University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 25/2021 (373)

Applying Hybrid ARIMA-SGARCH in Algorithmic Investment Strategies on S&P500 Index

Nguyen Vo
Robert Ślepaczuk

Warsaw 2021

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

Applying Hybrid ARIMA-SGARCH in Algorithmic Investment Strategies on S&P500 Index

Nguyen Vo[a], Robert Ślepaczuk[b*]

[a] *University of Warsaw, Faculty of Economic Sciences, Quantitative Finance Research Group*
[b] *University of Warsaw, Faculty of Economic Sciences, Quantitative Finance Research Group, Department of Quantitative Finance*
* *Corresponding author: rslepaczuk@wne.uw.edu.pl*

Abstract: This research aims to compare the performance of ARIMA as linear model with that of the combination of ARIMA and GARCH family models to forecast S&P500 log returns in order to construct algorithmic investment strategies on this index. We use the data collected from Yahoo Finance. The dataset has daily frequency and covers the period from 01/01/2000 to 31/12/2019. By using rolling window approach, we compare ARIMA with the hybrid models to examine whether hybrid ARIMA-SGARCH and ARIMA-EGARCH can really reflect the specific time-series characteristics and have better predictive power than simple ARIMA model. In order to assess the precision and quality of these models in forecasting, we decide to compare their equity lines, their forecasting error metrics (MAE, MAPE, RMSE, MAPE) as well as their performance metrics (annualized return compounded, annualized standard deviation, maximum drawdown, information ratio and adjusted information ratio). The results show that the hybrid models outperform ARIMA and the benchmark (Buy&Hold strategy on S&P500 index). These results are not sensitive to varying window sizes, the type of distribution and the type of the GARCH model.

Keywords: algorithmic investment strategies, ARIMA, ARIMA-SGARCH, ARIMA-EGARCH, hybrid model, forecast stock returns, model robustness

JEL codes: C4, C14, C45, C53, C58, G13

Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Introduction

Over the past few decades, time series forecasting in finance has been an interesting and important research area. It has attracted the attention of not only the researcher community but also investors, speculators, and governments. The main aim of time series modeling is to carefully measure and analyze the historical observations of the time series in order to develop the most appropriate models. The most important function of these models is to forecast future values of the time series, i.e. to predict the movements, behaviors, and changes, usually by reflecting the characteristics of the historical observations. It is obvious that to obtain adequately low forecasting errors, a proper process for model fitting needs to be taken into consideration. Along with the development of more advanced forecasting techniques, a lot of efforts have been put into improving forecasting accuracy by choosing, testing, and fitting more efficient models. As a result, various important theories and assumptions about modeling forecast have evolved. Especially, the analyses of the time series of an essential US stock index like the S&P500 have never failed to get attention and efforts from those interested in quantitative finance.

One of the most popular and frequently used time-series models is the **A**uto**r**egressive **I**ntegrated **M**oving **A**verage (ARIMA). In this model, there is a linear relationship between past observation values (autoregressive) and random errors (moving average) where random errors are assumed to be independent and identically distributed (i.i.d.) with a mean of zero and a constant variance σ_ε^2 over time. The term "constant variance" is also known as homoskedasticity. It is not so surprising that financial time series often do not follow this assumption, and S&P500 Index is not an exception. Its returns can be extremely volatile during booms and busts. This means, the existence of volatility clustering in time series can affect the forecasting performance of the mean models like ARIMA. Therefore, most researchers started to use symmetric **G**eneralized **A**utoregressive **C**onditional **H**eteroscedasticity (SGARCH) and its family models when modeling volatility in order to obtain accurate forecasts. This is the reason why in our study we evaluate and compare the forecasting performance among the ARIMA model, the hybrid ARIMA-SGARCH and ARIMA-EGARCH (EGARCH-Exponential **G**eneralized **A**utoregressive **C**onditional **H**eteroscedasticity- is a modified version of GARCH).

In that context, the paper addresses one main hypothesis (RH): *whether the ARIMA(p,1,q)-SGARCH(1,1) (hybrid model) with window size s = 1000 can generate a trading strategy that outperforms ARIMA(p,1,q).*

Based on this hypothesis, a few research questions are constructed:

RQ1. *Is the result obtained from the main test robust to varying family of GARCH model?*

RQ2. *Does the hybrid model with EGARCH outperform the one with SGARCH?*

RQ3. *Is the result obtained from the main test robust to varying window sizes?*

RQ4. *Is the hybrid model sensitive to different window sizes?*

RQ5. *Is the result obtained from the main test robust to varying distributions?*

RQ6. *Is the hybrid model sensitive to different distributions?*

In order to verify the main hypothesis and answer the research questions mentioned above, an empirical research is conducted based on the dataset of S&P500 index. The data are collected on a daily basis over the period from 01/01/2000 to 31/12/2019. Firstly, we conduct a rolling forecast based on the ARIMA model with window size s = 1000. The optimized combination of p and q which has the lowest Akaike Information Criterion (AIC) is used to predict the return for the next day. For the purpose of the out-of-sample results, the vector of forecasted values has the length of 3530 elements (starting on 20/12/2005). Secondly, we describe and review our implementation of ARIMA(p,1,q)-SGARCH(1,1) models with generalized error distribution (GED) and window size equal to 1000 where optimized ARIMA(p,1,q) is taken from the 1st step. Thirdly, we evaluate the performance of SGARCH with different window sizes as well as various distributions to check the sensitivity of the results obtained in the main test. EGARCH, known as another family of GARCH models, is also applied in the sensitivity analysis in order to check the robustness of our initial assumptions. In order to examine the precision and the quality of these models in predicting and their efficiency in algorithmic investment strategies, we compare their equity lines, their error metrics (Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE)) and their performance metrics: annualized return compounded (ARC), annualized standard deviation (ASD), maximum drawdown (MD), information ratio (IR) and adjusted information ratio (AIR). We expect that hybrid models can help build more efficient algorithmic investment strategies which outperform ARIMA.

The paper is structured as follows: Chapter 1 provides a literature review and an overview of ARIMA, SGARCH, EGARCH and the hybrid models. Chapter 2 describes all the details concerning the dataset. Chapter 3 explains the methodology of ARIMA and the hybrid models as well as the description of trading strategy construction and the technique of generating of buy/sell signals for the investment strategy. Chapter 4 discusses the empirical results while chapter 5 conducts some sensitivity analysis to determine whether the result is robust to initial assumptions. The last chapter draws conclusions and makes suggestions for future works.

1. Literature Review

According to Strickland (2014), there are various ways to classify methods for time series analysis. Some examples are frequency-domain and time-domain, parametric and non-parametric. The category used in this study is linear and non-linear regression. Regression analysis is a process of estimating the relationship between a dependent variable and one (or more than one) independent one(s). If there is only a single independent variable, this is known as simple linear regression, otherwise it is called multiple regression. Both of these models assume that the dependent variable is continuous.

1.1 Linear Forecasting Models - Autoregressive Integrated Moving Average Models

Box and Jenkins first introduced Autoregressive Integrated Moving Average Models (ARIMA) in 1976. This model describes the linear relationship between past observation values and random errors (also known as shocks or disturbances). In order to estimate ARIMA model correctly, we must identify and remove non-stationarity through differencing, hence the differences between a value and its lagged values $(y_t - Ly_t)$ need to be calculated. ARIMA model can be regarded as an extension of the ARMA model.

There are many related studies in modelling and forecasting stock prices using ARIMA models. For instance, Ariyo et al. (2014) revealed an extensive process describing how to obtain the most appropriate ARIMA model to anticipate stock prices (based on the smallest value of SIC). Later, Kamruzzaman et al. (2017) calculated returns by using Relative Difference method and chose ARIMA(2,1,2) model (based on the smallest value of AIC) as the most superior model for forecasting the stock market returns of DSE in Bangladesh. In addition, Abbasi et al. (2017) applied this linear process into flying cement industry and suggested that ARIMA(1,2,1) was a parsimonious model for forecasting cement stock prices in their case study.

1.2 Non-Linear Forecasting Models

1.2.1 The Autoregressive Conditional Heteroskedasticity – ARCH(q)

This model was first proposed by Engle (1982) to predict the conditional variance of return series. Despite the key strength as a simple model which produces volatility estimates with positive excess kurtosis[1], its weaknesses should also be taken into consideration. Firstly, due to possible large value of the lag q, it could lead to a large number of parameters to be estimated. Hence it may result in difficulties to determine parameters (Ghani I. M. Md et al., 2019).

[1] It means fat tails relative to the normal distribution which is in line with empirical observations about returns.

Secondly, as it is well known in practice, the stock prices or financial assets in general react differently to positive and negative shocks. However, ARCH models assume these kinds of shock have the same effects on the volatility as it depends on the square of the previous shocks. As a result, this weakness should be taken into consideration in forecasting when applying ARCH models (Chua C. L. & et al., 2019). Additionally, since ARCH models respond slowly to large shocks, they are likely to overpredict the volatility (Jansen S., 2020). Furthermore, the ARCH models do not provide any new insights for understanding the source of volatility of financial time series. They merely provide a mechanical way to describe the behavior of the conditional variance and give no indication of what causes such behavior to occur (Tsay, 2010).

