\ 0 & : X_t>0.\end{cases}$$

M Step: After Z_t is replaced by its estimate, we proceed to maximize $l\left(\Phi,\hat{Z}\right)$. First set

$$\frac{dl(\Phi)}{d\phi_k}=0\quad\text{for all } k\ .$$

If $\hat{\Phi}$ is the solution to the system of equations in (3.4) exists, then $S\left(\hat{\Phi}\right)=\underline{0}$, where $S(\Phi)$ is the fishers score matrix, and $\hat{\Phi}$is the vector that minimizes the log likelihood providing us with the estimate of the parameter vector $\Phi=\left(\Gamma^T,\Theta^T\right)^T$.

Since a closed form solution does not exist, we require an iterative procedure to find the estimates. Let us consider the first order Taylor expansion of $S\left(\tilde{\Phi}\right)$ evaluated at the value $\tilde{\Phi}$ around the initial parameter values Φ_0, yielding. $S\left(\tilde{\Phi}\right)\approx S\left(\Phi_0\right)+\dfrac{dS\left(\Phi\right)}{d\phi}\left(\tilde{\Phi}-\Phi_0\right)$. We also let the matrix of the second derivatives of the log likelihood function is defined as $H\left(\Phi\right)=\dfrac{d^2l\left(\Phi\right)}{d\Phi d\Phi^T}=\dfrac{dS\left(\Phi\right)}{d\Phi}$.

From the above, we obtain the first order approximation $\tilde{\Phi}=\Phi_0-H^{-1}\left(\Phi_0\right)S\left(\Phi_0\right)$, and this result provides the standard Newton-Raphson algorithm. For an appropriate chosen initial value Φ_0, the above Newton Raphson algorithm is used to obtain a sequence of improved estimates recursively. The improved estimates at i^{th} iteration are updated as the initial values for the next iteration as follows:

$$\hat{\Phi}^{(i+1)}=\hat{\Phi}^{(i)}-H^{-1}\left(\hat{\Phi}^{(i)}\right)S\left(\hat{\Phi}^{(i)}\right).$$

This Process is repeated until the differences between successive estimates are sufficiently close to zero. In our study, convergence of the EM procedure is determined by using the following criterion:

$$\left|\left(\hat{\phi}_j^{(i+1)}-\hat{\phi}_j^{(i)}\right)\Big/\left(\hat{\phi}_j^{(i)}\right)\right|\leq 10^{-6}.$$

3.2. MAXIMUM LIKELIHOOD ESTIMATION FOR THE TVZIP-INGARCH (1, 1)

The conditional likelihood function $L(\Phi)$ of the TVZIP-INGARCH model (2.1) is,

$$L(\Phi)=\prod\nolimits_{X(t)=0}\left[\omega_t+(1-\omega_t)e^{-\lambda_t}\right]\prod\nolimits_{X(t)>0}\left[(1-\omega_t)\frac{\lambda_t^{X_t}e^{-\lambda_t}}{X_t!}\right]. \tag{3.5}$$

The conditional log likelihood function, $l(\Phi)$ obtained from (3.5) is given by

$$l(\Phi)=\sum_{X(t)=0}\log\left|\omega_t+(1-\omega_t)e^{-\lambda_t}\right|+\sum_{X(t)>0}\left[\log(1-\omega_t)+X_t\log(\lambda_t)-\lambda_t-\log(X_t!)\right]. \tag{3.6}$$

Let $P_{0,t}=\omega_t+(1-\omega_t)e^{-\lambda_t}$ and $I\left(X(t)=0\right)=x_{0,t}$. Then,

$$\frac{dl(\Phi)}{d\omega_t}=\sum_{t=1}^{N}\left[\frac{x_{0,t}\left(1-e^{-\lambda_t}\right)}{P_{0,t}}-\frac{\left(1-x_{0,t}\right)}{\lambda_t}\right], \tag{3.7}$$

and

$$\frac{dl(\Phi)}{d\lambda_t}=\sum_{t=1}^{N}\left[\frac{x_{0,t}\omega_t}{P_{0,t}}+\frac{\left(X_t-\lambda_t\right)}{\lambda_t}\right]. \tag{3.8}$$

The first derivatives of the conditional log likelihood function (3.6) are as follows,

$$\frac{dl(\Phi)}{d\gamma_i}=\frac{dl(\Phi)}{d\omega_t}\frac{d\omega_t}{d\gamma_i}=\sum_{t=1}^{N}\left[\frac{x_{0,t}\left(1-e^{-\lambda_t}\right)}{P_{0,t}}-\frac{\left(1-x_{0,t}\right)}{\lambda_t}\right]\frac{d\omega_t}{d\gamma_i},i=0,1,\ldots,r, \tag{3.9}$$

$$\frac{dl(\Phi)}{d\theta_j}=\frac{dl(\Phi)}{d\lambda_t}\frac{d\lambda_t}{d\theta_j}=\sum_{t=1}^{N}\left[\frac{x_{0,t}\omega_t}{P_{0,t}}+\frac{\left(X_t-\lambda_t\right)}{\lambda_t}\right]\frac{d\lambda_t}{d\theta_j},j=0,1,\ldots,p+q, \tag{3.10}$$

$$\frac{dl(\Phi)}{d\phi_k}=\frac{dl(\Phi)}{d\omega_t}\frac{d\omega_t}{d\phi_k}+\frac{dl(\Phi)}{d\lambda_t}\frac{d\lambda_t}{d\phi_k},\phi_k\in\Phi=(\Gamma,\Theta),$$

$$\frac{d\omega_t}{d\phi_k}=0\text{, if }\phi_k\notin\Gamma\ \ \phi_k\notin\Gamma\text{ and }\frac{d\lambda_t}{d\phi_k}=0\text{, if }\phi_k\notin\Theta. \tag{3.11}$$

We can use Newton-Raphson (NR) iterative procedure to obtain the maximum likelihood estimated for the equation (3.5) by setting $\frac{dl(\Phi)}{d\phi_k} = 0$ for all k. With a reasonable initial starting value $\hat{\Phi}^{(0)}$, the i^{th} iteration is calculated using

$$\hat{\Phi}^{(i+1)} = \hat{\Phi}^{(i)} - H^{-1}\left(\hat{\Phi}^{(i)}\right) S\left(\hat{\Phi}^{(i)}\right), \text{ where } S\left(\hat{\Phi}\right) = \frac{dl(\Phi)}{d\phi_k}\Big|\hat{\Phi} \text{ and } H\left(\hat{\Phi}\right) = \frac{dl(\Phi)}{d\phi_k d\phi_k^{\,T}}\Big|\hat{\Phi}.$$

We stop the algorithm once pre specified convergence criteria is satisfied.

4. SIMULATION STUDY

We investigated the finite sample performance of estimators using a simulation study. We used 'poissrnd' function of Matlab software to generate the relevant data. Lengths of the time series studies were set to n=120 and n=360 and thousand (m=1000) simulations runs were carried out for each parameter sample size combination. We carried out two separate sets of simulation studies based on the zero inflation function was introduced in Section 2. We maximized the profile likelihood functions (3.1) and (3.5) using the constrained nonlinear optimization function 'fmincon' in Matlab. The zero inflation $\left(\omega_t = g\left(\mathbf{V}_t, \Gamma\right)\right)$ was allowed to vary cyclically or to be driven by an exogenous variable. Similar to the work Zhu (2012a) presents, the Mean Absolute Deviation Error (MADE) was utilized as the evaluation criterion. The MADE is defined as, $\frac{1}{m}\sum_{j=1}^{m}\left|\hat{\phi}_j - \phi_j\right|$ where m is the number of replications. Simulation results are reported in Tables 1 through Table 12.

4.1. SIMULATION RESULTS FOR CASE 1: SINUSOIDAL ZERO INFLATION FUNCTION

In this simulation study, the sinusoidal zero inflated function $\omega_t = g(\mathbf{V}_t, \Gamma)$ expressed in Equation (2.4) was used to generate cyclically varying zero inflation probabilities between zero and one. We set the following constraints to the parameter vector $\Gamma = \begin{pmatrix} A \\ B \\ C \end{pmatrix}$:

$$\omega_t = g(\Gamma, S) = A\sin\left(\frac{2\pi}{S}t\right) + B\cos\left(\frac{2\pi}{S}t\right) + C,$$

where $C = \sqrt{A^2 + B^2} + \delta > \sqrt{A^2 + B^2}$, and $A^2 + B^2 < \frac{1}{4}$ where $|A| < \frac{1}{2}$, $|B| < \frac{1}{2}$ and δ is a fixed value such that $\delta > 0$.Note that the above constraints were applied to bound the zero inflation probabilities between 0 and 1.

Tables 1 through 3 provide the simulation results for the MLE estimation techniques, while Tables 4 through 6 provide simulation results for the case where estimates were obtained using the EM algorithm. The frequency of the sinusoidal wave was set at S=12, mimicking a 12 month cycle in present in monthly data. The parameter vector for the simulation study was expressed as $\Phi = (A, B, \alpha_0, \alpha_1, \alpha_2, \beta_1)$, where A and B are the parameters in the sinusoidal model while (α_0, α_1), $(\alpha_0, \alpha_1, \alpha_2)$ and $(\alpha_0, \alpha_1, \beta_1)$ are the parameter combinations in TVZIP-INARCH (1), TVZIP-INARCH (2), TVZIP-INGARCH (1, 1), respectively. The parameter combination of $\Gamma = (A, B)^T$ was set at (0.10, 0.10), (0.25,-0.20) and (-0.35,-0.30) and represented, minimal to minimal, minimal to moderate, and minimal to maximum zero inflation, while maintaining a fixed $\delta = 0.0001$.

