Outlier Detection Based on Discrete Wavelet Transform with Application to Saudi Stock Market Closed Price Series

Khudhayr A. RASHEDI¹, Mohd T. ISMAIL², S. Al WADI³, Abdeslam SERROUKH⁴

Received: September 01, 2020 Revised: October 26, 2020 Accepted: November 05, 2020

Abstract

This study investigates the problem of outlier detection based on discrete wavelet transform in the context of time series data where the identification and treatment of outliers constitute an important component. An outlier is defined as a data point that deviates so much from the rest of observations within a data sample. In this work we focus on the application of the traditional method suggested by Tukey (1977) for detecting outliers in the closed price series of the Saudi Arabia stock market (Tadawul) between Oct. 2011 and Dec. 2019. The method is applied to the details obtained from the MODWT (Maximal-Overlap Discrete Wavelet Transform) of the original series. The result show that the suggested methodology was successful in detecting all of the outliers in the series. The findings of this study suggest that we can model and forecast the volatility of returns from the reconstructed series without outliers using GARCH models. The estimated GARCH volatility model was compared to other asymmetric GARCH models using standard forecast error metrics. It is found that the performance of the standard GARCH model were as good as that of the gjrGARCH model over the out-of-sample forecasts for returns among other GARCH specifications.

Keywords: MODWT Wavelets Transform, Saudi Arabia Stock Market, Outlier Detections, GARCH Models

JEL Classification Code: M15, M27, M20, C40

1. Introduction

Outliers in time series are defined as a type of data anomalies where unexpected observed values deviate from their expected values, and naturally correspond to critical events (Chandola et al., 2009a; Hawkins, 1980). The problem of outliers' detection has been considered in several

fields of applications such as customized marketing, credit card fraud detection, sensor event detection, fault diagnosis in industry, weather prediction, and financial applications (loan approval, stock market data). The occurrence of outliers in data may be due to several reasons such as poor data quality, low quality measurements (Hoaglin et al., 1986), and can as well indicate interesting and meaningful information that can be represented by periods of high or low volatility particularly in financial time series. Detecting outlier values is beneficial, ~~since these values~~ they contain important information in many application fields (Chandola et al., 2009b; Fileto et al., 2015; Giacometti & Soulet, 2016; Rasheed & Alhajj, 2013). In financial data and stock markets such as index prices, asset prices, outliers are defined as extreme points that deviate a lot from the other data points. Thus, before modeling financial time series data, we first try to identify data points that are unlikely if a certain fitted model is assumed to have generated the data. Financial time series data are frequently messed up with outliers due to the influence of unusual and non-repetitive events. Forecast accuracy in such situations is decreased dramatically due to a carry-over effect of the outliers on the point forecast and a bias in the model parameters estimation.

¹First Author and Corresponding Author. PhD Student, School of Mathematical Science, Universiti Sains Malaysia, Penang, Malaysia [Postal Address: Faculty of Science, University of Ha'il, PO Box 2440 Ha'il - 81451 Kingdom of Saudi Arabia]
Email: khudhayr2019@gmail.com

²Professor, School of Mathematical Science, Universiti Sains Malaysia, Penang, Malaysia. Email: mtahir@cs.usm.my

³Associate Professor, Department of Risk Management and Insurance, Faculty of Business, The University of Jordan, Jordan. Email: mohsleh1966@yahoo.com

⁴Associate Professor, Polydisciplinary Faculty of Larache, University Abdelmalek Essaadi, Morocco. Email: aserkh71@gmail.com

© Copyright: The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

There are many outlier detection methods available in literature, for more details we may refer to (Breunig et al., 2000; Janssens et al., 2011; Kriegel et al., 2008). In our approach we propose an outlier detection method based on discrete wavelet transform (DWT). The wavelet transform is a well-known method in signal processing and has many statistical applications. For a complete guide to wavelet methods for discrete time series we refer to (Percival & Walden, 2000). The DWT transforms a time series into two components: approximation and details. The approximation series is a smoother version of the original series, while detail wavelet coefficients capture detail features that describe frequent movements of the data. In addition the transform allows to obtain a reconstruction of the original series while preserving the original information. This attractive property makes them ideally appropriate to detect anomalies at different time scales. Our aim in this paper is to apply a modified version of the DWT called the maximum overlapping discrete wavelet transform (MODWT). Unlike the DWT which requires a series of power of two, the MODWT is a redundant non-orthogonal transform, designed to process a series of arbitrary size, and can be used to form a multiresolution analysis (MRA) in similar manner as in DWT case.