1.2.2 Generalized Autoregressive Conditional Heteroscedasticity – GARCH(p,q)

GARCH model is considered as an extension of ARCH model and was proposed by Bollerslev (1986) by developing the symmetric GARCH (SGARCH). As demonstrated by many researchers and studies, SGARCH (1,1) process is able to represent the majority of the time series (Engle, 2001). The dataset which requires a model of higher orders like SGARCH(1,2) or SGARCH(2,1) is very rare (Bollerslev, 1986). However, financial time series inherits many characteristics that SGARCH is not able to incorporate well. Therefore, extensive generalizations with further features have been put forward in the literature.

One of the most essential properties of volatility that should be taken into consideration is the leverage effect, which describes the fact there is a difference in reaction of volatility between notable price rises and notable price falls. This has led to the introduction of asymmetric GARCH models by purely adjusting the error term in the variance equation with a parameter to be responsible for this effect. These models were initially proposed by Engle R. F. & et al. (1990). Nowadays, there are a variety of such models including Higgins M. L. & et al. (1992) in which a nonlinear asymmetric GARCH (or N-GARCH) accounts for the leverage effect. Later, Glosten L. et al. (1993) introduced GJR-GARCH to precisely build up the volatility response from negative market shocks with an indicator function, while Q-ARCH launched by Sentana (1995) established asymmetric effects of both positive and negative shocks. The most recent model that allows for an asymmetric response due to leverage effects is Exponential GARCH (known as EGARCH) which is discussed in the next chapter.

1.2.3 The Conditional Variance Equation: Exponential GARCH.

Presented by Nelson (1991), EGARCH is supposed to avoid imposing constraints on the coefficients by specifying the logarithm of the conditional volatility. In reality, “bad news” typically has a larger impact on volatility than "good news". In other words, by applying EGARCH, we can mitigate the disadvantage of using GARCH in which negative innovations tend to increase the volatility more than positive innovations with the same magnitude. EGARCH is applied in this thesis in the section concerning the comparison of its performance with SGARCH's.

1.2.4 Underlying Return Distributions

GARCH models assume that the distribution of returns is normally distributed. However, this assumption has been proved in empirical financial market to be inaccurate. It is because the distribution of financial returns tends to be leptokurtic (Wilhelmsson (2006), Hafezian et al. (2015)). It means that the tails are heavier in comparison with normal distribution. As a consequence, several fat tail distributions are applied in order to overcome this shortcoming. For instance, Student-t was introduced by Bollerslev in 1987, the generalized error distribution (GED) by Nelson in 1991 and their skewed versions, which are all for leptokurtic distribution analysis.

1.3 The Hybrid ARIMA-GARCH

The class of ARIMA models with ARCH errors was proposed initially by Weiss (1984). The techniques were applied to U.S. macroeconomic time series. This approach later on was adopted and extended by many researchers for modelling time series in various fields (e.g. Jabłecki et al. (2015), Hauser and Kunst (1998), Kijewski and Ślepaczuk (2020))

Alongside the theory development of the hybrid models in forecasting economic time series, Yaziz et. al (2013) analyzed the performance of ARIMA-GARCH in forecasting gold price. The empirical results of 40-day gold price data series indicate that the hybrid ARIMA(1,1,1)-GARCH(0,2) model provides superior results and effectively improves evaluating and predicting precision in comparison to the linear models. Later, Sun (2017) proposed the hybrid models to model and predict the equity returns for three US benchmark indices namely Dow Transportation, S&P500 and VIX. Based on the observed results, suggested hybrid models are appropriate for anticipating the equity returns but have not been explored in the previous works. The latest work discussed in this chapter is by Mustapa and Ismail (2019). They presented the assessment in building the best fitted ARIMA-GARCH model to generate predicting values of the S&P500 stock prices. ARIMA(2,1,2)-GARCH(1,1) model was figured out to be the most appropriate model for forecasting stock prices.

Collectively, the presented papers reveal heteroskedasticity can affect the validity or power of statistical tests when using ARIMA models, the ARCH effect should be considered. Furthermore, mentioned studies also indicate the combination of ARIMA and the family of GARCH should be expected to perform well in modelling financial time series. In our study, we fit an optimal ARIMA-SGARCH as well as ARIMA-EGARCH in different window sizes and various return distributions in algorithmic investment strategies on S&P500 index. It is based on the AIC criterion, each day, using the rolling window approach.

2. Data Analysis

In this chapter, we provide a complete data analysis and a model-fitting procedure for the logarithmic returns of S&P500 Index.

2.1 Data Fetching and Preprocessing of Historical Data

We fetch the historical data from Yahoo Finance. Given the hybrid ARIMA-GARCH model proposed in Chapter 1, the data for S&P500 index are collected with the period of 19 years for sufficiently reliable model fitting and forecasting purposes.

The first step in the process of cleaning the data is to delete all missing and invalid data from the time series. If daily observations for a given instrument are missing, the daily stock return is set at 0%. Then we transform the adjusted price into daily logarithmic return which is calculated according to the following formula:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{2.1}$$

There are many reasons for choosing log returns instead of normal prices as a variable to forecast in this study. First, they can be added across time periods in order to create cumulative returns. Second, it is easy to convert between log return and simple return. Last but not least, log return follows normal distribution. Why is it an important advantage? You probably know this kind of distribution is solely dependent on the mean and the standard deviation of the sample. Based on these characteristics, any variable that exhibits normal distribution is feasible to be forecasted with higher accuracy, and our variable, in this case, is log return. Moreover, stock prices cannot be modelled by normal distribution because they have negative side, and stock prices cannot fall below zero. In other words, prices are lognormally distributed, then the logarithm of each price will have a normal distribution. These relationships can be expressed by equation below:

$$\ln\left(\frac{P_t}{P_0}\right) = \ln\,(P_{t)} - \ln\,(P_{0)} = \ln(1 + r_t) \tag{2.2}$$

2.2 Descriptive Statistics

Table 1 presents the descriptive statistics of the adjusted closing prices and log returns of S&P 500 Index for the whole dataset.

Table 1. Descriptive Statistics for S&P500 (Jan 2000 – Dec 2019)

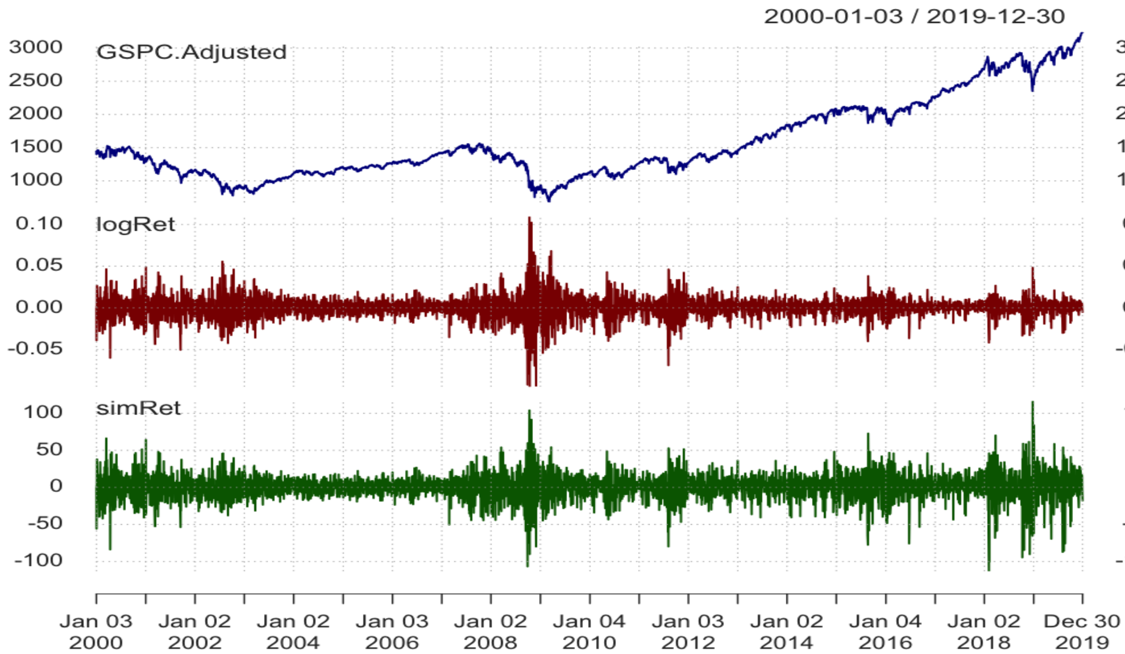
Descript Statistics	S&P500 Original Prices	Log Returns
Min	676.5300	-0.0947
1st Quantile	1151.5349	-0.0047
Median	1360.9550	0.0005
Arithmetic Mean	1574.6801	0.0002
3rd Quantile	1986.2225	0.0057
Max	3240.0200	0.1096
Skew	0.9886	-0.2295
Kurtosis	-0.0715	8.6448
Standard Error Mean (se)	8.2559	0.0002
Standard Deviation (sd)	585.5315	0.0119

Note: The table covers the period between 01/01/2000 and 31/12/2019.