The following models were considered:

(A) TVZIP-INARCH (1) models: $\Phi = (A, B, \alpha_0, \alpha_1)$

A1. (0.10, 0.10, 1.00, 0.40)

A2. (-0.25, -0.25, 2.00, 0.50)

A3. (-0.35, -0.30, 1.00, 0.70)

(B) TVZIP-INARCH (2) models: $\Phi = (A, B, \alpha_0, \alpha_1, \alpha_2)$

B1. (0.10, 0.10, 1.00, 0.20, 0.20)

B2. (-0.25, -0.25, 2.00, 0.30, 0.20)

B3. (-0.35, -0.30, 1.00, 0.40, 0.30)

(C) TVZIP-INGARCH (1,1) models: $\Phi = (A, B, \alpha_0, \alpha_1, \beta_1)$

C1. (0.10, 0.10, 1.00, 0.20, 0.20)

C2. (-0.25, -0.25, 2.00, 0.30, 0.20)

C3. (-0.35, -0.30, 1.00, 0.40, 0.30)

Based on the simulation study results, it was observed that both EM and MLE procedures produced similar results for the estimated values of parameter sets in TVZIP-INGARCH (1) and TVZIP-INARCH (2). However, in some parameter combinations for TVZIP-INGARCH (1, 1) process, there was a slight difference between the MLE and EM estimated results. In general, bigger sample sizes produced more accurate results. The MADE for the time varying Zero inflation parameters (A, B) were relatively lower than

Table 1: Means of MLE estimates and MADE (within parentheses), for TVZIP-INARCH (1) models where zero inflation is driven by sinusoidal function

Model	n	A	B	α_0	α_1
True Parameters		0.10	0.10	1.00	0.40
A1	120	0.0893 (0.0561)	0.0845 (0.0566)	1.0472 (0.1433)	0.3712 (0.0903)
	360	0.0986 (0.0321)	0.0951 (0.0299)	1.0172 (0.0802)	0.3917 (0.0479)
True Parameters		-0.25	-0.25	2.00	0.50
A2	120	-0.2488 (0.0401)	-0.2475 (0.0399)	2.0467 (0.2210)	0.4800 (0.0873)
	360	-0.2504 (0.0225)	-0.2478 (0.0223)	2.0203 (0.1311)	0.4925 (0.0484)
True Parameters		-0.35	-0.30	1.00	0.70
A3	120	-0.3468 (0.0457)	-0.2959 (0.0482)	1.0369 (0.1621)	0.6675 (0.1163)
	360	-0.3508 (0.0250)	-0.2970 (0.0263)	1.0178 (0.0963)	0.6875 (0.0629)

Table 2: Means of MLE estimates and MADE (within parentheses), for TVZIP-INARCH (2) models where zero inflation is driven by sinusoidal function

Model	n	A	B	α_0	α_1	α_2
True Parameters		0.10	0.10	1.00	0.20	0.20
B1	120	0.0872 (0.0559)	0.0838 (0.0538)	1.0529 (0.1602)	0.1877 (0.0873)	0.1765 (0.0864)
	360	0.0955 (0.0317)	0.0971 (0.0290)	1.0239 (0.0945)	0.1943 (0.0496)	0.1906 (0.0503)
True Parameters		-0.25	-0.25	2.00	0.30	0.20
B2	120	-0.2485 (0.0430)	-0.2476 (0.0395)	2.0524 (0.2563)	0.2842 (0.0952)	0.1906 (0.0931)
	360	-0.2514 (0.0234)	-0.2470 (0.0224)	2.0254 (0.1478)	0.2989 (0.0500)	0.1901 (0.0536)
True Parameters		-0.35	-0.30	1.00	0.40	0.30
B3	120	-0.3478 (0.0455)	-0.2941 (0.0486)	1.0357 (0.1712)	0.3839 (0.1305)	0.2781 (0.1324)
	360	-0.3510 (0.0282)	-0.2974 (0.0273)	1.0149 (0.1004)	0.3975 (0.0709)	0.2914 (0.0762)

Table 3: Means of MLE estimates and MADE (within parentheses), for TVZIP-INGARCH (1, 1) models where zero inflation is driven by sinusoidal function

Model	n	A	B	α_0	α_1	β_1
True Parameters		0.10	0.10	1.00	0.20	0.20
C1	120	0.0939 (0.0521)	0.0882 (0.0541)	0.9420 (0.2205)	0.2226 (0.0789)	0.2078 (0.1399)
	360	0.1057 (0.0300)	0.0983 (0.0295)	0.9511 (0.1544)	0.2295 (0.0554)	0.1977 (0.1162)
True Parameters		-0.25	-0.25	2.00	0.30	0.20
C2	120	-0.2522 (0.0396)	-0.2472 (0.0402)	1.7725 (0.3498)	0.3609 (0.0979)	0.1929 (0.1178)
	360	-0.2516 (0.0231)	-0.2500 (0.0221)	1.8088 (0.2330)	0.3843 (0.0884)	0.1540 (0.0859)
True Parameters		-0.35	-0.30	1.00	0.40	0.30
C3	120	-0.3589 (0.0414)	-0.2937 (0.0456)	0.9463 (0.2131)	0.4666 (0.1294)	0.2327 (0.1689)
	360	-0.3591 (0.0259)	-0.2984 (0.0265)	0.9426 (0.1442)	0.4958 (0.1088)	0.2022 (0.1372)

Table 4: Means of EM estimates and MADE (within parentheses), for TVZIP-INARCH (1) models where zero inflation is driven by sinusoidal function

Model	n	A	B	α_0	α_1
True Parameters		0.10	0.10	1.00	0.40
A1	120	0.0898 (0.0557)	0.0850 (0.0561)	1.0472 (0.1433)	0.3712 (0.0903)
	360	0.0986 (0.0321)	0.0951 (0.0299)	1.0172 (0.0802)	0.3917 (0.0479)
True Parameters		-0.25	-0.25	2.00	0.50
A2	120	-0.2487 (0.0401)	-0.2476 (0.0398)	2.0467 (0.2210)	0.4800 (0.0873)
	360	-0.2504 (0.0225)	-0.2478 (0.0222)	2.0203 (0.1311)	0.4925 (0.0484)
True Parameters		-0.35	-0.30	1.00	0.70
A3	120	-0.3468 (0.0457)	-0.2959 (0.0482)	1.0369 (0.1621)	0.6675 (0.1164)
	360	-0.3508 (0.0250)	-0.2970 (0.0263)	1.0178 (0.0963)	0.6876 (0.0629)

Table 5: Means of EM estimates and MADE (within parentheses), for TVZIP-INARCH (2) models where zero inflation is driven by sinusoidal function

Model	n	A	B	α_0	α_1	α_2
True Parameters		0.10	0.10	1.00	0.20	0.20
B1	120	0.0875 (0.0556)	0.0841 (0.0536)	1.0529 (0.1601)	0.1877 (0.0874)	0.1765 (0.0863)
	360	0.0955 (0.0317)	0.0972 (0.0290)	1.0239 (0.0945)	0.1943 (0.0496)	0.1906 (0.0503)
True Parameters		-0.25	-0.25	2.00	0.30	0.20
B2	120	-0.2485 (0.0430)	-0.2476 (0.0395)	2.0524 (0.2563)	0.2842 (0.0952)	0.1906 (0.0931)
	360	-0.2514 (0.0234)	-0.2470 (0.0224)	2.0254 (0.1478)	0.2989 (0.0500)	0.1901 (0.0536)
True Parameters		-0.35	-0.30	1.00	0.40	0.30
B3	120	-0.3480 (0.0453)	-0.2940 (0.0485)	1.0358 (0.1712)	0.3839 (0.1305)	0.2781 (0.1324)
	360	-0.3510 (0.0282)	-0.2974 (0.0273)	1.0149 (0.1004)	0.3975 (0.0709)	0.2914 (0.0762)

Table 6: Means of EM estimates and MADE (within parentheses), for TVZIP-INGARCH (1, 1) models where zero inflation is driven by sinusoidal function

Model	n	A	B	α_0	α_1	β_1
True Parameters		0.10	0.10	1.00	0.20	0.20
C1	120	0.0952 (0.0534)	0.0898 (0.0558)	0.9376 (0.2188)	0.2218 (0.0793)	0.2142 (0.1415)
	360	0.1058 (0.0301)	0.0983 (0.0296)	0.9436 (0.1589)	0.2292 (0.0556)	0.2042 (0.1206)
True Parameters		-0.25	-0.25	2.00	0.30	0.20
C2	120	-0.2515 (0.0395)	-0.2466 (0.0404)	1.7578 (0.3261)	0.3616 (0.0972)	0.1936 (0.1175)
	360	-0.2516 (0.0231)	-0.2500 (0.0221)	1.8070 (0.2347)	0.3843 (0.0884)	0.1539 (0.0859)
True Parameters		-0.35	-0.30	1.00	0.40	0.30
C3	120	-0.3583 (0.0415)	-0.2931 (0.0453)	0.9280 (0.2186)	0.4671 (0.1288)	0.2364 (0.1683)
	360	-0.3590 (0.0258)	-0.2984 (0.0264)	0.9350 (0.1494)	0.4961 (0.1086)	0.2018 (0.1367)

that of the INGARCH (or INARCH) related parameters $(\alpha_0, \alpha_1, \alpha_2, \beta_1)$ in the model. Therefore, parameter combination of A and B were estimated with higher accuracy, while other parameter estimates were reasonable. This phenomenon was common for all three minimal to minimal, minimal to moderate and minimal to maximum zero inflation cases.

4.2. SIMULATION STUDY FOR CASE 2: ZERO INFLATION FUNCTION DRIVEN BY AN EXOGENOUS VARIABLE

In this study, we allowed the exogenous variable to generate zeros through a logistic model as described in Equation (2.5). The parameter vector for the simulation study under this scenario was $\Phi = (\delta_0, \delta_1, \alpha_0, \alpha_1, \alpha_2, \beta_1)$, where δ_0 and δ_1 are the parameters in the logistic model, while (α_0, α_1) $(\alpha_0, \alpha_1, \alpha_2)$ and $(\alpha_0, \alpha_1, \beta_1)$ are the parameter combination in TVZIP-INARCH (1), TVZIP-INGARCH (2) and TVZIP-INGARCH (1, 1) models, respectively. The parameter combination δ_0 and δ_1 were set to (-2, 0), (-1, -1) and (2, 1) representing minimal, moderate, and large zero inflations, respectively. We generated an exogenous stationary AR (12) time series using $\eta = 0.25$, and the following models were considered:

(A) TVZIP-INARCH (1) models: $\Phi = (\delta_0, \delta_1, \alpha_0, \alpha_1)$

A1. (-2.00, 0.00, 1.00, 0.40)

A2. (-1.00, -1.00, 2.00, 0.50)

A3. (2.00, 1.00, 1.00, 0.70)

(B) TVZIP-INARCH (2) models: $\Phi = (\delta_0, \delta_1, \alpha_0, \alpha_1, \alpha_2)$

B1. (-2.00, 0.00, 1.00, 0.20, 0.20)

B2. (-1.00, -1.00, 2.00, 0.30, 0.20)

B3. (2.00, 1.00, 1.00, 0.40, 0.30)

(C) TVZIP-INGARCH (1,1) models: $\Phi = (\delta_0, \delta_1, \alpha_0, \alpha_1, \beta_1)$

C1. (-2.00, 0.00, 1.00, 0.20, 0.20)

C2. (-1.00, -1.00, 2.00, 0.30, 0.20)

C3. (2.00, 1.00, 1.00, 0.40, 0.30)

Tables 7 through 9 provide the simulation results for the MLE estimation techniques, while Tables 10 through 12 provide EM (Expectation Maximization) algorithm simulation results.