Assuming we have a time series data with a relatively small number of outliers, in the first step we apply the proposed wavelet transform to the original series, and then we detect and correct any existing outliers in the obtained details using lower and upper fences as described below in section 2.1. After cleaning the details from outliers, we reconstruct the original series using the additive decomposition of the multiresolution analysis of the transform.

This paper is structured as follows: in section 2 we present a brief review of some of the research work on wavelet based outliers detection and define the upper and lower fences. In section 3, we run the MODWT to the closed price series of the Saudi stock market. GARCH model fitting the volatility is presented in section 4. Forecasting performances are reported in section 5, and conclusion is given in section 6.

2. Wavelets and Outlier Detections

The problem of outliers detection using statistical methods has been studied by several authors in literature using different approaches. To name a few, we can cite (Liu et al., 2017) who proposed an online anomaly detection method that analyzed residuals of AR model using an improved recursive wavelet algorithm (IRWT) that can realize on-line wavelet decomposition. The detection algorithm does not use the standard wavelet transform, and is based on the Hidden Markov Model. Their methodology is more data dependent and more suitable for data detection

in the process of unstable regulation. (Hosseinioun, 2016), considered an adaptive ensemble models of Extreme Learning Machines (ELMs.) combined with wavelet transform to forecast the stock market index value outliers. The methods applied to market indexes of Tehran Over-the-Counter Market (OTC) for Petroleum sector, and tried to predict whether abnormal fluctuation will appear or not. (Grané & Veiga, 2010) proposed a wavelet based method for outlier detection and correction that can be applied to the residuals of different volatility models such as: the GARCH, the GJR-GARCH and the autoregressive stochastic volatility model (ARSV) for more details refer to (Go & Lau, 2014; Hongsakulvasu & Liamukda, 2020; Trinh et al., 2020). The procedure runs the standard DWT using the Haar wavelet only, and outliers are identified starting with their maximum as those observations in the original series whose detail coefficients are greater in absolute value than a certain threshold.

As far as we know no research work has been found compared to our wavelet based approach for detecting outliers in the Saudi stock market (Tadawul). In our procedure, we aim to apply the simple traditional method suggested by (Tukey, 1977) combined with the MODWT to detect patches of outliers in the closed price series of Saudi Arabia stock market between Oct. 2011 and Dec. 2019.

Upper and lower fences are built on the Interquartile Range (IQR) which is the difference between third quartile (Q_3) and first quartile (Q_1), thus $IQR = Q_3 - Q_1$. Outliers are defined as the observations that lie outside the interval $[Q_1 - 3IQR, Q_3 + 3IQR]$, whereas, observations that fall 1.5 times interquartile range apart from the first and third quartile, $[Q_1 - 1.5IQR, Q_3 + 1.5IQR]$ are regarded as suspected outliers. The constant of fences which is fixed as 1.5 is considered too liberal for detecting outliers in random normally distributed data (Hoaglin et al., 1986; Schwertman et al., 2004).

3. The MODWT Transform

The discrete wavelet transform is a fairly new way of decomposing time series data and widely used in many applications. In contrast to the Fourier transform which is localized only in the frequency, the wavelet transform is well localized in both time and frequency domain and introduces the concept of multiresolution analysis (Mallat, 1989). The algorithm we propose uses the discrete wavelet transform particularly we use the maximal-overlap discrete wavelet transform (MODWT) which is a modified version of the standard discrete wavelet transform (DWT) and has been discussed in the wavelet literature see (Coifman & Donoho, 1995; Daubechies & Bates, 1993; Percival & Walden, 2000). This transform makes use of the compactly

supported orthonormal wavelets refer to (Percival & Walden, 2000) a complete review of wavelet methods for time series. We note as well that we will be using the “wavelet” R package that implements their MODWT algorithm.