As can be seen from Figure 1, there are a few periods such as 2008, 2011, 2015, and 2018 that show high volatility of returns. Therefore, we can expect to build more accurate forecasting models if we are able to mitigate and "smooth" such periods. This is further explained in Methodology section.

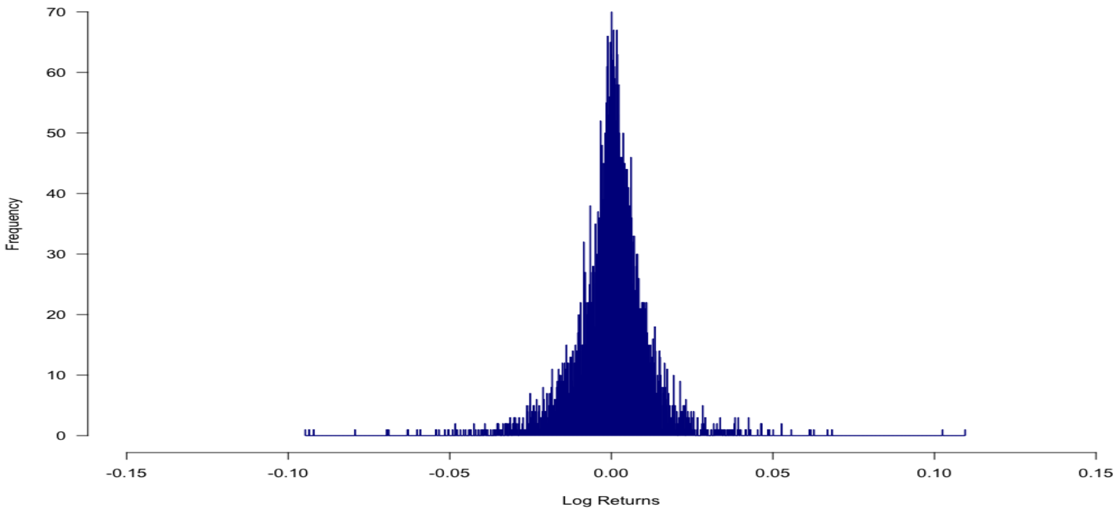
Next, when we consider the central tendency, Table 1 shows two types of estimation as mean and median. The central tendency of a distribution is an estimation of the "center" of distribution, in this case of stock prices and log returns. If mean (or average) is computed by added up all the values and divided by the number of values, median, on the other hand, is the middle value or midpoint in data (also known as 50th percentile). In a normal distribution, these two metrics fall at the same midline point. In other words, mean and median are equal. In this study, with mean of 1574.6801 and median of 1360.9550, our initial assessment about stock prices of S&P500 is that they are not normally distributed. And it makes sense since normal distribution has 2 sides while stock prices cannot be negative (below zero).

Figure 1. S&P500 index prices with its 1st differences and log returns



Note: The fluctuations of S&P500 index prices, its first differences and log returns of S&P500 index prices in the period between 01/01/2000 and 31/12/2019.

Figure 2. Histogram for log returns – S&P500

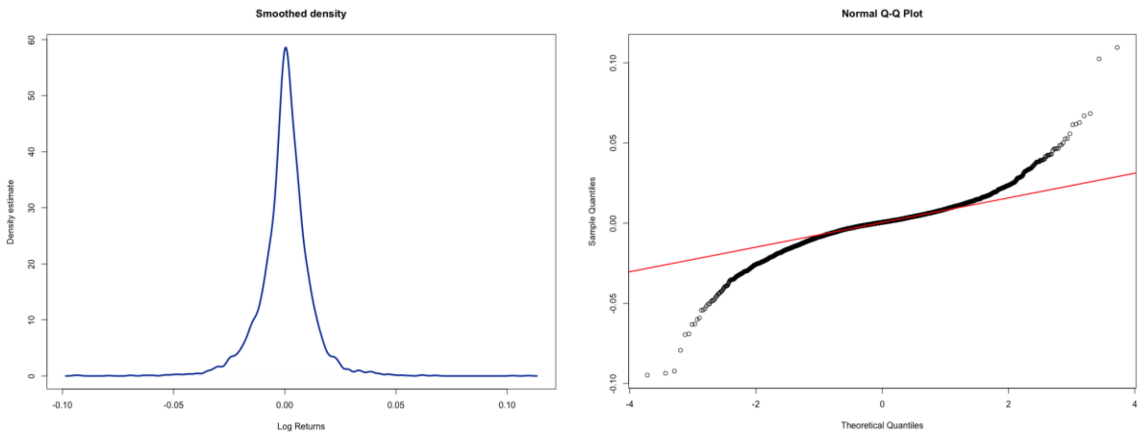


Note: The histogram of S&P500 index log returns in the period between 01/01/2000 and 31/12/2019.

Figure 2 shows that log return series, on the other hand, has different characteristics of distribution in comparison to stock prices dataset. Intuitively, log returns are normally distributed because mean and median values are close to each other (0.0002 & 0.0005, respectively). Furthermore, values of 1st & 3rd quantile (-0.0047 & 0.0057, respectively) as well as min & max values (-0.0947 & 0.1096, respectively) are quite symmetric. It is the main reason why we use log returns to build models.

However, kurtosis value of 8.6448 is larger than 3 (hence it is named leptokurtic) and skew value of -0.2295 (near but below zero), we could say the log returns series is similar to double exponential distribution[2]. This kind of distribution is symmetric, but compared to the normal one, it has a stronger (higher and sharper) peak, more rapid decay, and heavier tails. Furthermore, looking at the histograms in Figure 3, it is perfectly clear that they show certain similarities to the normal distribution.

Figure 3. Smoothed density & Q-Q plot for log returns – S&P500



Note: Smoothed density & Q-Q plot for S&P500 index log returns in the period between 01/01/2000 and 31/12/2019.

Finally, when taking dispersion into consideration, if standard error (SE) of S&P 500 prices is far away from zero (8.2559), log return series' one, on the other hand, is quite close to zero (0.0002). You might notice SE can be approximated by the following formula:

$$SE = \frac{\sigma}{\sqrt{n}} \tag{2.3}$$

where: σ is standard deviation and n is the number of observations (sample size).

Equation (2.3) informs us that the larger the sample size (more data points) involved in the calculation, the smaller the SE tends to be. That is, if SE is small, the data are said to be more representative of the true mean. So, with the value of 8.2559, SE of S&P500 prices shows that data may have some notable irregularities as the sample is less accurate (due to high value

[2]https://www.itl.nist.gov/div898/handbook/eda/Chapter3/eda35b.htm#:~:text=Skewness%20is%20a%20measure%20of,relative%20to%20a%20normal%20distribution.

of SE). Obviously, with SE of 0.0002 (≈ 0), log return dataset can be expected to build more accurate models.

3. Methodology

3.1 Fundamental Concepts and Definitions

3.1.1 Autoregressive Moving Average Models - ARMA(p,q)

The ARMA process is the combination of autoregressive model and moving average (Box and Jenkins, 1976) designed for stationary time series. Autoregression (AR) describes a stochastic process and AR(p) can be denoted as below:

$$\text{AR(1)}: \quad y_t = \phi y_{t-1} + \varepsilon_t$$

$$\text{AR(p)}: \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t \tag{3.1}$$

where: ϕ_p denotes the weights given to past observations at each lag and p is a positive integer providing the number of lags to be included, and ε_t is white noise.

We now introduce the *lag* i.e. in the interest of notational convenience, which simply produces the previous element of the series, as below:

$$Ly_t = y_{t-1} \tag{3.2}$$

So, an AR (p), using lag notation, is now:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) y_t = c + \varepsilon_t, \quad p = 1,2, \ldots \tag{3.3}$$

The Moving Average process of order q is denoted as MA(q) and the created time series contains a mean of q lagged white noise variables shifting along the series.

$$\text{MA(1)}: \quad y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$\text{MA(q)}: \quad y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{3.4}$$

where: μ is the mean of the series, and θ_q are the weights given to each white noise value. Using lag notation, MA(q) can be written:

$$y_t = \mu + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t, \quad q = 1,2, \ldots \tag{3.5}$$

ARMA(p,q) is now expressed as below:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} + \ldots - \theta_q \varepsilon_{t-q} \tag{3.6}$$

where $\boldsymbol{\varepsilon_t}$ is independent of $y_{t-1}, y_{t-2}, \ldots$

3.1.2 Autoregressive Integrated Moving Average Models - ARIMA(p,d,q)

ARIMA (Box and Jenkins, 1976) model can be regarded as an extension of the ARMA model. This process can be written as:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d y_t = c + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \tag{3.7}$$

where:

- d is the number of differencing done to the series to achieve stationarity[3] with I (d):

$$(1-L)^d y_t = \mu + \varepsilon_t \tag{3.8}$$

- p is the number of autoregressive term (AR)
- q is the number of moving average term (MA)

3.1.3 The Autoregressive Conditional Heteroskedasticity – ARCH(q)

ARCH (Engle, 1982) can be expressed as:

$$y_t = C + \varepsilon_t,\ \varepsilon_t = z_t \sigma_t \tag{3.9}$$

where:

- y_t is an observed data series
- C is a constant value
- ε_t is residual
- z_t is the standardized residual, independently and identically distributed with mean is equal to 0 and variance tends toward 1 as sample size tends towards infinity
- σ_t is square root of the conditional variance, and it is a non-negative process.