Table 7: Means of MLE estimates and MADE (within parentheses), for TVZIP-INARCH (1) models where zero inflation is driven by exogenous variable

Model	n	δ_0	δ_1	α_0	α_1
True Parameters		-2.00	0.00	1.00	0.40
A1	120	-2.4301 (0.7134)	-0.0331 (0.5515)	1.0169 (0.1381)	0.3808 (0.0837)
	360	-2.0982 (0.3258)	-0.0162 (0.2559)	1.0143 (0.0817)	0.3892 (0.0487)
True Parameters		-1.00	-1.00	2.00	0.50
A2	120	-1.0467 (0.2382)	-1.0711 (0.2580)	2.0193 (0.2143)	0.4864 (0.0836)
	360	-1.0209 (0.1338)	-1.0182 (0.1400)	2.0073 (0.1199)	0.4932 (0.0476)
True Parameters		2.00	1.00	1.00	0.70
A3	120	2.0047 (0.4342)	1.1592 (0.4153)	0.9914 (0.2978)	0.4605 (0.3699)
	360	1.9976 (0.2197)	1.0365 (0.1947)	1.0037 (0.1694)	0.5662 (0.2739)

Table 8: Means of MLE estimates and MADE (within parentheses), for TVZIP-INARCH (2) models where zero inflation is driven by exogenous variable

Model	n	δ_0	δ_1	α_0	α_1	α_2
True Parameters		-2.00	0.00	1.00	0.20	0.20
B1	120	-2.4576 (0.7430)	-0.0369 (0.5504)	1.0217 (0.1560)	0.1947 (0.0741)	0.1846 (0.0729)
	360	-2.1479 (0.3557)	0.0024 (0.2779)	1.0147 (0.0950)	0.1956 (0.0474)	0.1904 (0.0453)
True Parameters		-1.00	-1.00	2.00	0.30	0.20
B2	120	-1.0505 (0.2390)	-1.0593 (0.2436)	2.0354 (0.2494)	0.2911 (0.0829)	0.1930 (0.0721)
	360	-1.0132 (0.1258)	-1.0191 (0.1345)	2.0092 (0.1418)	0.2989 (0.0476)	0.1977 (0.0453)
True Parameters		2.00	1.00	1.00	0.40	0.30
B3	120	1.9814 (0.4398)	1.1959 (0.4396)	0.9418 (0.2881)	0.3127 (0.2752)	0.2656 (0.2253)
	360	1.9801 (0.2139)	1.0532 (0.1973)	0.9797 (0.1665)	0.3584 (0.2285)	0.2952 (0.1940)

Table 9: Means of MLE estimates and MADE (within parentheses), for TVZIP-INGARCH (1, 1) models where zero inflation is driven by exogenous variable

Model	n	δ_0	δ_1	α_0	α_1	β_1
True Parameters		-2.00	0.00	1.00	0.20	0.20
C1	120	-2.1527 (0.4716)	-0.0183 (0.3284)	0.9318 (0.2158)	0.2257 (0.0790)	0.2090 (0.1381)
	360	-2.0880 (0.3031)	0.0025 (0.2342)	0.9411 (0.1509)	0.2280 (0.0512)	0.1887 (0.1070)
True Parameters		-1.00	-1.00	2.00	0.30	0.20
C2	120	-1.0226 (0.2184)	-1.0412 (0.2332)	1.7477 (0.3285)	0.3924 (0.1142)	0.1620 (0.1008)
	360	-1.0083 (0.1244)	-1.0131 (0.1340)	1.8084 (0.2163)	0.3992 (0.1021)	0.1337 (0.0869)
True Parameters		2.00	1.00	1.00	0.40	0.30
C3	120	1.9804 (0.3863)	1.0538 (0.2881)	0.8406 (0.3229)	0.3927 (0.3003)	0.2271 (0.2395)
	360	1.9927 (0.2193)	1.0375 (0.1920)	0.8296 (0.2337)	0.4782 (0.2648)	0.2271 (0.2242)

Table 10: Means of EM estimates and MADE (within parentheses), for TVZIP-INARCH (1) models where zero inflation is driven by exogenous variable

Model	n	δ_0	δ_1	α_0	α_1
True Parameters		-2.00	0.00	1.00	0.40
A1	120	-2.5278 (0.8110)	-0.0402 (0.5931)	1.0160 (0.1386)	0.3806 (0.0837)
	360	-2.1031 (0.3307)	-0.0154 (0.2575)	1.0144 (0.0817)	0.3892 (0.0487)
True Parameters		-1.00	-1.00	2.00	0.50
A2	120	-1.0467 (0.2382)	-1.0711 (0.2580)	2.0193 (0.2143)	0.4864 (0.0836)
	360	-1.0209 (0.1338)	-1.0182 (0.1400)	2.0073 (0.1198)	0.4932 (0.0476)
True Parameters		2.00	1.00	1.00	0.70
A3	120	2.0083 (0.4307)	1.1504 (0.4065)	0.9923 (0.2968)	0.4606 (0.3699)
	360	1.9976 (0.2196)	1.0365 (0.1947)	1.0037 (0.1694)	0.5662 (0.2739)

Table 11: Means of EM estimates and MADE (within parentheses), for TVZIP-INARCH (2) models where zero inflation is driven by exogenous variable

Model	n	δ_0	δ_1	α_0	α_1	α_2
True Parameters		-2.00	0.00	1.00	0.20	0.20
B1	120	-2.1663 (0.4741)	-0.0180 (0.3259)	1.0296 (0.1554)	0.1957 (0.0746)	0.1852 (0.0730)
	360	-2.0915 (0.3012)	0.0072 (0.2335)	1.0165 (0.0946)	0.1958 (0.0474)	0.1907 (0.0453)
True Parameters		-1.00	-1.00	2.00	0.30	0.20
B2	120	-1.0505 (0.2390)	-1.0593 (0.2436)	2.0354 (0.2494)	0.2911 (0.0829)	0.1930 (0.0721)
	360	-1.0132 (0.1258)	-1.0191 (0.1345)	2.0092 (0.1418)	0.2989 (0.0476)	0.1977 (0.0453)
True Parameters		2.00	1.00	1.00	0.40	0.30
B3	120	1.9865 (0.4348)	1.1890 (0.4326)	0.9420 (0.2878)	0.3137 (0.2752)	0.2649 (0.2248)
	360	1.9801 (0.2139)	1.0532 (0.1973)	0.9797 (0.1655)	0.3585 (0.2285)	0.2944 (0.1936)

Table 12: Means of EM estimates and MADE (within parentheses), for TVZIP-INGARCH (1, 1) models where zero inflation is driven by exogenous variable

Model	n	δ_0	δ_1	α_0	α_1	β_1
True Parameters		-2.00	0.00	1.00	0.20	0.30
C1	120	-2.1170 (0.4717)	-0.0190 (0.3224)	0.9289 (0.2108)	0.2251 (0.0794)	0.2120 (0.1371)
	360	-2.1166 (0.3504)	0.0029 (0.2631)	0.9347 (0.1536)	0.2275 (0.0514)	0.1925 (0.1082)
True Parameters		-1.00	-1.00	2.00	0.30	0.20
C2	120	-1.0236 (0.2191)	-1.0413 (0.2330)	1.7423 (0.3318)	0.3924 (0.1141)	0.1612 (0.0994)
	360	-1.0083 (0.1244)	-1.0131 (0.1340)	1.8084 (0.2163)	0.3992 (0.1021)	0.1337 (0.0869)
True Parameters		2.00	1.00	1.00	0.40	0.30
C3	120	1.9798 (0.3665)	1.0452 (0.2845)	0.8236 (0.3205)	0.3963 (0.3005)	0.2297 (0.2359)
	360	1.9871 (0.2187)	1.0339 (0.1909)	0.8092 (0.2478)	0.4805 (0.2631)	0.2202 (0.2147)

Both EM and MLE methods produced fairly accurate estimates for the parameters across all three types of models, especially for those parameters associated with zero inflation function. In general, for larger sample sizes, all the three models produce more accurate estimates for both methods. In the moderate zero inflation case, estimates for the parameter combination of (δ_0, δ_1) had higher accuracy (i.e., lower MADE) than the minimal and maximum zero inflation cases, and this behavior persisted for all three TVZIP-INARCH (1), TVZIP-INARCH (2) and TVZIP-INGARCH (1, 1) models.

In the next sections, we compared the model selection performance of two versions of ZIP-INGARCH (p, q) models namely, Zhu's Fixed ZIP-INGARCH (p, q) model and proposed TVZIP-INGARCH (p, q) model. The underline model can either be FZIP-

INGARCH (p, q) or the proposed TVZIP-INGARCH (p, q) model. For an example a time series for FZIP-INGARCH (p, q) model was generated with sample size $n=360$ using 'poissrnd' function of Matlab software. Using the generated data, both the FZIP-INGARCH (p, q) and the TVZIP-INGARCH (p, q), were fitted to the data, and the model parameters were estimated for both the models using the EM algorithm. Next, information criteria were used to compare the model performance in this proposed study. Thereafter, the AIC values were compared for FZIP-INGARCH (p, q) and TVZIP-INGARCH (p, q) models; and thereby, the model with the lowest AIC value was selected. Likewise, we also simulated $m=1000$ time series and measure the accuracy as a percentage defined below:

$$Accuracy = \frac{c}{m} * 100\% \tag{4.1}$$

Here, c is the number of correct classifications and m is the total number of time series generated. The same procedure was repeated for a simulated dataset using the TVZIP-INGARCH (p, q) model. Simulation results are reported in Tables 13 through Table 16.

4.3. MODEL SELECTION FOR CASE 1: SINUSOIDAL ZERO INFLATION FUNCTION

In this section, time series data were generated using two time series. First, Zhu's FZIP-INGARCH (p, q) model was used to generate the data and then fitted both FZIP-INGARCH (p, q) and TVZIP-INGARCH (p, q) process to the data. Then, checked which process fitted better to the simulated data based on the AIC values. In similar way, we also consider TVZIP-INGARCH (p, q) model as the underline process to generate time series data and tried to fit both the time varying and the constant INGARCH models to discern the number of correctly classified cases based on AIC values. Finally, the accuracy was calculated using the equation 4.1, for both aforementioned methods. In this section, the sinusoidal zero inflation function $\omega_t = g\left(V_t, \Gamma\right)$, mentioned in the equation (2.4), was used to generate the zero inflated data for the proposed TVZIP-INGARCH (*p*, *q*) model. Table 13 provides the model selection results for underline FZIP-INGARCH (*p*, *q*) process, and Table 14 presents that of the TVZIP-INGARCH (*p*, *q*) models.

Table 13: Model selection accuracy for the underline FZIP-INGARCH (*p*, *q*) process

Model	FZIP-INGARCH (*p*, *q*) True Parameter (Φ)					Accuracy (%)
	ω	α_0	α_1	α_2	β	
FM1	0.15	1.00	0.90			74.7
FM2	0.25	2.00	0.50			99.7
FM3	0.80	2.00	0.70			100
FM4	0.15	1.00	0.20	0.70		86.8
FM5	0.25	2.00	0.20	0.50		99.9
FM6	0.80	2.00	0.10	0.80		100
FM7	0.15	2.00	0.30		0.20	95.1
FM8	0.25	2.00	0.40		0.20	99.8
FM8	0.80	2.00	0.30		0.10	100

According to the Table 13, there are nine time series models (FM1-FM9) and they can be mainly categorized into three types: FZIP-INARCH (1), FZIP-INARCH (2) and

FZIP-INGARCH (1, 1). Models listed under FZIP-INARCH (1) are FM1, FM2 and FM3. Models denoted by FM4, FM5 and FM6 belong to FZIP-INARCH (2). Models represented by FM7, FM8 and FM9 exhibit FZIP-INGARCH (1, 1) type behavior. We considered three distinct fixed values, 0.15, 0.25 and 0.80, for each category to represent minimal, moderate, and high zero inflation for the zero-inflation parameter $\omega_t = \omega$. All three categories time series models with fixed minimal zero inflation have lower accuracies and higher constant zero inflation values when compared to other time series models. If the constant zero inflation component happened to take fixed minimum values, it was possible that time varying zero inflation components with low values of A and B could pick it, thus it was misclassified as a TVZIP-INGARCH (*p*, *q*) type process.