Let $X = (X_1, \dots, X_n)$ be an observed time series, then by applying the MODWT to the series X up to J level, we obtain a decomposition of the series respectively in term of details $d_1, d_2, \dots, d_j$ and smooth (approximation) series $a_1, a_2, \dots, a_j$ such that

$$a_{j-1} = d_j + a_j, j = 2, 3, \dots \quad (1)$$

and

$$X = d_1 + a_1 = d_1 + d_2 + a_2 = d_1 + d_2 + d_3 + a_3 \quad \text{for } J=3 \quad (2)$$

The smooth series is a smoother approximation of the original series and keep track of many features of the data set such as trends, while the detail coefficients remove any trend in the series, but more importantly they are sensitive to spikes and jumps in data. Therefore we will focus on detecting outliers in using only details of the series.

3.1. The MODWT of The Closed Price Series

In order to detect outliers, we run the MODWT up to level $J=2$ to the closed price series X of the Saudi Stock market (Taduwal) of length $n=2027$. Note that this transform does not require the sample size to be a power of 2 as in the standard DWT. The MODWT was run using the LA(8) (Least Asymmetric Daubechies wavelet filter of length 8). The obtained detail and smooth series are of the same length as X .

According to the upper fence 10845 and lower fence 4568 we detected 24 outliers in the series X from date 2014-08-24 to 2014-10-02 as shown by the red vertical dashed lines in Fig-1.

3.2. Outliers Removal Procedure

Our proposed procedure for detecting and removing outliers is based on three main steps, we start by Applying the Tukey method to the original series to check for outliers. Second if such outliers do exist, we then run the MODWT transform and remove them, and third if successful in the previous step, we retain the reconstructed series, otherwise we repeat step two to the reconstructed series until we reach a series without outliers.

Assume that we detected outliers in the original series, We then run the MODWT up to some J level and start detecting outliers in the obtained detail d_j for $j=1, \dots, J$. The choice of the level J depends on the sample size n of the series and the unknown level J_0 up to which they are no more outliers in d_{J_0} . Assuming that the sample size n is large enough and allow us to reach such level J_0 , we set $J = J_0$. We then remove all detected outliers in d_j by locating their positions and setting their values to zero or to the median of the corresponding series. After removal of outliers in d_j , we obtain the modified details $\hat{d}_j, j = 1, \dots, J$ which are then used to reconstruct the original series by applying the multiresolution decomposition as follows:

$$a_{j-1} = \hat{d}_j + a_j, \quad a_{j-2} = \hat{d}_{j-1} + a_{j-1}, \dots \quad (3)$$

$$\hat{X} = \hat{d}_1 + a_1$$

Figure 1: Daily closed Price 2011-2019 and outliers indicated with vertical dashed lines

Figure 2: Plots from top to bottom are the rescaled original series $Y = \frac{x}{1000}$ in blue and the reconstructed in red for $J=2, 4, 5, 7$

Although fast, the implementation of this choice may fail in the case where the sample size n is not large enough to reach the level J_0 of the wavelet transform. To overcome this problem we propose to search recursively for outliers by running again the MODWT transform to the reconstructed series $\hat{X}$ based on the allowed maximum J_{max} level. The main idea of the algorithm is to keep searching for outliers and stops when there are no outliers in the reconstructed series. Based on that, the best choice for J would be to set $J=J_{max}$, which will depend on the size n and the filter length been used in the transform. We should note here that this choice of J does not mean that the algorithm will always run the transform to its maximum level. In fact the algorithm steps are designed to stop when we reach the first reconstruction without outliers. This may happen at any level $J=1, 2, \dots, J_{max}$. Usually the algorithm stops after very few steps. In addition to that when the sample size n is large, the implementation

of the algorithm offers more control by allowing the users to choose a desired level J without recursive search. We can start with the first level and then move to the next higher level until no outliers are detected.

4. Model Fitting

We now consider the time series model that best fit the reconstructed series which is free from outliers. Fig. 1 shows that the daily closed price series X_t is not stationary in the mean. In order to get a stationary series we propose to consider a transformation of the reconstructed series.

$$R_t = 100 \times \ln \left(\frac{X_t}{X_{t-1}} \right) \quad t = 2, \dots, N \quad (5)$$

The return series R_t consists of 2026 observations due to differencing operations. The series R_t has nearly a constant mean but with non-constant variance and high variability between 2014 and 2016. The autocorrelation sample (ACF) of R_t and its square R_t^2 are shown in Fig 3 respectively. The estimated autocorrelation coefficient of R_t at lags 1, 2, 6 and 26 are well outside the test bounds, so the series has some degree of dependence. Also the ACF of the squared R_t^2 shows that they are many lags that are outside the test bounds. The amount of dependence in both series is of practical importance, and an ARMA model of smaller order might be appropriate to account for the amount of autocorrelation in R_t .