ARCH(q) can be expressed in the following equation:

$${\sigma^2}_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i {r^2}_{t-i} \tag{3.10}$$

With $\alpha_0, \alpha_i \ \geq\ 0\ (i\ =\ 1,\ldots,q)\ so\ {\sigma^2}_t$ is non-negative.

3.1.4 Generalized Autoregressive Conditional Heteroscedasticity – GARCH(p,q)

GARCH (Bollerslev, 1986) model is considered to be an extension of an ARCH model. Unlike ARCH which involves only the most recent returns, Generalized ARCH (GARCH) enhances the accuracy of forecasting by adding all the past squared returns with higher weights on more recent data and lower ones for faraway lags. Furthermore, GARCH is more restrained in comparison to ARCH, hence it can avoid overfitting and permits an infinite number of past squared errors to impact the current conditional variance (Brooks, 2002). So now, the conditional variance σ^2 is expressed by GARCH(p,q) as:

$${\sigma^2}_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i {r^2}_{t-i} + \sum_{j=1}^{p} \beta_j {\sigma^2}_{t-j} \tag{3.11}$$

GARCH(1,1) can be expressed by the equation below:

$${\sigma^2}_t = \alpha_0 + \alpha_1 {r^2}_{t-1} + \beta_1 {\sigma^2}_{t-1} \text{ with } 0 < \alpha_1 + \beta_1 < 1 \tag{3.12}$$

[3] Shumway R. H. and Stoffer D. S., Time Series Analysis and Its Applications With R Examples, 2nd Edition.

and the rate of decay governed by $\alpha_1 + \beta_1$ where the closer $\alpha_1 + \beta_1$ is to 1, the slower the decay of the autocorrelation is. As being proved by Bollerslev et. al (1992), the valuations of GARCH(1,1) for stock returns usually yield $\alpha_1 + \beta_1$ very close to 1.

3.1.5 The Conditional Variance Equation: Exponential GARCH

The EGARCH (Nelson, 1991) model is defined as $a_t = e_t\sigma_t$, in which:

$$\ln({\sigma^2}_t) = \alpha_0 + g(e_{t-1}) + \beta_1 ln({\sigma^2}_{t-1} \tag{3.13}$$

the function $g(e_{t-1})$ determines the asymmetry and is defined as the weighted innovation:

$$g(e_{t-1}) = \alpha_1 e_{t-1} + \gamma_1 [|e_{t-1}| - E(|e_{t-1}|)] \tag{3.14}$$

where: α_1 and γ_1 are real constants. This means the model can be written:

$$\ln({\sigma^2}_t) = \alpha_0 + \alpha_1 e_{t-1} + \gamma_1 [|e_{t-1}| - E(|e_{t-1}|)] + \beta_1 ln({\sigma^2}_{t-1}) \tag{3.15}$$

The equation (3.15) informs us that a positive shock has the effect $(\alpha_1 + \gamma_1)e_{t-1}$ while the effect of the negative one has $(\alpha_1 - \gamma_1)e_{t-1}$. In reality, "good news" typically exerts a smaller impact on the volatility than "bad news". For this reason, the use of $g(e_{t-1})$ allows the model to respond asymmetrically to "new information" in the market.

The equation (3.16) expresses the general EGARCH(s,r) models.

$$ln({\sigma^2}_t) = \alpha_0 + \sum_{i=1}^{s} g_i\, e_{t-i} + \sum_{j=1}^{r} \beta_j ln({\sigma^2}_{t-j}) \tag{3.16}$$

3.1.6 The Hybrid ARIMA-GARCH

As discussed above, ARIMA models are proposed for stationary time series with the assumption of constant variance, defined as "homoskedasticity" while financial time series data often do not follow these assumptions. In practice, stock prices can be tremendously volatile during economic growth as well as recessions. In such scenarios, when homoskedasticity presumption is violated, it is said that the errors are heteroskedastic (a phenomenon known as heteroskedasticity). In other words, since heteroskedasticity is present, ARIMA or linear regression in general gives equal weights to all observations when observations with larger disturbance variance contain less information than the ones with smaller disturbance variance (Allison, 1999). Given that heteroskedasticity can affect the validity or power of statistical tests when using ARIMA models, the ARCH effect should be considered.

Furthermore, according to Mandelbrot (1963), large changes tend to be followed by large changes and vice versa. If volatility of a series exhibits such characteristic, it suggests that past variances might be predictive of the current variance. Hence, ARCH and GARCH models are the appropriate options in not only capturing the variance of each error term and correcting the deficiencies of heteroskedasticity for least squares but also dealing with the issue of volatility clustering. If one mechanism can simultaneously predict both the conditional mean and the conditional heteroscedasticity of the process, it is suggested as hybrid ARIMA-GARCH. It combines an ARIMA specification for modelling the mean behavior with the family of GARCH functions for simulating, estimating, and forecasting the variance behavior of the residuals from the ARIMA model. The hybrid ARIMA(p,d,q)-GARCH(r,s) can be specified as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} + \ldots - \theta_q \varepsilon_{t-q}$$

$$\sigma^2{}_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i r^2{}_{t-i} + \sum_{j=1}^{p} \beta_j \sigma^2{}_{t-j}$$

$$\varepsilon_t = z_t \sigma_t \ (\varepsilon_t \colon \mathrm{N}(0, \sigma^2{}_t)) \tag{3.17}$$

3.2 Overview the methodology and input parameters

In order to achieve goals which were mentioned in Introduction section, the methodology of this research is structured in the following way:

- Firstly, we conduct rolling forecast based on ARIMA model with window size(s) equal to 1000. The optimized combination of p and q which has the lowest AIC is used to predict return for the next point. At the end, the vector of forecasted values has the length of 3530 elements with starting point at 20/12/2005).
- Next, we describe and review our implementation of dynamic ARIMA(p,1,q)-SGARCH(1,1) models with GED distribution and window size(s) equal to 1000 and where optimized ARIMA(p,1,q) is taken from the 1st step. Then, we evaluate the results based on error metrics, performance metrics and equity curves.
- After that, in the sensitivity analysis section, we build hybrid models with different input parameters: window size(s) equal to 500 and then 1500, the following distributions: SNORM, SSTD, SGED.
- Finally, we replace SGARCH by EGARCH. We also conduct forecasting ARIMA on different window sizes in order to have a final conclusion whether hybrid model outperforms ARIMA in different input variables. And we use the same criteria as in the main test to compare and evaluate the performance of each model.

To sum up, the forecasting models are centered around five sets displayed in Table 2 below:

Table 2. Sets of input parameters (ARIMA/hybrid ARIMA-xGARCH)

Parameters	Values
Sample sizes	s ∈ {500, **1000**, 1500} (days)
Distribution	**Generalized Error Distribution (GED)** Skewed Normal Distribution (SNORM) Skewed Generalised Error Distribution (SGED) Skewed Student t Distribution (SSTD)
xGARCH MODEL	x ∈ {**SGARCH**, eGARCH} (x represents the type of tested GARCH model. In other words, x is either **symmetric** (s)GARCH or **e**xponential (e)GARCH.

Note: The letters in bold represent the parameters in the main test.