Table 14 presents the model selection accuracy for nine different time varying ZIP-INGARCH (*p*, *q*) models. TVM1, TVM2 and TVM3 represent time varying zero inflation INARCH (1) processes; there are three TVZIP-INARCH (2) types of time series models, and they are labeled as TVM4, TVM5 and TVM6, whereas TVM7, TVM8 and TVM9 are the ZIP-INGARCH (1, 1) models. Three pairs of A and B values were used to represent minimal-minimal, minimal-moderate, and minimal-maximum coverage of zero inflation values. When compared to the results presented in the Table 13, there is a 99% minimum accuracy that it can be classified correctly if the data were simulated using TVZIP-INGARCH (*p*, *q*) process. This level of accuracy was an anticipated result because it was difficult to model the time varying zero inflation component when only using a constant zero inflation component.

Table 14: Model selection accuracy for the underline TVZIP-INGARCH (p, q) process with sinusoidal zero inflation function

Model	TVZIP-INGARCH (p, q) True Parameter (Φ)						Accuracy (%)
	A	B	α_0	α_1	α_2	β	
TVM1	0.05	0.05	1.00	0.90			99.2
TVM2	-0.25	-0.25	2.00	0.50			100
TVM3	0.35	0.30	2.00	0.70			100
TVM4	0.05	0.05	1.00	0.20	0.70		99.3
TVM5	-0.25	-0.25	2.00	0.20	0.50		100
TVM6	0.35	0.30	2.00	0.10	0.80		100
TVM7	0.05	0.05	2.00	0.30		0.20	99.4
TVM8	-0.25	-0.25	2.00	0.40		0.20	100
TVM9	0.35	0.30	2.00	0.30		0.10	100

4.4. MODEL SELECTION FOR CASE 2: ZERO INFLATION FUNCTION IS DRIVEN BY EXOGENOUS VARIABLE

In this section, zero inflated time series data was simulated, based on two methods. First, Zhu's constant ZIP-INGARCH (p, q) type process, presented in the above section, was used to generate the time series data and tried to classify whether the simulated data belonged to the appropriate process, which was an underline time series (i.e. in this case FZIP-INGARCH (p, q) process) or proposed TVZIP-INGARCH (p, q) model, where time varying component is modeled by an exogenous process. The results, based on this approach, are presented in the Table 15. In the next section, we considered proposed TVZIP-INGARCH (p, q) process to simulate time series data, and then tried to classify whether the simulated count data belong to underline TVZIP-INGARCH (p, q) process or the Zhu's FZIP-INGARCH (p, q) process. Summarized results are presented in the Table 16. It was assumed that the zero inflation function $\omega_t = g(V_t, \underline{\Gamma})$ was driven by an exogenous variable, as described in equation (2.5), and V_t was AR (12) process with $\eta = 0.25$.

Table 15: Model selection accuracy for the underline FZIP-INGARCH (p, q) process

Model	FZIP-INGARCH (p, q) True Parameter (Φ)					Accuracy (%)
	ω	α_0	α_1	α_2	β	
FM1	0.15	1.00	0.40			60.9
FM2	0.30	2.00	0.50			78.9
FM3	0.80	1.00	0.60			72.3
FM4	0.15	1.00	0.20	0.20		59.6
FM5	0.30	2.00	0.30	0.20		79.1
FM6	0.80	1.00	0.40	0.30		73.0
FM7	0.15	1.00	0.20		0.20	59.7
FM8	0.30	2.00	0.30		0.20	79.1
FM9	0.80	1.00	0.40		0.30	72.6

Both FZIP-INGARCH (*p*, *q*) and TVZIP-INGARCH (*p*, *q*) are fit into the process to investigate which model better explains when the data are generated by using FZIP-INGARCH (*p*, *q*) time series. Table 15 shows the model selection accuracy, and, in this case, accuracy is low with compare to the case 1 results presented in the above Table 13. If the δ_1 parameter in the TVZIP component is close to zero, and the time varying component behaves more like a constant. This may lead to the misclassification of FZIP-INGARCH (*p*, *q*) model, as a TVZIP-INGARCH (*p*, *q*) model and reduce the model selection accuracy. As in case 1, three models are considered based on the minimal, moderate, and higher constant zero inflation component in each of the FZIP-INGARCH (1), FZIP-INGARCH (2) and FZIP-INGARCH (1, 1) cases.

Table 16 presents the model selection accuracy when the time series of count data were originally simulated from a TVZIP-INGARCH (*p*, *q*) type process with ω_t, being modeled by an exogenous time series variable. According to the Table 16, the coefficient of model selection accuracy had lower values for relatively small δ_1. When smaller δ_1 values were presented time varying zero inflation function ω_t, behaved more like fixed zero inflation function ω (i.e. $\lim_{\delta_1 \to 0} \omega_t = \omega$), and hence it was difficult to correctly categorized as a TVZIP-INGARCH (*p*, *q*) process.

Table 16: Model selection accuracy for the underline TVZIP-INGARCH (p, q) process with zero inflation function is driven by exogenous variable

Model	TVZIP-INGARCH (p, q) True Parameter (Φ)						Accuracy (%)
	δ_0	δ_1	α_0	α_1	α_2	β	
TVM1	-2.00	0.50	1.00	0.40			82.9
TVM2	-1.00	-1.00	2.00	0.50			100
TVM3	2.00	1.00	1.00	0.60			100
TVM4	-2.00	0.50	1.00	0.20	0.20		79.2
TVM5	-1.00	-1.00	2.00	0.30	0.20		100
TVM6	2.00	1.00	1.00	0.40	0.30		100
TVM7	-2.00	0.50	1.00	0.20		0.20	82.1
TVM8	-1.00	-1.00	2.00	0.30		0.20	100
TVM9	2.00	1.00	1.00	0.40		0.30	100

5. REAL DATA EXAMPLE

In this section, the proposed TVZIP-INGARCH (p, q) process was applied to a real-world dataset and compared the performance of these models to that of Zhu (2012a). The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were employed to select the best model among the collection of competing models. In the following subsections, we considered two examples: the first example was based on

"Influenza A associated pediatric deaths". The data set was downloaded from the Center for Disease Control and Prevention (https://www.cdc.gov) web page. This data was modeled by using the TVZIP-INGARCH (p, q) model with sinusoidal zero inflation function. In the second example, we used "Pediatric mortality caused by Influenza B" data set, which was downloaded from the same web page (https://www.cdc.gov) to demonstrate the performance of TVZIP-INGARCH (p, q) process where the zero inflation was driven by an exogenous variable.

5.1. REAL DATA EXAMPLE - USE OF A SINUSOIDAL ZERO INFLATION FUNCTION

In this section, "Influenza A associated pediatric mortality" data set was applied to the proposed TVZIP-INARCH (1), TVZIP-INARCH (2) and TVZIP-INGARCH (1, 1) models with sinusoidal zero inflation and compared to the results of those from fixed zero inflation, as discussed by Zhu (2012a). We chose the weekly U.S. data of pediatric death caused by virus type A over the time period of week 40from year 2015 to week 43 from year 2018. The data was taken from weekly U.S. Influenza Surveillance report, which was published by Center for Disease Control and Prevention (CDC) (https://gis.cdc.gov/GRASP/Fluview/PedFluDeath.html). Data sets contain 160 weekly observations of pediatric death counts. Summary statistics of the data showed the mean of 1.506 and a variance of6.352, suggestive of over dispersion. Figure 1 illustrates the frequency of each pediatric mortality case caused by virus type "A" using a bar chart. Observe that there are 86 zeros, which comprises 53.8% of the total time points.

Figure 2 illustrates the original time series of the data set followed by its ACF plot and PACF plot, respectively.The bar chart shows the excess number of zeros in the data, and the time series plot demonstrates prolonged periods of zero counts, which supports the use of zero inflated poisson time series models to analyze this data. Furthermore, we can observe an annual seasonality in the peaks and the low periodsof the time series.

Since we chose the zero inflated poisson model to fit the data, we need to find the best model that describs the zero inflation behaviour. To understand the zero inflation behaviour of this data set, we aggregate weekly data in to its corresponding calendar month and constructed the total monthly zero mortality counts.It was converted into a monthly

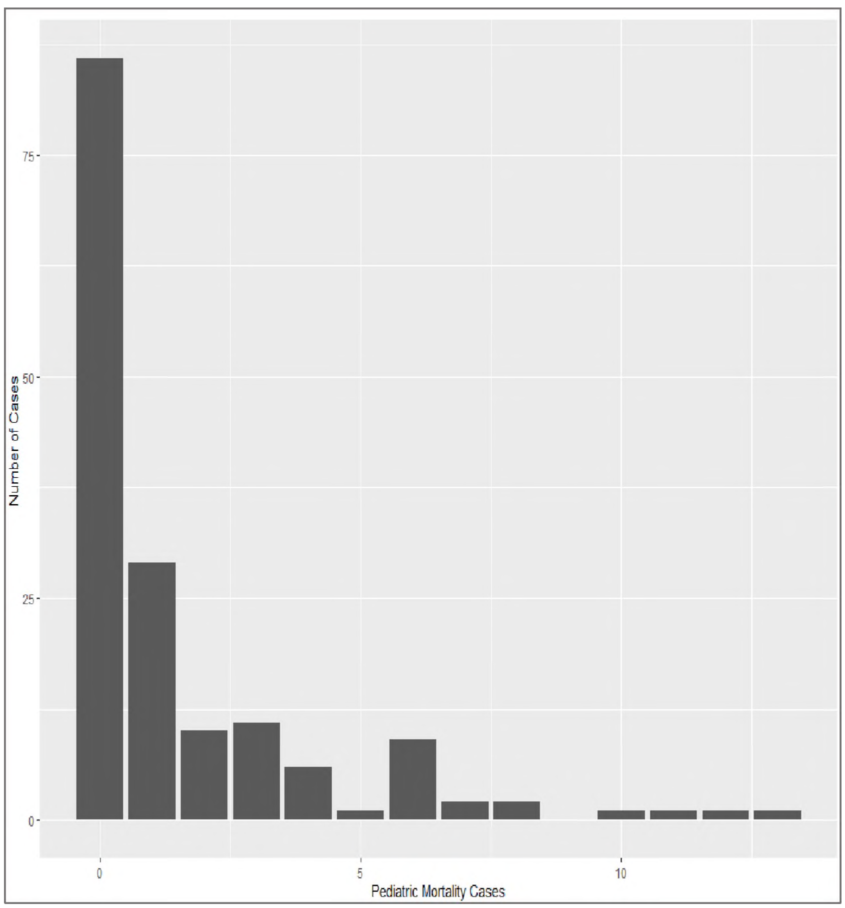

Figure 1: Bar chart of the pediatric deaths caused by virus type “A”

proportion by dividing each monthly count by its maximum. The plot of the monthly proportion of zero counts exhibited a sinosidual behaviour throughout the observed time span. Thus, the general sinosidual function mentioned in Equation (2.4) was used to model the zero inflation behaviour of this data.