The Quantile-Quantile plot in Fig. 3 c) shows that R_t does not come from a normal distribution. The series Kurtosis is 6.9372, that's positive and larger than 3 which indicates a fat-tail distribution. The Skweness value is -0.2459 which implies that the distribution of the series R_t is slightly skewed to the left. The sample ACF of R_t suggest to fit an ARMA(p,q) model with smaller order, but the pattern in the sample ACF of the squared R_t^2 exhibits a higher length of serial autocorrelation and show that there is correlation between the magnitude of change in observations. This mean there is a serial dependence in the variance of the series and this should be considered as evidence that support the use of

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models.

4.1. ARMA Model

Assuming that the returns series R_t is stationary, we choose in the first step to see how well an ARMA model would fit our data. The R package “rugarch” is used in our whole study to identify the best model ARMA(p,q) using the Akaike information criterion (AIC) and the (BIC) criterion.

Table 1 shows that AIC and BIC for ARMA models with p, $q \leq 2$. We see that AIC is minimized by the ARMA(2,2) and BIC is minimal for AR(1), and all other models are nearly equally well, but the AR(1) seems to be a good choice as a model with the smallest numbers of parameters. On the other hand the analysis of residuals from any of these models shows that the sample ACF do not resemble a white noise. This indicate that fitting any of these ARMA models isn't enough and would fail to adequately capture the returns. In addition to that, the normal plot of residuals shows a heavy tail. These results confirm the serial dependence in the variance of the series, and suggest to combine an ARMA with a (GARCH) model which is capable to model both the conditional heteroskedasticity and the heavy-tailed distributions of financial market data.

Figure 3: The sample ACF a) of R_t and b) its square R_t^2 respectively, and c) the sample quantiles relative to the normal distribution

Table 1: Some of the best ARMA (p,q) models and their corresponding AIC and BIC

P	q	AIC	BIC
2	2	2.7851	2.7990
0	2	2.7912	2.7995
2	0	2.7917	2.8000
1	1	2.7938	2.8021
1	0	2.7953	2.8000

4.2. GARCH Model

We know consider to model the series R_t by choosing AR(1) for the mean model and GARCH model for the variance. The “ugarchfit” function from “rugarch” package is used again to estimate a GARCH(1,1) model to this series.

Assume that our model is given by

$$R_t = \mu_t + a_t \quad \text{with} \quad a_t = \sigma_t \varepsilon_t \quad (6)$$

where a_t is a GARCH(1,1)

$$E(a_t^2 / \mathcal{F}_{t-1}) = \sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (7)$$

where ε_t is an i.i.d. white noise (independent and identically distributed random variables), $\mathcal{F}_{t-1}$ is the information given by the past values $a_{t-1}, a_{t-2}, \dots$. The parameters ω, α_1 and β_1 are such that

$$\omega > 0, \alpha_1 \geq 0, \beta_1 \geq 0 \text{ and } \alpha_1 + \beta_1 < 1$$

The conditional mean model AR(1) is given by

$$R_t = \mu + \phi_1 (R_{t-1} - \mu) + a_t \text{ with } a_t = \sigma_t \varepsilon_t \quad (8)$$

where μ is the mean value of R_t and ϕ_1 is the AR(1) parameter.

Table 2 below shows the computed AIC and BIC for different combinations of GARCH (p_G, q_G) model with $p_G, q_G \leq 2$ to the variance σ_t^2 with ε_t a Gaussian white noise and the p-values of the corresponding model parameters.

Note that the AR(1) parameters has p-value of 0.0000 in all these models.

From table 2 we can see that the best model that minimize both the AIC and BIC is the GARCH(1,1), and summary of all the estimated parameters of GARCH(1,1) is given by table 3. The sum of the estimated parameters equals $\hat{\alpha}_1 + \hat{\beta}_1 = 0.9966$ which indicates that volatility shocks are persistent.