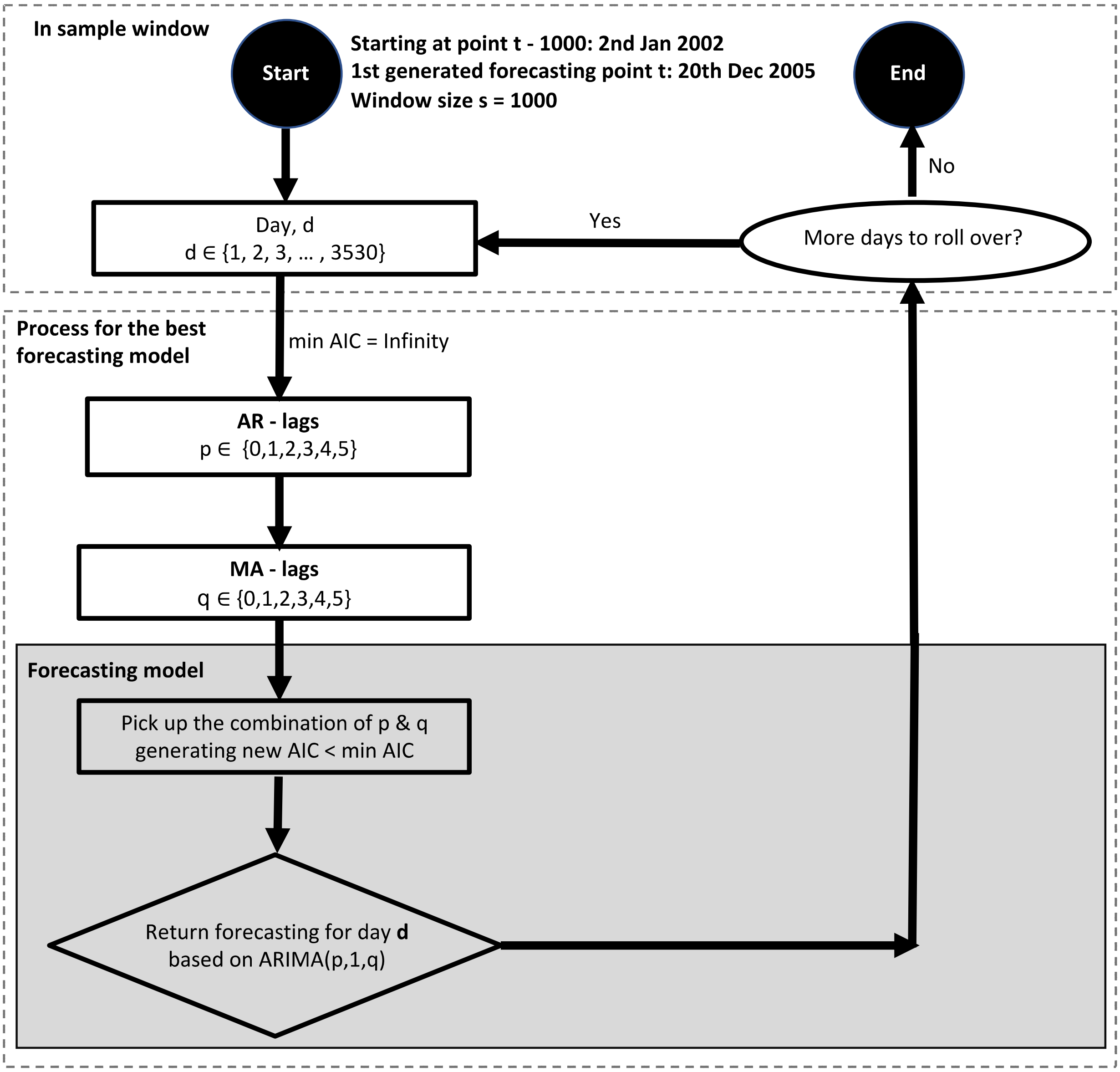
3.3 The implementation of forecasting models

3.3.1 ARIMA (p,1,q)

This section gives an in-depth outline of the actual implementation of ARIMA(p,1,q). As can be seen from the flowchart in Figure 4, this process spans fitting and forecasting from selecting sample size until the one-day-ahead return is obtained. Besides, the model assessment framework will be provided including return generating properties and an overview of the model's computational complexity.

Particularly, one optimal ARIMA(p,1,q) forecasting model is fitted using a rolling window approach with different combinations of p and q for the values of the input variables. This optimized model which has the lowest value of AIC is used to generate one-day-ahead return. Since rolling window approach is applied, the next data point is estimated based on the sample size equal to the length of window. The mechanism of this method is illustrated more specifically in Figure 5 with three iterations and sample size(s) equal to 1000. With the starting point at 20/12/2005, we have 3530 forecasted values for ARIMA model, based on which we construct equity lines for each strategy.

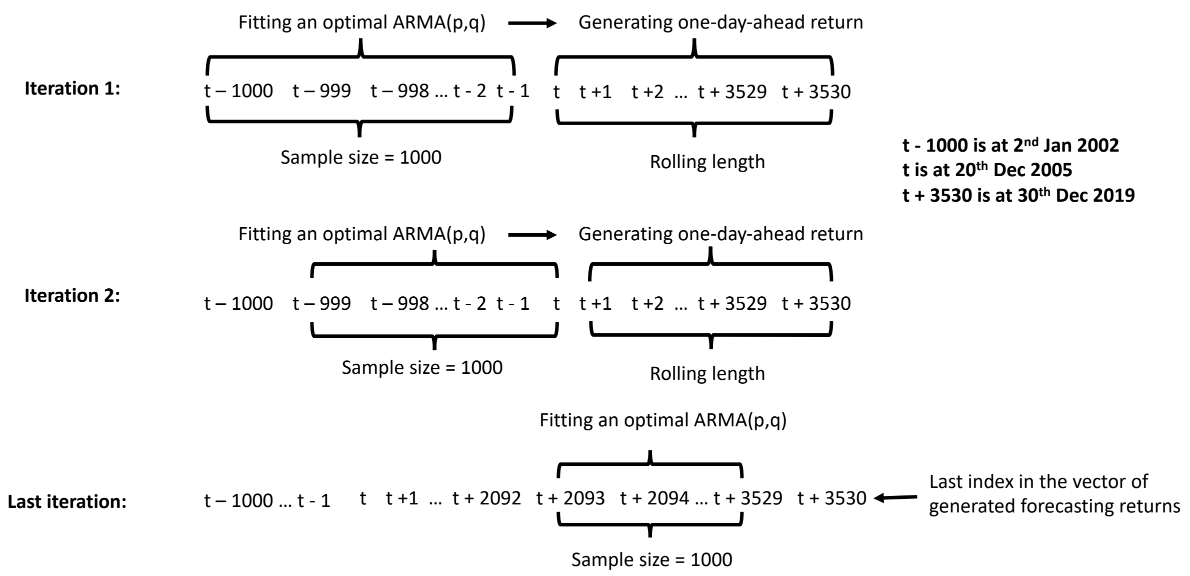
Figure 4. Flowchart of the forecasting model ARIMA(p,1,q)



Note: This flowchart is for models with window size s = 1000. For various values of window size such as 500 or 1500, the 1st generated forecasting point t is the same, hence all models have the same length of forecasted log return vector of 3530. The difference here is the starting point where models select 1st point to start a training model. For s = 500, starting point should be at t – 500 and similarly, for s = 1500, starting point should be at t – 1500.

Testing one combination of p and q is referred to as one iteration. An important condition we set up in this loop is that p and q cannot be equal to 0 at the same time. It means ARIMA(0,1,0) is excluded. The best fitting model is selected based on the lowest AIC. In each iteration, the inner loop compares 6 * 6 – 1 = 35 models together to pick up the best one with lowest value of AIC. To put it in another way, excluding the situation p = q = 0, with 6 values from 0 to 5, p and q generate 35 combinations. Hence with starting point t at 20/12/2005, each point in vector of 3530 elements is forecasted based on the most optimized ARIMA within these 35 models. For the whole process, the loop generates predicting returns vector by checking 3530*35 =123 550 models in total.

Figure 5. Rolling window illustration for sample size s = 1000

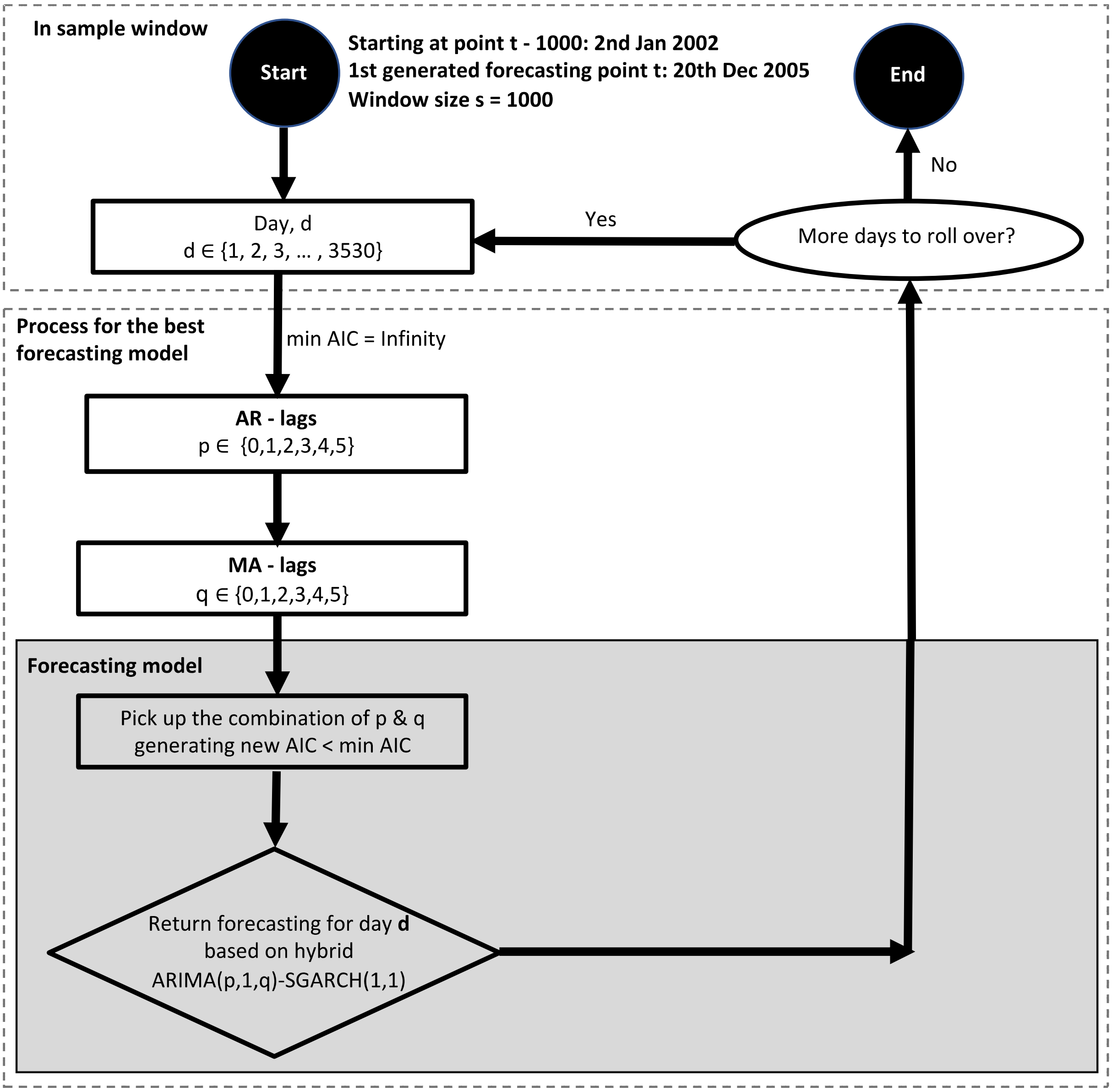


Note: this illustration is for models with window size s = 1000 and applied the same for both ARIMA and hybrid model in the next section. For various values of window size such as 500 or 1500, the process is the same except the sample size understanding as window length. For s = 500, starting point should be at t – 500 and similarly, for s = 1500, starting point should be at t – 1500.