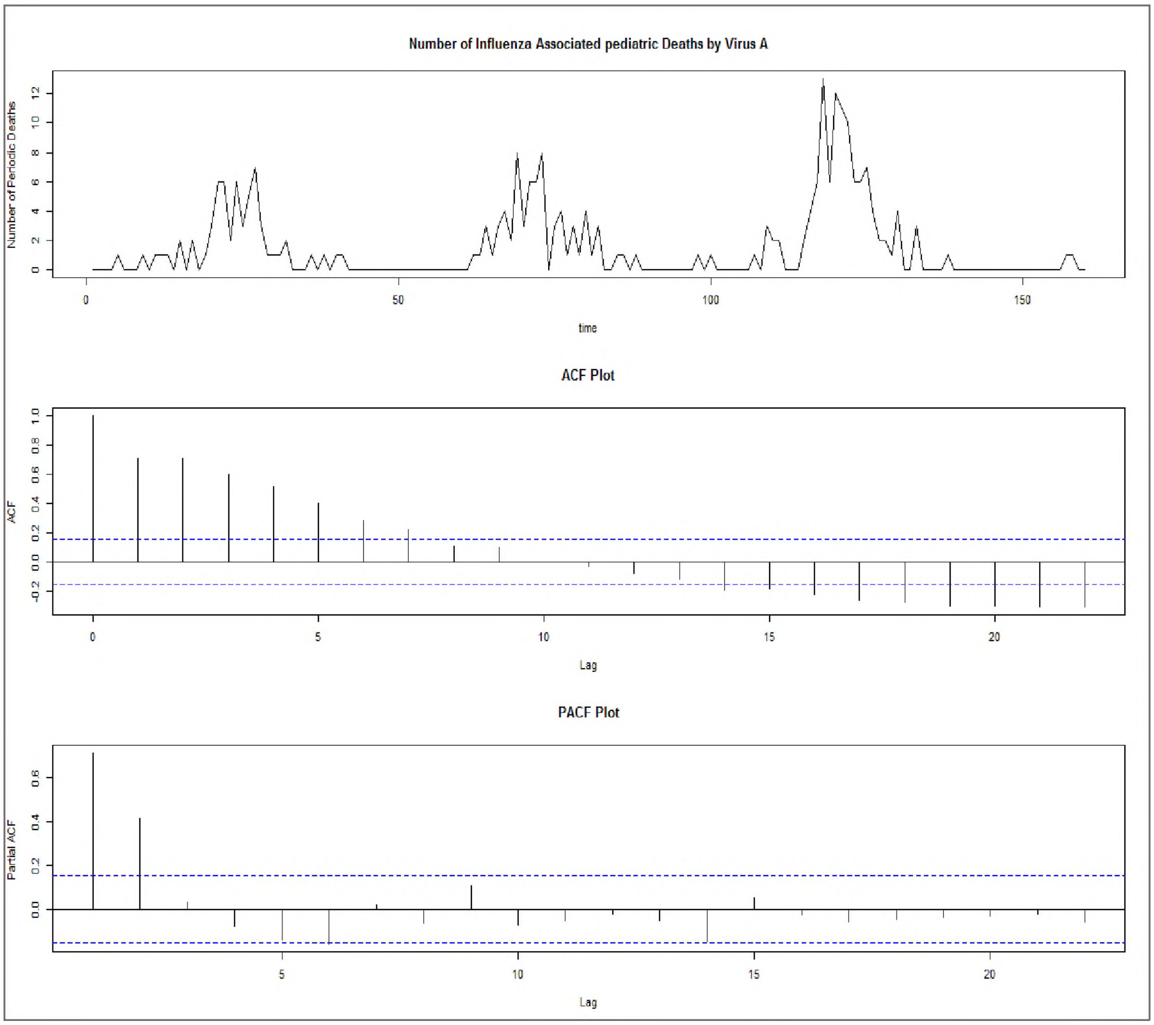

Figure 2: Pediatric mortality time series plot, sample auto covariance plot and sample partial auto covariance plot

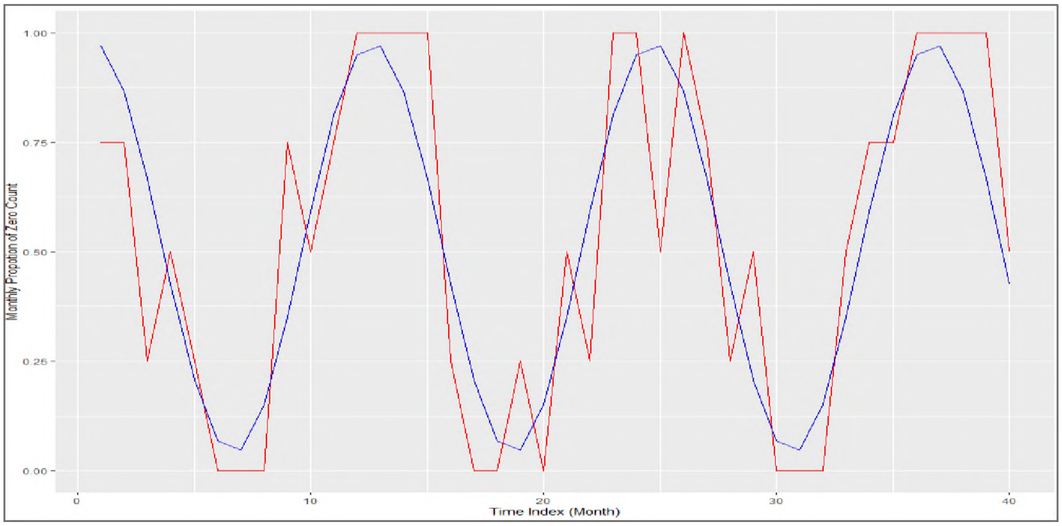

Figure 3: Monthly proportion of zero mortality counts (red) versues the fitted sinusoidal zero inflated function (blue)

Three time series models were fitted namely; TVZIP-INARCH(1) (***M1***),TVZIP-INARCH (2) (***M2***) and TVZIP-INGARCH (1,1) (***M3***), for three different senarios (***S1***, ***S2,*** and ***S3***) . In senario ***S1,*** it was assumed that there was a constant zero inflated value throughout the time period as assumed by Zhu (2012a). In the model listed as ***S2,*** we assumed a constant zero inflation for short term periods (weeks) of a given long term period (month),while the zero inflation component of the long term period (monthly) was allowed to vary, according to a sinusoidal zero inflation function.The models under ***S3*** used a sinusoidal function to exemplify the time varying zero inflated function at each unit time. The EM algorithems were used to estimate the model parameters and the results were recorded in the Table 17. We used both AIC and BIC model selection criterion to identify the best fit for the data.

Table 17: Estimated parameters, AIC and BIC for the pediatric death counts cause by virus A

Model	ω	A	B	α_0	α_1	α_2	β_1	AIC	BIC
S1M1	0.1806			0.4714	0.8416			451.6289	460.8544
S1M2	0.1012			0.2539	0.4967	0.4187		419.8733	432.1740
S1M3	0.0909			0.0807	0.4804		0.5057	420.6112	432.9119
S2M1		0.1023	0.4467	1.1030	0.7165			411.1064	423.4071
S2M2		0.0775	0.4266	0.7446	0.4715	0.3595		**397.4778**	**412.8537**
S2M3		0.0733	0.4219	0.3730	0.4832		0.4413	401.4099	416.7857
S3M1		-0.3116	0.3214	1.0292	0.7193			420.2297	432.5304
S3M2		-0.3074	0.2613	0.6460	0.4544	0.3927		402.9449	418.3208
S3M3		-0.3009	0.2420	0.2726	0.4698		0.4773	407.6126	422.9885

According to results in Table 17, in general, models with cyclically varying zero inflation function had lower AIC and BIC values, when compared to the results for models Zhu (2012a) proposes. Put simply, models introduced under the ***S2*** and ***S3*** fitted the data better than the models discussed in ***S1***. Among all the TVZIP-INARCH (1), TVZIP-INARCH (2) and TVZIP-INGARCH (1, 1) models TVZIP-INARCH (2) process had lower AIC and BIC values, when compared to the results for models Zhu (2012a) proposes. Put simply, models introduced under the ***S2*** and ***S3*** fitted the data better than the models discussed in ***S1***. Among all the TVZIP-INARCH (1), TVZIP-INARCH (2) and TVZIP-INGARCH (1, 1) models TVZIP-INARCH (2) process had lower AIC and BIC values in all the three scenarios. Finally, based on these information criteria, the TVZIP-INARCH (2) model with zero inflation model ***S2,*** provided the best fit to the data. In this model, we assumed that week's within any given month has a constant zero inflation, yet monthly zero inflation varies cyclically.

5.2. REAL DATA EXAMPLE - ZERO INFLATION FUNCTION IS DRIVEN BY EXOGENOUS VARIABLE

In this section, we examined the performance of the TVZIP-INGARCH (p, q) models where the zero inflated function was driven by an exogenous variable. "Influenza B associated pediatric mortality" data set was used, and the data is applied to the TVZIP-INARCH (1), TVZIP-INARCH (2) and TVZIP-INGARCH (1, 1) models. We selected the weekly average of nationwide low temperatures as the exogenous variables that drive the zero inflation probability. Influenza B associated pediatric mortality data was taken from the weekly U.S. Influenza Surveillance Report (https://www.cdc.gov/flu/weekly/index.htm), and the temperature data was taken from weather prediction center (https://www.wpc.ncep.noaa.gov). Both data sets spanned over the time period of week 40 for year 2014 to week 39 for year 2018. Both data sets contained 209 weekly observations of pediatric death counts and number of tests done. Summary statistics of infant mortality cases due to influenza B showed the mean as 0.8517 and variance as 2.4538. Since the empirical variance was higher than the empirical mean, the data exhibited an over dispersion. Figure 4 illustrates the frequency of each pediatric mortality case caused by virus type "B" using a bar chart. According to the bar chart, there are 128 zeros, which comprises 61.2% of total of the time points. This suggested that the pediatric mortality data were zero inflated.

The time series plot, ACF and PACF plots are given in Figure 5. Based on time series plot, we can see that there is an annual seasonality exhibited in this data set. Moreover, it shows that there were periods with clusters of zeros. Therefore, as discussed in Section 5.1, we suggested TVZIP-INGARCH (p, q) model the count data series.