In order to check the adequacy of the fitted model we explore several graphical and statistical diagnostic tests. For a properly fitted GARCH(1,1) model, we consider

the standardized residuals $\hat{\varepsilon}_t = \frac{\hat{a}_t}{\hat{\sigma}_t}$ which are the ordinary

residuals divided by their estimated conditional standard deviations based. If the model fits well, then neither $\hat{\varepsilon}_t$ nor $\hat{\varepsilon}_t^2$ should exhibit serial correlation. The output of “ugarchfit” includes the Weighted Ljung-Box Test applied to the standardized and squared standardized residuals. All the approximate p-value of these tests are larger than 0.05, and hence strongly support that there is no serial correlation in $\hat{\varepsilon}_t$ and $\hat{\varepsilon}_t^2$. In addition, the output provides the Weighted ARCH- LM Tests which is a test for autoregressive conditional heteroskedasticity (ARCH effect) in the residuals (Engle 1982), the approximate p-values show no autocorrelation among the squared residuals $\hat{\varepsilon}_t^2$ which mean there is no ARCH effect in the model. Furthermore all p-values of the Nyblom stability test are larger than 0.05 and hence the parameter values are constant

Table 2: Best GARCH (p_G, q_G) under Gaussian white noise, their AIC and BIC and p-values

Model	AIC	BIC	$\hat{\mu}$	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
ARCH(1,0)	2.6959	2.7080	0.4801	0.0000	0.0000	X	X	X
GARCH(1,1)	2.5123	2.5274	0.0041	0.0025	0.0000	X	0.0000	X
GARCH(1,2)	2.5135	2.5316	0.0053	0.0024	0.0001	X	0.0000	0.0000
GARCH(2,1)	2.5139	2.5320	0.0102	0.1180	0.1406	0.999	0.0000	X
GARCH(2,2)	2.5146	2.5357	0.0055	0.0025	0.0025	1.000	0.0000	0.0000

Table 3: Parameter estimate for AR(1) and GARCH(1,1) under Gaussian white noise

Parameter	Estimate	Std. Error	t value	Pr(> t)
μ	0.05859	0.020465	2.86316	0.00419
ϕ_1	0.16912	0.025457	6.64326	0.00000
ω	0.00603	0.002004	3.01161	0.00259
α_1	0.06538	0.008990	7.27291	0.00000
β_1	0.93123	0.008146	114.309	0.00000

The output include also the Goodness-of-Fit tests which compare the empirical distribution of the standardized residuals with the specified theoretical Gaussian distribution for the white noise. The small p -values strongly reject our assumption of Gaussian white noise process ε_t . This is confirmed by the normal quantile plot of the standardized residuals.

In order to improve the fitting, we refit our model under the assumption that the white noise has a t-distribution (Student). The output gives AIC=2.3817 and BIC=2.3998 which are smaller to those obtained under a Gaussian white noise. The weighted Ljung-Box and ARCH-LM test statistics and their approximate p -values all indicate strongly that the estimated model for the conditional mean and variance are adequate.

Table 4 shows the model parameter estimate under t-distribution white noise, and indicates that all parameters are significant. The sum of the estimated parameters equals $\hat{\alpha}_1 + \hat{\beta}_1 = 0.9861$, and the unconditional variance reduced from 1.786015 to 0.9344001.

Fig.4 plots the sample ACF of the standardized and squared standardized residuals. This shows that there are two lags slightly outside the test bounds in the squared standardized residuals.

The p-values of the goodness-of-fit test statistics are very large and strongly reject that the white noise process ε_t is Gaussian. This is supported by Fig. 5 b) which shows that a t-plot with $df=4$ for the standardized residuals $\hat{\varepsilon}_t$ is more adequate.