3.3.2 Dynamical ARIMA(p,1,q)-SGARCH(1,1)

In this section, we describe and review the implementation of dynamic ARIMA(p,1,q)-SGARCH(1,1). The steps applied to select parameters are similar to those used to fit ARIMA models, and they are described as in Figure 6. It has similar steps as mentioned in Figure 4 regarding optimizing input parameters to fit model as well as rolling process in Figure 5. The return distribution named GED, SGARCH(1,1) and with window size s = 1000 are used in building hybrid models. We have ARIMA(p,1,q)-SGARCH(1,1) as an optimal outcome per iteration to forecast the next value of log return where SGARCH is applied to model the nonlinear patterns of the residuals. In other words, the error term ε_t of ARIMA model in this process follows SGARCH(1,1) instead of being assumed constant like ARIMA process in Figure 4.

Figure 6. Flowchart of the forecasting model ARIMA(p,1,q)-SGARCH(1,1)



Note: This flowchart is for models with window size s = 1000. For various values of window size such as 500 or 1500, the 1st generated forecasting point t is the same, hence all models have the same length of forecasted log return vector of 3530. The difference here is the starting point where models select 1st point to start training model. For s = 500, starting point should be at t – 500 and similarly, for s = 1500, starting point should be at t – 1500.

3.4 Trading strategy criteria

In general, the rule for going long (buy) or short (sell) is as follow: if forecasted log return is positive at time t + 1, we go long (buy stocks) at time t (direction would be +1); if forecasted log return is negative at time t + 1, we go short (sell stocks) at time t (direction would be -1); and if forecasted direction at time t+1 is the same as at time t then there are no changes.

The initial investment is assumed to be $1259.92 at the beginning. It is also the Adjusted Closing Price of S&P500 on 19/12/2005 (at t - 1) which is used as the starting point of equity curves. For the benchmark, we just use Buy&Hold strategy on S&P 500 index in the period 20/12/2005 – 31/12/2019 and compare our strategy's performance with this benchmark.

3.5 Criteria and Evaluation of Statistic Fit and Forecasting

3.5.1 Akaike Information Criterion (AIC)[4]

When a statistical model is selected to represent the process that generated the data, one might bear in mind that it will not be completely accurate. In other words, some "information" will be lost by applying this model in forecasting and it might lead astray if the missing information is of great importance and has a huge effect on adopted data. However, there is a trade-off between the goodness of fit (how well the model fits a set of observations) and the number of parameters (more parameters -> more information) in the model. In order to avoid the risks of overfitting and underfitting, we apply the Akaike information criterion – AIC. In general terms, AIC is an estimator of the relative quality of statistical models for a given dataset and also provides means for the model selection which is expressed by the following formula:

$$\text{AIC} = 2\text{k} - 2\ln(\hat{L}) \tag{3.18}$$

where: k is the number of estimated parameters in the model and $\hat{L}$ is the maximum value of the likelihood function for the model.

3.5.2 Error Metrics

In order to evaluate forecast form estimated models we calculate the following error metrics:

- **Mean Absolute Error (MAE)**

$$\textbf{MAE} = \frac{1}{n}\sum_{t=1}^{n}|A_i - F_i| \tag{3.19}$$

where: n is the number of errors; A_i is the actual value and F_i is the forecasted value computed by the given model.

- **Mean Square Error (MSE)**

$$\textbf{MSE} = \frac{1}{n}\sum_{t=1}^{n}(A_i - F_i)^2 \tag{3.20}$$

- **Root Mean Square Error (RMSE)**

$$\textbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2} \tag{3.21}$$

[4] https://www.statisticshowto.com/akaikes-information-criterion/

- **Mean Absolute Percentage Error (MAPE)**

$$\mathbf{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \tag{3.22}$$

3.5.3 Performance Statistics

Moreover, in order to evaluate the efficiency of algorithmic investment strategies built based on the signals from econometric models we calculated the performance metrics based on the created equity lines and formulas from Kość et al. (2019) and Zenkova and Ślepaczuk (2018).

- **Annualized Return Compounded (ARC)**

ARC is expressed as percentage (%) and computed as:

$$\mathrm{ARC} = \prod_{i=1}^{N}(1 + R_i)^{252/N} - 1 \tag{3.23}$$

where: R_i is the percentage rate of return; N is the sample size

- **Annualized Standard Deviation (ASD)**

ASD is expressed as percentage (%) and computed as:

$$\mathrm{ASD} = \sqrt{252} * \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(R_i - \bar{R})^2} \tag{3.24}$$

where: R_i is the percentage rate of return; $\bar{R}$ is the average rate of return; N is the sample size

- **Maximum Drawdown (MD)**

MD is the difference between the global maximum and the consecutive global minimum of the equity curve. The importance here is the time order which means the global maximum must occur before the global minimum. It is expressed as below:

$$\mathrm{MD}\ (\mathrm{S})_{t_1}^{t_2} = \max_{(x,y)\,\in\,\{[t_1,t_2]^2:\, x\,\le\, y\}} \frac{S_x - S_y}{S_x} \tag{3.25}$$

where: S is the price process; $[t_1, t_2]$ is the period between time t_1 *and* t_2

- **Information Ratio (IR)**

IR is the ratio between ARC and ASD informing us about risk adjusted returns for tested strategy

$$\mathrm{IR} = \frac{\mathrm{ARC}}{\mathrm{ASD}} \tag{3.26}$$

- **Adjusted Information Ratio (IR*)**

IR* is similar to IR but it also takes into account MD as one of the risk factors, then we have:

$$IR^* = \frac{ARC^2 * sign\{ARC\}}{ASD * MD} \tag{3.27}$$

where: sign {ARC} is the sign of ARC and can take values of 0, -1 or +1.

4. Empirical Results & Discussion

The performance of ARIMA (ARIMA 1000) and hybrid model ARIMA(p,1,q)-SGARCH(1,1) with GED distribution (SGARCH.GED 1000) as well as benchmark (Buy&Hold–S&P500) for window size(s) equal to 1000 are presented in Table 3. As the result shows, the hybrid model outperforms ARIMA and benchmark strategy evaluated based on error metrics and performance statistics. In particular, SGARCH.GED 1000 is more accurate than ARIMA 1000 in predicting returns and has the lowest values of MAE, MSE, RMSE and MAPE (11.831 against 12.122; 303.044 against 310.372; 17.408 against 17.617; 0.00754 against 0.00775, respectively).

Table 3. Forecasting performance of ARIMA(p,1,q) & ARIMA(p,1,q)-SGARCH(1,1)

	ERROR METRICS				PERFORMANCE STATISTICS				
METHOD	**MAE**	**MSE**	**RMSE**	**MAPE**	**ARC**	**ASD**	**MD**	**IR**	**IR***
BUY&HOLD S&P 500					6.931%	**18.826%**	56.775%	0.368	0.045
ARIMA 1000	12.122	310.372	17.617	0.00775	8.084%	18.878%	50.007%	0.428	0.069
SGARCH.GED 1000	**11.831**	**303.044**	**17.408**	**0.00754**	**14.026%**	18.893%	**25.885%**	**0.742**	**0.402**

Note: In order to simplify the structure of the table, SGARCH.GED 1000 is understood as ARIMA(p,1,q)-SGARCH(1,1) with GED distribution and window size equal to 1000 days. MAE: Mean Absolute Error; MSE: Mean Squared Error; RMSE: Root Mean Squared Error; MAPE: Mean Absolute Percentage Error; ARC: RC: Annualized Return Compounded; ASD: Annualized Standard Deviation; MD: Maximum Drawdown; IR = ARC/ASD: Information Ratio; IR = (ARC^2 * sign(ARC))/(ASD * MD): adjusted Information Ratio. Figures in bold indicate the best results.*

Concerning performance statistics, the hybrid model generates the highest Information Ratio (IR) among the 3 methods with 0.742, the second one is ARIMA 1000 with IR of 0.428 and the last one is the benchmark with IR of 0.368. Although hybrid model gives the highest Annualized Return Compounded (ARC) equal to 14.026%, its Annualized Standard Deviation (ASD) is also the highest with 18.893%. However, it is not significantly different from the lowest value of 18.826% belonging to the benchmark. In terms of adjusted Information Ratio (IR*), we can see that SGARCH.GED 1000 also outperforms ARIMA 1000. The difference between IR and IR* is that we additionally take into account Maximum Drawdown (MD) as a measure of risk beside ASD. We can see that MD of ARIMA (50.007%) is almost 2 times higher than the MD of the hybrid model (25.884%) while the ARC in the numerator of IR* (8.084%) is lower than the hybrid model's one nearly 1.75 times. As a result, IR* of ARIMA

1000 is approximately 6 times smaller than SGARCH.GED 1000's one (0.069 against 0.402). With this value of IR*, the hybrid method again beats the market with IR* of 0.045 (approximately 10 times higher).