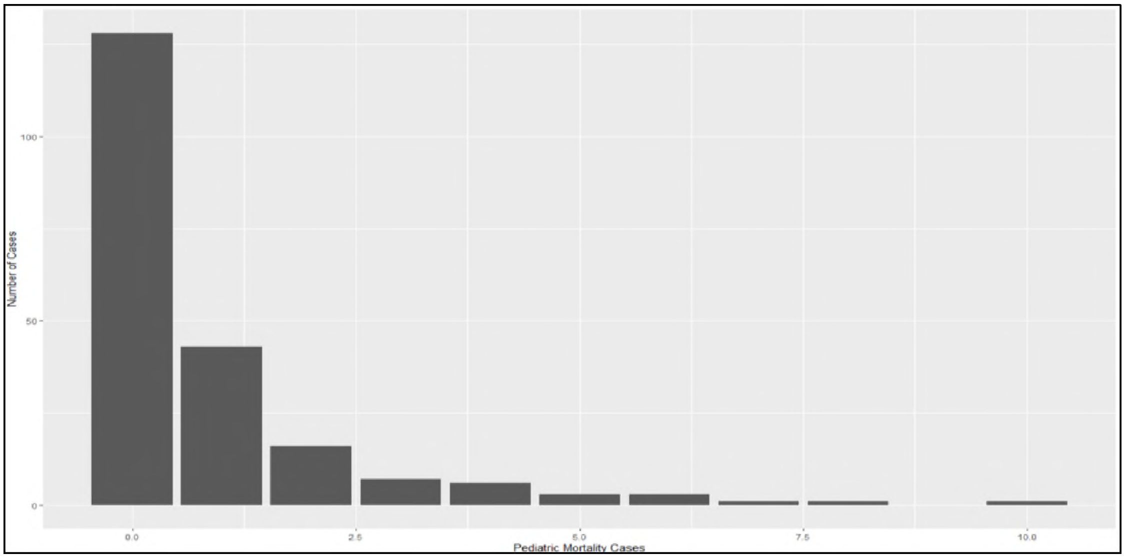

Figure 4: Bar chart of the pediatric deaths caused by virus type “A”

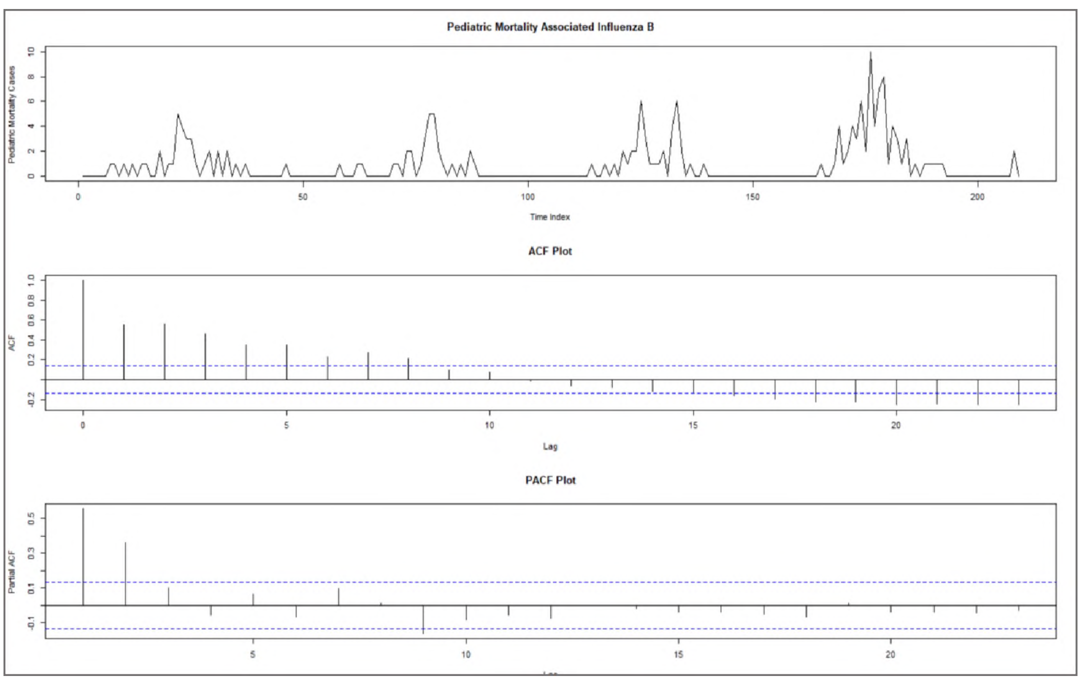

Figure 5: Pediatric mortality time series plot, sample auto covariance plot and sample partial auto covariance plot

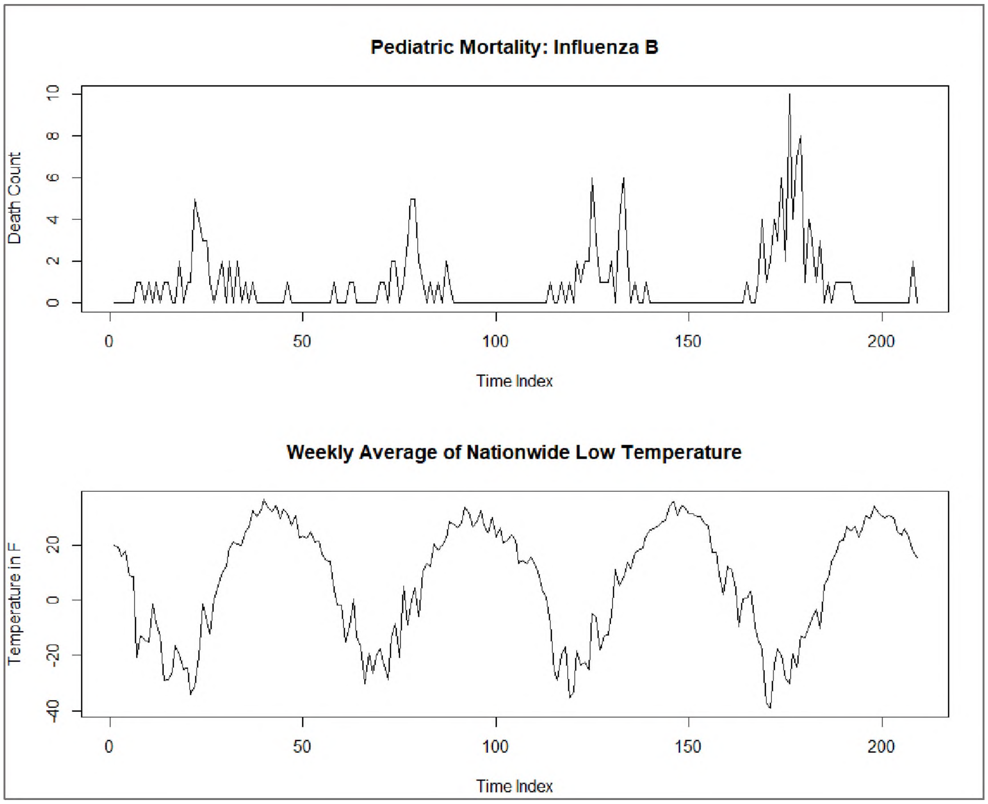

Figure 6: Time series plots of pediatric mortality cases (upper panel) and weekly average of nationwide low temperature

In this example, the time varying zero inflation was modeled by considering an exogenous time series. We assumed that the excess zeros were driven by another independent time series. We used the weekly average of nationwide low temperature. Comparison between two time series plots is given in Figure 6.

Figure 6 shows that periods with higher values of low temperature coincide with periods of zero pediatric mortality caused by Influenza B. Hence, it was established that periods with high zero counts (low pediatric mortality) are notably related to periods with higher values of low temperature. Thus, we used averaged low temperature data to model the time varying zero inflated component in the pediatric mortality data set. In this example, time varying zero inflation process was modeled using the formulation from Equation (2.5).

We fitted three different time series models TVZIP-INARCH(1) (***M1***), TVZIP-INARCH (2) (***M2***) and TVZIP-INGARCH (1,1) (***M3***) for two different senarios (***S1*** and ***S2***) . In senario ***S1,*** we considered that there was a constant zero inflated value throughout the time period, as assumed by Zhu (2012a). For the models listed under ***S2,*** we assumed there was a time varying zero inflation,and we modeled it by using a logistic regression model with low temperature as the independent variable. The EM algorithm was used to estimate the model parameters,and the results arerecorded in Table 18. We used AIC and BIC model selection criterion to identify the best fit for the data.

Table 18: Estimated parameters, AIC and BIC for the pediatric death counts cause by virus B

Model	ω	δ_0	δ_1	α_0	α_1	α_2	β_1	AIC	BIC
S1M1	0.2183			0.4517	0.6746			488.2955	498.3225
S1M2	0.0575			0.1737	0.4095	0.4454		442.2950	455.6644
S1M3	0.1478			0.0001	0.3766		0.7631	441.4626	454.8319
S2M1		-1.7436	0.1115	0.7258	0.5789			453.1119	466.4812
S2M2		-2.5485	0.1190	0.3478	0.4027	0.4101		**427.7886**	**444.5003**
S2M3		-1.6346	0.1048	0.2385	0.3591		0.8305	428.6728	445.3844

Table 18 shows the models that fall under the ***S2*** exhibited low AIC and BIC values compared to the models under ***S1***. Our novel approach of modeling time varying zero inflation improved on the AIC and BIC values when compared to the constant zero inflation method. Based on the AIC and BIC values, TVZIP-INARCH (2) with time varying zero inflation function provided a better fit to the pediatric mortality data when compared to a model that assumes constant zero inflation probability.

6. CONCLUSIONS

We proposed a time varying zero inflated Poisson integer GARCH Model (TVZIP-INGARCH) with two distinct formulations to model the time varying zero inflation component. Based on the Monte-Carlo simulation study results, the Expected Maximization (EM) and Maximum Likelihood Estimation (MLE) methods produce similar results with respect to parameter estimates. It is seen that both EM and MLE techniques estimate the parameters of the predefined time varying zero inflated function with good accuracy. In cases where the zero inflation function is cyclically varying with minimal to large or minimal to moderate zero inflations, the zero inflation parameters were estimated with lower Mean Absolute Deviation Error (MADE) when compared to the scenarios of minimal zero inflation. Further, in cases where zero inflation was low and it is influenced by an exogenous process, MLE method produces far better estimates for the parameters, especially for those associated with zero inflation process. In contrast, the EM method produces slightly better estimates for the zero inflated parameters based on the MADE values, than the MLE method when the zero inflation is high and driven by the exogenous process. Furthermore, the level of accuracy in correctly identifying the data generating process was found to be high for both types of zero inflation cases, given that the data is simulated via a TVZIP-INGARCH process. When tested on two real-life data sets, the TVZIP-INGARCH models performed better than those proposed by Zhu (2012a), illustrating the utility of the proposed models. In addition, the flexibility of the zero inflation component of the formulation that allowed modeling through deterministic cyclical functions, or through exogenous time series, provided the proposed model for added versatility.