The fitted GARCH model assumes a symmetric volatility response to the market news. The simpler GARCH model assumes that positive and negative shocks have the same effect on volatility. It has been suggested in the financial literature that negative shocks in the market have a larger impact on volatility than positive shocks of the same magnitude, and as a result alternative asymmetric GARCH models should be considered. (Glosten et al., 1993) developed the GJR and EGARCH models, respectively, to take into account this asymmetry. The GJR model is similar to the EGARCH model, both of them embody the asymmetries in volatility in response to negative and positive shocks. Another interesting model specification is the APARCH (Asymmetric–Power–ARCH(p,q)) of Ding, Granger, and Engle (1993). The forecasting performance of the our standard GARCH are evaluated in term of prediction errors against these alternative asymmetric models in the next section.

Table 4: Parameter estimate for AR(1) and GARCH(1,1) under t-distribution white noise

Parameter	Estimate	Std. Error	t value	Pr(> t)
μ	0.05345	0.01811	2.9505	0.00317
ϕ_1	0.15656	0.02277	6.8758	0.00000
ω	0.01292	0.00500	2.5816	0.00983
α_1	0.07295	0.01914	3.8098	0.00013
β_1	0.91321	0.02075	44.0024	0.00000

Figure 4: The sample ACF of the a) standardized residuals $\hat{\varepsilon}_t$ and b) their squares $\hat{\varepsilon}_t^2$ under the t-distribution

Figure 5: a) The empirical density estimates of the standardized residuals $\hat{\varepsilon}_t$ and its t-plot. compared to the Gaussian distribution in red line, b) the t-distribution quantiles plot of $\hat{\varepsilon}_t$

5. Forecasting

An important task of modeling conditional volatility is to generate accurate forecasts of future volatility, especially over short term. Assume that the basic GARCH(1,1) model as given in (7) is estimated over the time period $t = 1, 2, \dots, n$. Let T be the time forecast origin, the GARCH(1,1) model at time $T+1$ can be written as

$$\sigma_{T+1}^2 = \omega + \alpha_1 a_T^2 + \beta_1 \sigma_T^2 \quad (9)$$

The optimal, in terms of mean-squared error, forecast of σ_{T+1}^2 given information at time T is given by $E(\sigma_{T+h}^2 / \mathcal{F}_T)$. For $h = 1$ the 1-step-ahead volatility forecast is

$$\hat{\sigma}_T^2(1) = \hat{\omega} + \hat{\alpha}_1 \hat{a}_T^2 + \hat{\beta}_1 \hat{\sigma}_T^2 \quad (10)$$

now $a_T^2 = \sigma_T^2 \varepsilon_T^2$, which gives $\sigma_{T+1}^2 = \omega + (\alpha_1 + \beta_1) \sigma_T^2 + \alpha_1 \sigma_T^2 (\varepsilon_T^2 - 1)$, and $E(\sigma_{T+1}^2 / \mathcal{F}_T) = \omega + (\alpha_1 + \beta_1) \sigma_T^2$ because $E((\varepsilon_T^2 - 1) / \mathcal{F}_T) = 0$ and $E(\sigma_T^2 / \mathcal{F}_T) = \sigma_T^2$

At time $T+2$ this gives $\sigma_{T+2}^2 = \omega + (\alpha_1 + \beta_1) \sigma_{T+1}^2 + \alpha_1 \sigma_{T+1}^2 (\varepsilon_{T+1}^2 - 1)$

Thus at time $T+2$, the 2-step volatility forecast is

$$\hat{\sigma}_T^2(2) = \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1) \hat{\sigma}_T^2(1)$$

The general h -step ahead forecast for σ_{T+h}^2 is computed using a simple recursion,

$$\hat{\sigma}_T^2(h) = \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1) \hat{\sigma}_T^2(h-1) \quad (11)$$

In order to test and compare the forecasting ability of the fitted models, we have to evaluate how well the model performs on the out-of-sample data. The in-sample data is the first portion of the series that was used to fit the model. The good in-sample estimation does not guarantee an accurate forecast. The forecasting performance is often judged by the out-of-sample data, and the forecasting ability can be measured using the corresponding forecast errors given by

$$e_{T+h} = \sigma_{T+h}^2 - \hat{\sigma}_T^2(h) \quad (12)$$

An important practical problem is that the h -step ahead volatility σ_{T+h}^2 is not directly observable. As a common practice α_{T+h}^2 is used to proxy σ_{T+h}^2 , and the variance σ_{T+h}^2 is set equal to the squared daily residual $\hat{a}_{T+h}^2 = (r_{T+h} - \hat{\mu}_{T+h})^2$. These residuals are computed starting at the 1-step-ahead as the difference between the 1-step-ahead predicted conditional mean $\hat{\mu}_{T+1}$ and the observed return series r_{T+1} . The measure of forecast accuracy can be summarized using different forecast error metrics; the root mean square error (RMSE) and the root mean absolute error (RMAE) given as follows:

$$MSE = \frac{1}{T^*} \sum_{t=1}^{T^*} \left(\sigma_{T+t}^2 - \hat{\sigma}_T^2(t) \right)^2 \text{ and } MAE = \frac{1}{T^*} \sum_{t=1}^{T^*} \left| \sigma_{T+t}^2 - \hat{\sigma}_T^2(t) \right|$$

Where T^* is the size of the out-of-sample and $RMSE = \sqrt{MSE}$ and $RMAE = \sqrt{MAE}$. Vilhelmsson (2006) characterizes the MSE as being sensitive to outliers and argues that the mean absolute error (MAE) is more robust to outliers. Table 5 shows four different GARCH models that are used to compare the forecast performances.

Table 5: Forecast Performance of the Estimated GARCH Models

Forecast Criteria	sGARCH(1,1)	eGARCH(1,1)	gjrGARCH(1,1)	apARCH(1,1)
RMSE	2.862586	2.85154	2.866016	2.856134
RMAE	1.12044	1.122422	1.113841	1.123775

Figure 6: The returns R_t (blue solid line) and the fitted (dashed red line) series over the period Aug.- Dec. 2019

6. Conclusion

The main purpose of this study was to detect and remove outliers in the first stage, and then find a volatility model which is the best fit for the purpose of forecasting. We start by running the MODWT up to the maximum allowed level $J = J_{\max} = 8$ to the closed price series X_t of length $n = 2027$. We applied our algorithm as described in section 3.2 and set any detected outlier to the median to obtain the modified detail $\hat{d}_j$. For illustration we choose to reconstruct the original series using arbitrary level $J=2, 4, 5$ and 7 of the transform as shown in Fig. 2. In each panel and for a particular value of J , we plot the original series X_t and the reconstructed series $\hat{X}_t$. The plots for panel $J=2, 4$ and 5 show that the reconstructed series is hardly different from the original one and is still contaminated with outliers. We get a clean series from outliers until we reach the levels $J_{\max}=8$ of the transform. It important to note here that the search and removal of outliers in the details $\hat{d}_j$ has shown that the detected number of outliers was not getting smaller as we move from lower to higher scales, but was varying arbitrary and we did observe zero outliers until the last level $J = J_{\max}$.

The reconstructed clean series $\hat{X}_t$ is then used as a substitute for the original one and was then subject to further analysis. In section 4, we explore the statistical modeling of

the returns series computed from the reconstructed one. The plot of the series X_t in Fig. 1 shows that X_t is not stationary. In an attempt to get a stationary series we take the first difference of the logarithm of the series to get the returns R_t . Initially we applied ARMA models in the aim to model the series and remove the serial correlations in R_t . But the analysis of the sample ACF of residuals and their squares indicated that the ARMA model alone was not good enough to adequately capture the returns.

The presence of serial dependence in the variance of the series R_t suggest to combine an ARMA with a GARCH model which is capable to model both the conditional heteroskedasticity and the heavy-tailed distributions. GARCH models have the ability to capture volatility clustering which is a characteristic of financial time series. GARCH model analysis involves estimation of two distinct models, one for the conditional mean and the other for conditional variance.

Based on the ARMA analysis of our series in section 4.1, we fitted an AR(1) model to the conditional mean. A GARCH(1,1) model has been fitted to the conditional variance, and in order to evaluate our model in terms of its ability to forecast future returns, the data set has been split into in-sample from the period 13-08-2011 to 11-03-2019 and out-of-sample of size 200 from the period 12-03-2019

to 31-12-2019. The in-sample data was used to estimate the GARCH(1,1) under normal and then t-Student distributions. The plot in Fig. 6, shows a portion of the predicted returns using the whole series and indicate the goodness of our fitted model. The last step was to check the adequacy of the model, and different volatility models that capture the asymmetrical effect due to responses for good and bad news in financial markets. Table 5 provides the measures of forecast evaluation in terms of RMSE and RMAE for each model based on the out-of-sample returns.