Figure 7. Equity curves of ARIMA(p,1,q) & ARIMA(p,1,q)-SGARCH(1,1)

Note: In order to simplify the structure of the legend, SGARCH.GED 1000 is understood as ARIMA(p,1,q)-SGARCH(1,1) with GED distribution and window size equal to 1000 days; ARIMA 1000 is ARIMA(p,1,q) with window size s = 1000; BUY&HOLD-S&P500 is the benchmark strategy.

To visualize the performance of ARIMA(p,1,q) and the hybrid model with GED distribution as well as the benchmark, the cumulative returns of these strategies are shown in Figure 7. The equity curves of ARIMA and the hybrid model remain below the Buy&Hold strategy for almost 2 years, but during the financial market crisis of 2008-2009, they behave tremendously well. Especially, from latter half of 2010, ARIMA proofs as a good candidate which even outperforms the hybrid model impressively in almost 2 years (2010-2011). However, at the end of 2011, ARIMA performance witnesses a dramatical decline and then remains below hybrid model till the end of our data period. After financial market crisis of 2008-2009, the hybrid model undergoes an upward trend with small breaks to the end of 2019. In general, in spite of being under ARIMA in short periods of time from 2008-2011, it is depicted as the most superior model in the whole discussed data frame. At the onset it is clear it captures well all the movements of time series and is much better when compared with the benchmark.

In general, error metrics, performance statistics and equity curves imply that hybrid model outperforms ARIMA and the benchmark. Referring to the main hypothesis of this paper, we can conclude that the combination between ARIMA(p,1,q) and SGARCH(1,1) is efficient.

5. Robustness Test

In this Chapter, we verify if the result we obtain above is robust to varying family of GARCH, various distributions as well as different window lengths. In the previously obtained results, we conduct rolling forecasting on hybrid model ARIMA(p,1,q)-SGARCH with GED distribution and window size equal to 1000. In order to check the sensitivity of this result, we change input parameters to conduct three extra tests. In particular, the first robustness test is to substitute SGARCH to EGARCH (keep the same GED distribution and window size of 1000 days used in the main test). The second one is changing GED to variety distributions such as SNROM, SSTD, SGED (the other conditions of the main remain unchanged). The last one is replacing the window size of 1000 to 500 and 1500 (the remaining conditions of the main test are kept the same).

5.1 Varying family of GARCH models

Table 4 informs us that ARIMA(p,1,q) has the worst performance with the highest values of MAE, MSE, RMSE and MAPE in comparison to EGARCH.GED 1000 (12.122 against 11.828, 310.372 against 301.745, 17.617 against 17.371, 0.00775 against 0.00753, respectively). Moreover, with the highest values of all Key Performance Indicators (KPIs) where ARC = 11.010%, IR = 0.582 and IR* = 0.220, EGARCH.GED 1000 beats not only ARIMA with ARC = 8.084%, IR = 0.428 and IR* = 0.069 but also the benchmark with ARC = 6.931%, IR = 0.368 and IR* = 0.045.

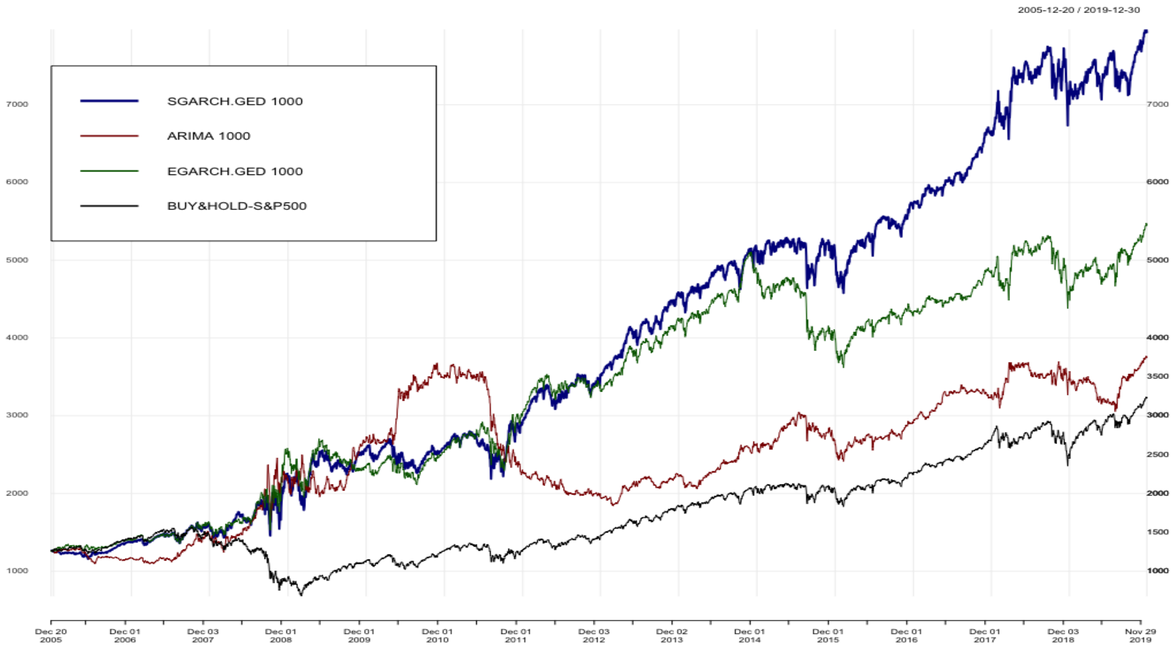
Table 4. Forecasting performance of ARIMA(p,1,q) & ARIMA(p,1,q)-EGARCH(1,1)

	ERROR METRICS				PERFORMANCE STATISTICS				
METHOD	**MAE**	**MSE**	**RMSE**	**MAPE**	**ARC**	**ASD**	**MD**	**IR**	**IR***
BUY & HOLD S&P 500					6.931%	**18.826%**	56.775%	0.368	0.045
ARIMA 1000	12.122	310.372	17.617	0.00775	8.084%	18.878%	50.007%	0.428	0.069
SGARCH.GED 1000	11.831	303.044	17.408	0.00754	**14.026%**	18.893%	**25.885%**	**0.742**	**0.402**
EGARCH.GED 1000	**11.828**	**301.745**	**17.371**	**0.00753**	11.010%	18.901%	29.150%	0.582	0.220

Note: In order to simplify the structure of the table, EGARCH.GED 1000 is understood as ARIMA(p,1,q)-EGARCH(1,1) with GED distribution and window size equal to 1000 days. MAE: Mean Absolute Error; MSE: Mean Squared Error; RMSE: Root Mean Squared Error; MAPE: Mean Absolute Percentage Error; ARC: RC: Annualized Return Compounded; ASD: Annualized Standard Deviation; MD: Maximum Drawdown; IR = ARC/ASD: Information Ratio; IR = (ARC^2 * sign(ARC))/(ASD * MD): adjusted Information Ratio. The figures in bold indicate the best results.*

The equity curves of all models and the benchmark are plotted in Figure 8. In general, in spite of being under ARIMA in some periods of time from 2008-2011, EGARCH.GED 1000 is depicted as a more superior model in the whole analyzed data period. It leads to the conclusion that the transformation from SGARCH to EGARCH seems to be insensitive. That is to say, we can conclude that the combination of ARIMA(p,1,q) and EGARCH(1,1) outperforms ARIMA in the similar way as the combination of ARIMA(p,1,q) and SGARCH(1,1), and we can treat it as the answer to the first research question of this paper.

Figure 8. Equity curves of ARIMA(p,1,q) & ARIMA(p,1,q)-EGARCH(1,1)



Note: In order to simplify the structure of the legend, SGARCH.GED 1000 is understood as ARIMA(p,1,q)-SGARCH(1,1) with GED distribution and window size equal to 1000 days; EGARCH.GED 1000 is understood as ARIMA(p,1,q)-EGARCH(1,1) with GED distribution and window size equal to 1000; ARIMA 1000 is ARIMA(p,1,q) with window size s = 1000; BUY&HOLD-S&P500 is the benchmark strategy.