REFERENCES

Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive INAR (1) process. *Journal of Time Series Analysis* 8: 261-275.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, 31:307-327.

Davis, R. A., Dunsmuir W.T.M., and Streett, S.B. (2003). Observation-Driven Models for Poisson Counts. *Biometrika* 90:777–790.

Engle, R. and Jeffrey, R. (1998). Autoregressive Conditional Duration: A New Model for Irregular Spaced Transaction Data. *Econometrica*, 66:1127-1162.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, 27:923– 942.

Ghahramani, M. and Thavaneswaran, A. (2009). On some properties of autoregressive conditional Poisson (ACP) models. *Economics Letters*, 105:273–275.

Heinen, A. (2003). Modeling Time Series Count Data: An Autoregressive Conditional Poisson Model. Technical Report MPRA paper 8113, University Library of Munich, Munich, Germany. Available at https://mpra.ub.uni-muenchen.de/8113/.

Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*. John Wiley & Sons. Hoboken, New Jersey.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, 21(4):645-650.

Quddus, M. (2008). Time series count data models: an empirical application to traffic accidents. *Accident Analysis and Prevention*, 40 (5):1732-1741.

Rydberg, T. H. and Shephard, N. (1999). BIN models for trade-by-trade data. Modelling the number of trades in a fixed interval of time. Working paper, Nuffield College, Oxford.

Weiß, C. H. (2008). Thinning operations for modeling time series of counts-A survey. *ASTA-Advances in Statistical Analysis*, 92(3):319– 341.

Yang, M. (2012). Statistical models for count time series with excess zeros. (Doctoral dissertation). University of Iowa.

Ye, F., Garcia, T.P., Pourahmadi, M., and Lord, D. (2012). Extension of negative binomial GARCH model analyzing effects of gasoline price and miles traveled on fatal crashes involving intoxicated drivers in Texas. *Transportation Research Record: Journal of the Transportation Research Board*, 2279:31–39.

Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, 32:54-67.

Zhu, F. (2012a). Zero-inflated Poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference*, 142:826–839.

Zhu, F. (2012b). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications*, 389 (1):58-71.

Zhu, F. (2012c). Modeling time series of counts with COM-Poisson INGARCH models. *Mathematical and Computer Modelling*, 56:191-203.

Zhu, F. and Wang, D. (2011). Estimation and testing for a Poisson autoregressive model. *Metrika*, 73:211–230.

APPENDIX A.

DERIVATIONS OF CONDITIONAL MEAN AND CONDITIONAL VARIANCE

Let $\{X_t\}_{t=1}^{N}$ be a discrete time series of count data and conditional distribution of $X_t|\mathrm{F}_{t-1} \sim ZIP(\lambda_t, \omega_t)$ as described in (2.1). As explained ZIP process has two types of zeros namely, zeroes coming from the Poisson distribution with the rate parameter λ_t and the zeros that are generated. Let us define $\{Z_t\}$ to be a Bernoulli random variable with two values such that $Z_t = 1$ if X_t is a generated zero and $Z_t = 0$ if it is coming from ordinary Poisson distribution. Therefore $Z_t|\mathrm{F}_{t-1} \sim Bernoulli(\omega_t)$ such that,

$$P(Z_t = z \mid \mathrm{F}_{t-1})\begin{cases} \omega_t : z = 1. \\ 1-\omega_t : z = 0. \end{cases}$$

Let define the conditional probability mass function of $X_t|Z_t, \mathrm{F}_{t-1} \sim Poisson((1-Z_t)\lambda_t)$.

Hence the conditional distribution of $X_t|\mathrm{F}_{t-1} \sim ZIP(\lambda_t, \omega_t)$ can be expressed as:

$$P(X_t = k|\mathrm{F}_{t-1}) = \sum_{z=0}^{1} P(X_t = k \mid Z_t, \mathrm{F}_{t-1}) P(Z_t = z \mid \mathrm{F}_{t-1}),$$

The Conditional expectation of $X_t|\mathrm{F}_{t-1}$ is:

$$\begin{aligned} E(X_t|\mathrm{F}_{t-1}) &= E\left[E(X_t|Z_t, \mathrm{F}_{t-1})\right], \\ &= E\left[(1-Z_t)\lambda_t|\mathrm{F}_{t-1}\right], \\ &= \lambda_t E\left[(1-Z_t)|\mathrm{F}_{t-1}\right], \\ &= \lambda_t(1-\omega_t). \end{aligned}$$

The Conditional variance of $X_t|\mathrm{F}_{t-1}$ is:

$$
\begin{aligned}
Var\left(X_t \middle| \mathrm{F}_{t-1}\right) &= Var\left[E\left(X_t \middle| Z_t, \mathrm{F}_{t-1}\right)\right] + E\left[Var\left(X_t \middle| Z_t, \mathrm{F}_{t-1}\right)\right], \\
Var\left[E\left(X_t \middle| Z_t, \mathrm{F}_{t-1}\right)\right] &= Var\left[\left(1-Z_t\right)\lambda_t \middle| \mathrm{F}_{t-1}\right] = \lambda_t^2\left(1-\omega_t\right)\omega_t, \\
E\left[Var\left(X_t \middle| Z_t, \mathrm{F}_{t-1}\right)\right] &= E\left[\left(1-Z_t\right)\lambda_t \middle| \mathrm{F}_{t-1}\right] = \lambda_t\left(1-\omega_t\right), \\
Var\left(X_t \middle| \mathrm{F}_{t-1}\right) &= \lambda_t^2\left(1-\omega_t\right)\omega_t + \lambda_t\left(1-\omega_t\right) = \lambda_t\left(1-\omega_t\right)\left(1+\lambda_t\omega_t\right).
\end{aligned}
$$

APPENDIX B.

THE CONDITIONAL LOG LIKELIHOOD FUNCTION

The conditional probability mass function of $Z_t \middle| \mathrm{F}_{t-1} \sim Bernoulli(\omega_t)$ is, $P(Z_t = z_t) = \omega_t^{z_t}(1-\omega_t)^{1-z_t}$.

The conditional probability mass function of $X_t \middle| Z_t, \mathrm{F}_{t-1} \sim Poisson\left((1-Z_t)\lambda_t\right)$.

$$P\left(X_t = x_t \middle| Z_t = z_t, \mathrm{F}_{t-1}\right) = z_t + (1-z_t)\frac{\lambda_t e^{-\lambda_t}}{x_t!} = \left(\frac{\lambda_t e^{-\lambda_t}}{x_t!}\right)^{(1-z_t)}.$$

Therefore, conditional log likelihood function of $P\left(Z_t \middle| \mathrm{F}_{t-1}\right)P\left(X_t \middle| Z_t, \mathrm{F}_{t-1}\right)$ is $L(\Phi)$,

$$\prod_{\forall t} P\left(Z_t \middle| \mathrm{F}_{t-1}\right)P\left(X_t \middle| Z_t, \mathrm{F}_{t-1}\right) = \prod_{\forall t} \omega_t^{z_t}(1-\omega_t)^{(1-z_t)}\left(\frac{\lambda_t e^{-\lambda_t}}{x_t!}\right)^{(1-z_t)} = \prod_{\forall t} \omega_t^{z_t}\left((1-\omega_t)\frac{\lambda_t e^{-\lambda_t}}{x_t!}\right)^{(1-z_t)}.$$

APPENDIX C.

PARAMETER CONSTRAINTS OF THE SINUSOIDAL FUNCTION

Let define sinusoidal zero inflation function $\omega_t = g(\mathbf{V}_t, \underline{\Gamma}) \in (0,1)$ as given bellow.

$$\omega_t = g(\mathbf{V}_t, \underline{\Gamma}) = A\sin\left(\frac{2\pi}{S}t\right) + B\cos\left(\frac{2\pi}{S}t\right) + C,$$

Case 1: If $A = 0, B = 0$ and $C \in (0,1)$ then,

$\omega_t = g(\mathbf{V}_t, \underline{\Gamma}) = C$, is a constant over the time span,

Case 2: If $A \neq 0, B \neq 0$ and $C \in (0,1)$ then,

$$\begin{aligned}
\omega_t &= A\sin\left(\frac{2\pi}{S}t\right) + B\cos\left(\frac{2\pi}{S}t\right) + C,\\
&= \sqrt{A^2+B^2}\left(\frac{A}{\sqrt{A^2+B^2}}\sin\left(\frac{2\pi}{S}t\right) + \frac{B}{\sqrt{A^2+B^2}}\cos\left(\frac{2\pi}{S}t\right) + \frac{C}{\sqrt{A^2+B^2}}\right),\\
&= \sqrt{A^2+B^2}\left(\cos(\varphi)\sin\left(\frac{2\pi}{S}t\right) + \sin(\varphi)\cos\left(\frac{2\pi}{S}t\right) + \frac{C}{\sqrt{A^2+B^2}}\right),\\
&= \sqrt{A^2+B^2}\left(\sin\left(\frac{2\pi}{S}t + \varphi\right) + \frac{C}{\sqrt{A^2+B^2}}\right),\\
&= \sqrt{A^2+B^2}\left(h(t,S,A,B,) + \frac{C}{\sqrt{A^2+B^2}}\right).
\end{aligned}$$

Here $\cos(\varphi) = \dfrac{A}{\sqrt{A^2+B^2}}$ and $h(t,S,A,B,) \in [-1,1]$. For a fixed A, B and C

$$\begin{aligned}
&\max(\omega_t) = \sqrt{A^2+B^2} + C < 1,\\
&\min(\omega_t) = -\sqrt{A^2+B^2} + C > 0,\\
&\sqrt{A^2+B^2} < C < 1 - \sqrt{A^2+B^2}.
\end{aligned}$$

If $C = \sqrt{A^2+B^2}$,

$$\omega_t = \sqrt{A^2+B^2}\left(h(t,S,A,B,) + 1\right), \omega_t \in [0,1) \Rightarrow \sqrt{A^2+B^2} < \frac{1}{2}, |A| < \frac{1}{2}, |A| < \frac{1}{2}.$$

If $C > \sqrt{A^2 + B^2} = \sqrt{A^2 + B^2} + \delta$, any $\delta > 0$, choose A, B such that:

$$\sqrt{A^2 + B^2} < \frac{1}{2}, |A| < \frac{1}{2}, |A| < \frac{1}{2}$$

$$\Rightarrow \omega_t \in (0,1).$$

APPENDIX D.

INITIAL VALUE ESTIMATION

Let consider the case ZIP-INGARCH (1,1) where p=1 and q=1.