The results indicate that there is a relatively small difference among forecasting performance measures, Taken into account that the mean absolute error (MAE) is more robust to outliers than the MSE (Wilhelmsson, 2006), The forecasting results show that standard GARCH (1,1) model is found to be the most preferred among all the models to study the volatility behavior of our returns.

References

- Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*.
- Chandola, V., Banerjee, A., & Kumar, V. (2009a). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1-58.
- Chandola, V., Banerjee, A., & Kumar, V. J. A. c. s. (2009b). Anomaly detection: A survey. 41(3), 1-58.
- Coifman, R. R., & Donoho, D. L. (1995). Translation-invariant denoising. In: *Wavelets and statistics* (pp. 125-150). Springer.
- Daubechies, I., & Bates B. J. (1993). Ten Lectures on Wavelets. *The Journal of the Acoustical Society of America*, 93, 1671. <https://doi.org/10.1121/1.406784>
- Fileto, R., May, C., Renso, C., Pelekis, N., Klein, D., & Theodoridis, Y. (2015). The Baquara2 knowledge-based framework for semantic enrichment and analysis of movement data. *Data & Knowledge Engineering*, 98, 104-122.
- Giacometti, A., & Soulet, A. (2016). Anytime algorithm for frequent pattern outlier detection. *International Journal of Data Science and Analytics*, 2(3-4), 119-130.
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5), 1779-1801.
- Go, Y. H., & Lau, W. Y. (2014). Asymmetric information spillovers between trading volume and price changes in Malaysian futures market. *Journal of Asian Finance, Economic and Business*, 1(3), 5-16. <https://doi.org/10.13106/jafeb.2014.vol1.no3.5>.
- Grané, A., & Veiga, H. (2010). Wavelet-based detection of outliers in financial time series. *Computational Statistics & Data Analysis*, 54(11), 2580-2593.
- Hawkins, D. M. (1980). *Identification of outliers* (Vol. 11). New York, NY: Springer.
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396), 991-999.
- Hongsakulvasu, N., & Liamukda, A. (2020). Asian Stock Markets Analysis: The New Evidence from Time-Varying Coefficient Autoregressive Model. *Journal of Asian Finance, Economics, and Business*, 7(9), 95-104. <https://doi.org/10.13106/jafeb.2020.vol7.no9.095>
- Hosseinioun, N. (2016). Forecasting outlier occurrence in stock market time series based on wavelet transform and adaptive ELM algorithm. *Journal of Mathematical Finance*, 6(1), 127-133.
- Janssens, J. H., Postma, E. O., & van den Herik, J. H. (2011). Maritime anomaly detection using stochastic outlier selection. *MAD 2011 Workshop Proceedings*.
- Kriegel, H. P., Schubert, M., & Zimek, A. (2008). Angle-based outlier detection in high-dimensional data. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*.
- Liu, F., Su, W., Zhao, J., & Liang, X. (2017). On-line Detection Method for Outliers of Dynamic Instability Measurement Data in Geological Exploration Control Process. *Sains Malaysiana*, 46(11), 2205-2213.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674-693.
- Percival, D. B., & Walden, A. T. (2000). *Wavelet methods for time series analysis* (Vol. 4). Cambridge, UK: Cambridge University Press.
- Rasheed, F., & Alhajj, R. (2013). A framework for periodic outlier pattern detection in time-series sequences. *IEEE Transactions on Cybernetics*, 44(5), 569-582.
- Schwertman, N. C., Owens, M. A., & Adnan, R. (2004). A simple more general boxplot method for identifying outliers. *Computational Statistics & Data Analysis*, 47(1), 165-174.
- Trinh, Q. T., Nguyen, A. P., Nguyen, H. A., & NGO, P. T. (2020). Determinants of Vietnam Government Bond Yield Volatility: A GARCH Approach. *Journal of Asian Finance, Economics, and Business*, 7(7), 15-25. <https://doi.org/10.13106/jafeb.2020.vol7.no7.015>
- Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2). Reading, MA.
- Wilhelmsson, A. (2006). GARCH forecasting performance under different distribution assumptions. *Journal of Forecasting*, 25(8), 561-578.