Let u_t be the martingale difference, which is defined as,

$u_t = X_t - E(X_t \mid \mathrm{F}_{t-1}) = X_t - \lambda_t(1-\omega_t)$ where the λ_t is expressed as INGARCH (1, 1) process and therefore we can write,

$$\begin{aligned}
&\lambda_t = \alpha_0 + \alpha_1 X_{t-1} + \beta_1 \lambda_{t-1},\\
&u_t = X_t - (\alpha_0 + \alpha_1 X_{t-1} + \beta_1 \lambda_{t-1})(1-\omega_t),\\
&X_t - (1-\omega_t)\alpha_1 X_{t-1} - \beta_1(1-\omega_t)\lambda_{t-1} = (1-\omega_t)\alpha_0 + u_t,\\
&X_t - (1-\omega_t)\alpha_1 X_{t-1} - \beta_1 X_{t-1} = (1-\omega_t)\alpha_0 + u_t - \beta_1 u_{t-1},\\
&X_t \sim ARMA(1,1) \Rightarrow X_t - \phi_t X_{t-1} = \gamma_t + u_t + \theta u_{t-1}.
\end{aligned}$$

Case 1: If $\omega_t = \omega \ \forall t$, then,

$\hat{\alpha}_1 = \dfrac{\hat{\phi}+\hat{\theta}}{(1-\omega)}, \hat{\beta}_1 = -\hat{\theta}$ and $\hat{\alpha}_0 = \dfrac{(1-\hat{\phi})}{(1-\omega)}\hat{\mu}$, where $\hat{\mu}$ is the mean of the ARMA (1, 1) model.

Case 2: If $\{X_t\}_{t=1}^{N}$ be a discrete time series and it has m subseries with each subseries have M counts. Within each subseries let assume $\omega_t = \omega_{.,l}$ where $l = 1,2,\dots,m.$ Then for a given subseries l, $\hat{\alpha}_{1,l} = \dfrac{\hat{\phi}_l + \hat{\theta}_l}{(1-\omega_{.,l})}, \hat{\beta}_{1,l} = -\hat{\theta}_l$, $\hat{\alpha}_{0,l} = \dfrac{(1-\hat{\phi}_l)}{(1-\omega_{.,l})}\hat{\mu}_l$. Here $\hat{\mu}_l$ is the mean of the subseries l. Therefore, $\alpha_i = \sum_{l=1}^{m} \dfrac{\alpha_{i,l}}{m}, i = 0,1$ and $\beta_l = \sum_{l=1}^{m} \dfrac{\beta_{1,l}}{m}$.

Simillarly we can derive equations to estimate parameters of the ZIP-INARCH(1) and ZIP-INARCH (2) models by fitting ARMA(1,0) and ARMA(2,0) process respectively.

SECTION

2. CONCLUSIONS

My doctoral dissertation extended the CARR and the Feedback Asymmetric CARR (FACARR) models using three papers. In addition, another paper presented a generalized zero-inflated Poisson model whose formulation is similar to that of the GARCH model. First three papers mainly focused on modeling and forecasting financial volatilities using conditional heteroscedastic time series models defined in the ranged based setting. The final paper proposed a conditional heteroscedastic time series process which is derived from the Autoregressive Conditional Poisson (ACP) to model the time series of count data.

Paper I proposed the Composite CARR (CCARR), which is a composite range, based component model used to analyze the long-term and short-term volatility components in the daily price range data. The long-term volatility component is modeled using a stochastic volatility component, which itself exhibits the conditional volatility. Both the long-term and short-term components are driven by past values of price range data. The MLE technique was used to estimate the model parameters and the CCARR model performance was gauged by using the S&P 500 and the FTSE 100 stock indices.

Paper II generalized the Feedback Asymmetric CARR (FACARR) model of Xi (2018) and introduced the Generalized FACARR (GFACARR) model. The Weak stationarity conditions for the GFACARR model were derived. Furthermore, two version of the GFACARR models were discussed namely, Bivariate Exponential GFACARR type a (BEGFACARR-a) model which is useful when the upward and downward disturbance term behave independently and the Bivariate Exponential GFACARR type b (BEGFACARR-b) process which is suitable when there is a correlation between upward and downward terms. The parameters of the proposed BEGFACARR-a and BEGFACARR-b were estimated using the MLE procedure and a simulation study was conducted to evaluate the finite sample performance. The performance of the BEGFACARR-b model was compared to FACARR using the S&P500, FTSE 100 and CAC 40 stock indices and found that the BEGFACARR-b model fitted better than the FACARR model to both in-sample and out-of-sample price range data.

In paper III, a new class of non-linear asymmetric range-based conditional heteroscedastic model was proposed, and this model is called as Threshold Asymmetric CARR (TACARR) model. The disturbance term of the TACARR process is assumed to follow a threshold distribution with positive support. The study assumed that the conditional expected range process switches between two market regimes namely upward market and downward market. Since the proposed model using past upward and downward price range data to adjust the threshold and having a nonlinear conditional heteroscedastic structure, the TACARR model is a good alternative to the CARR, Asymmetric CARR (ACARR), Feedback ACARR (FACARR), and Threshold Autoregressive Range (TARR) models. The performance of the TACARR model was assessed using IBM index data. Empirical results showed that the proposed TACARR model was useful in-sample prediction and out-of-sample forecasting volatility.

The final paper proposed a time varying zero-inflated Poisson process to model the time series of count data with serial dependence. The proposed Time Varying Zero-Inflated Poisson Integer GARCH (TVZIP-INGARCH) model was a generalization of the

Zero-Inflated Poisson Integer GARCH (ZIP-INGARCH) model of Fukang Zhu in 2012, which, itself was a generalization of the Integer GARCH (INGARCH) model proposed by Ferland, Latour, and Oraichi in 2006. In this relaxed the Zhu's fixed zero inflation parameter and allowed it to vary over time according to a deterministic function or be driven by an exogenous variable. Two applications based on the real world data were discussed and the proposed time varying ZIP-INGARCH (TVZIP-INGARCH) model fitted better with the data compared to Zhu's constant ZIP-INGARCH model.

BIBLIOGRAPHY

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. Journal of Finance, 57(3), 1047–1092.

Baillie, R., Bollerslev, T., and Mikkelsen, H.O., (1996). Fractionally integrated generalized autoregressive conditional heteroscedasticity, Journal of Econometrics, 74, (1), 3-30.

Beckers, S., (1983). Variance of Security Price Return Based on High, Low and Closing Prices. Journal of Business 56, 97–112.

Bollerslev, T., (1986). Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. Journal of Econometrics, 73, 151-184.

Chiang, M., Chou, R. Y., & Wang, L (2016). Outlier detection in the lognormal logarithmic conditional autoregressive range model. Oxford Bulletin of Economics and Statistics, 78(1), 126-144.

Chen C.W.S, Gerlach, R. and Lin, E.M.H. (2008) "Volatility Forecasting Using Threshold Heteroskedastic Models of the Intra-day Range" Computational Statistics and Data Analysis, Special Issue on Statistical and Computational Methods in Finance,52, 2990-3010.

Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model. Journal of Money, Credit and Banking, Vol. 37, pp. 561-582.

Chou, R. (2006). Modeling the asymmetry of stock movements using price ranges. Advances in Econometrics, 20, 231–257.

Engle, R.F., (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation. Econometrica 50, 987–1008.

Engle, R.F., (2003). Risk and Volatility: Econometric Models and Financial Practice, Nobel Lecture, http://nobelprize.org/economics/laureates/2003/engle-lecture.pdf.

Engle, R. F., and Bollerslev, T. (1986). Modelling the persistence of conditional variances. Econometric Reviews, 5, 1-500.

Engle, R. & Lee, G. (1999). A long-run and short-run component model of stock return volatility. In Engle, R. & White, H. (eds.), Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger, pp. 475–497. Oxford University Press.

Engle, R.F., & Russell, J.R., (1998). Autoregressive Conditional Duration: A New Model for Irregular Spaced Transaction Data. Econometrica 66, 1127–1162.

Engle, R. & Rangel, J. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. Review of Financial Studies 21, 1187–1222.

Engle, R. F., Ghysels, E., & Sohn, B (2013). Stock market volatility and macroeconomic fundamentals, Review of Economics and Statistics, 95(3), 776–797.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. Journal of Time Series Analysis, 27:923– 942.

Ghysels, E., Sinko, A., & Valkanov, R. (2006) MIDAS regressions: Further results and new directions. Econometric Reviews 26, 53–90.

Glosten, L., Jagannathan, R. and Runkle, D. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance 48: 1779-1801.

Heinen, A. (2001) Modeling Time Series Count Data: The Autoregressive Conditional Poisson Model. PhD thesis, San Diego, CA: Department of Economics, University of California

Heinen, A. (2003). Modeling Time Series Count Data: An Autoregressive Conditional Poisson Model. Technical Report MPRA paper 8113, University Library of Munich, Munich, Germany. Available at https://mpra.ub.uni-muenchen.de/8113/.

Kunitomo, N. (1992). Improving the Parkinson method of estimating security price volatilities. Journal of Business, 65(2), 295–302.

Ling, S., & McAleer, M. (2002). Necessary and Sufficient Moment Conditions for the GARCH (r, s) and Asymmetric Power GARCH (r, s) Models. Econometric Theory, 18(3), 722-729.

Nelson, D. (1991). Conditional heteroscedasticity in Asset Returns: A New Approach. Econometrica, 59(2), 347-370.

Nelson, D.B. and Cao, C.Q. (1992) 'Inequality Constraints in the Univariate GARCH Model', Journal of Business and Economic Statistics, Vol. 10, No. 3, pp. 229-235.

Parkinson, Michael (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. Journal of Business 53, 61–65.

Sentana, E. (1995). Quadratic ARCH Models. The Review of Economic Studies, 62(4), 639–661.

Xie H. (2018). Financial volatility modeling: The feedback asymmetric conditional autoregressive range model. Journal of Forecasting. 2019; 38:11–28.

Xie, H. B., & Wu, X. Y. (2017). A conditional autoregressive range model with gamma distribution for financial volatility modeling. Economic Modelling, 64, 349–356.

Zakoian, J. M. (1994). Threshold heteroskedastic models. Journal of Economic Dynamics and Control, 18(5), 931-955.

Zhu, F. (2012a). Zero-inflated Poisson and negative binomial integer-valued GARCH models. Journal of Statistical Planning and Inference, 142:826–839.

Zhu, F. (2012b). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications*, 389 (1):58-71.

Zhu, F. (2012c). Modeling time series of counts with COM-Poisson INGARCH models. *Mathematical and Computer Modelling*, 56:191-203.

Zhu, F. and Wang, D. (2011). Estimation and testing for a Poisson autoregressive model. *Metrika*, 73:211–230.

VITA

Ratnayake Mudiyanselage Isuru Panduka Ratnayake attended Kingswood College, Kandy, Sri Lanka from 1994 to 2007. He entered the University of Peradeniya, Sri Lanka in 2008 and completed his bachelor's degree specializing in Statistics in 2013. Soon after graduating he started teaching at the University of Peradeniya. A year later he joined to the University of Sri Jayewardenepura as a lecturer. Isuru entered the Missouri University of Science and Technology for graduate studies in fall 2015, and successfully defended his thesis in December 2020, and he received his PhD degree in Mathematics with Statistics emphasis from Missouri S&T in May 2021.