However, as can be seen in Table 4, although error metrics of EGARCH.GED.1000 have the lowest values, it cannot beat the SGARCH.GED 1000 in terms of performance statistics. With the ARC of 14.026%, IR* of 0.402, the hybrid model with SGARCH is the most superior strategy in comparison with the other 3 methods in Table 4. Furthermore, Figure 8 shows that from the beginning of 2016, we observe a big difference in the cumulative returns of these two hybrid models. EGARCH is introduced as more advanced than SGARCH since it takes the magnitude of volatility into consideration. In other words, EGARCH mitigates the disadvantage of GARCH by putting more weights on negative innovation since it tends to increase the volatility. However, the result based on IR*, which is selected as the most important performance statistics to evaluate the model, does not support this theory. It leads to the observation that the best model is not necessarily the same when the selection is based on the best error metrics or the best performance statistics. In a nutshell, as a response to the second research question in this paper, based on IR* as the main indicator for selecting the best model, SGARCH.GED 1000 outperforms EGARCH.GED 1000.

5.2 Varying window sizes

Table 5 demonstrates the performance of ARIMA and SGARCH.GED with window sizes of 500 and then 1500. The result of the main test with window size = 1000 is also included. Error metrics for the window size of 500 show that SGARCH.GED 500 with the lower values of MAE, MSE, RMSE and MAPE (11.91 against 12.216, 307.812 against 318.342, 17.545 against 17.842, 0.00758 against 0.00777, respectively) outperforms ARIMA 500. Moreover, with the higher values of KPIs in performance statistics (ARC = 5.912%, IR = 0.313 and IR* = 0.052), SGARCH.GED 500 also beats ARIMA 500 (ARC = -0.574%, IR = -0.03 and almost zero for IR*).

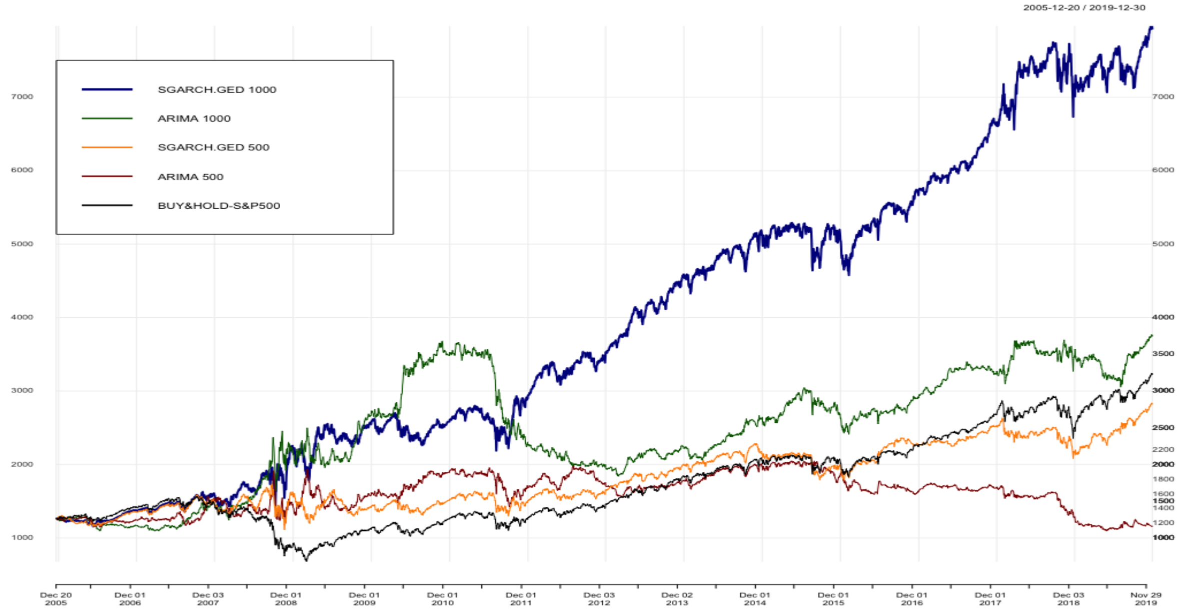
Table 5. Performance of ARIMA(p,1,q) & hybrid models in different window sizes

	ERROR METRICS				PERFORMANCE STATISTICS				
METHOD	**MAE**	**MSE**	**RMSE**	**MAPE**	**ARC**	**ASD**	**MD**	**IR**	**IR***
BUY & HOLD S&P 500					6.931%	**18.826%**	56.775%	0.368	0.045
ARIMA 500	12.216	318.342	17.842	0.00777	-0.573%	18.830%	46.471%	-0.030	0,000
SGARCH.GED 500	11.91	307.812	17.545	0.00758	5.912%	18.871%	35.666%	0.313	0.052
ARIMA 1000	12.122	310.372	17.617	0.00775	8.084%	18.878%	50.007%	0.428	0.069
SGARCH.GED 1000	11.831	**303.044**	**17.408**	**0.00753**	**14.026%**	18.893%	**25.885%**	**0.742**	**0.402**
ARIMA 1500	12.069	308.983	17.578	0.00771	5.005%	18.852%	50.733%	0.265	0.026
SGARCH.GED 1500	**11.825**	303.298	17.415	0.00753	12.186%	18.896%	25.885%	0.645	0.304

Note: In order to simplify the structure of the table, SGARCH.GED 500 is understood as ARIMA(p,1,q)-SGARCH(1,1) with GED distribution and window size s = 500 days, similar for s = 1000 and = 1500 days. MAE: Mean Absolute Error; MSE: Mean Squared Error; RMSE: Root Mean Squared Error; MAPE: Mean Absolute Percentage Error; ARC: RC: Annualized Return Compounded; ASD: Annualized Standard Deviation; MD: Maximum Drawdown; IR = ARC/ASD: Information Ratio; IR = (ARC^2 * sign(ARC))/(ASD * MD): adjusted Information Ratio. The figures in bold indicate the best results.*

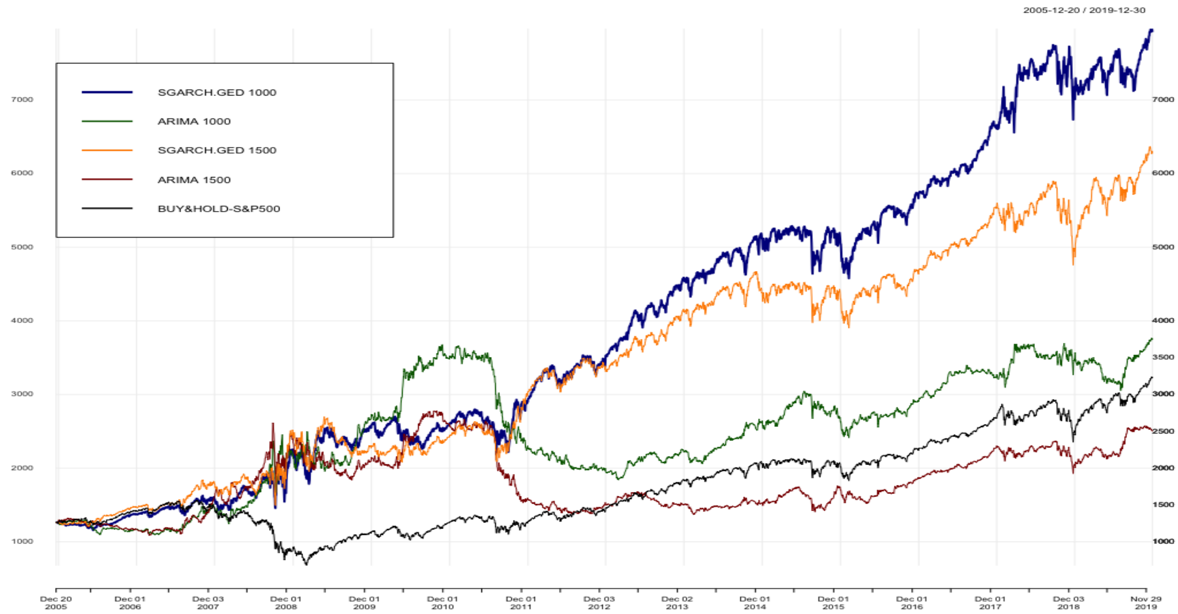
When window size is switched from 500 to 1500, we have the same results where the hybrid models demonstrate to be more superior than ARIMA. Particularly, SGARCH.GED 1500 has ARC of 12.186%, which is almost 2.5 times higher than that of ARIMA (5.005%), and ARIMA's IR* = 0.026 is nearly 10 times lower than the hybrid models' (0.304). The difference in the result of IR* is because of lower value of ARC as the numerator and higher value of MD as the denominator leading to lower final value of ARIMA's IR* in comparison with hybrid models'.

Figure 9. Equity curves of ARIMA(p,1,q) & hybrid models-window sizes=500 & 1000



Note: In order to simplify the structure of the legend, SGARCH.GED 500 / SGARCH.GED 1000 is understood as ARIMA(p,1,q)-SGARCH(1,1) with GED distribution and window size(s) equal to 500 /1000 days.

Figure 10. Equity curves of ARIMA(p,1,q) & hybrid models-window sizes=1000 & 1500

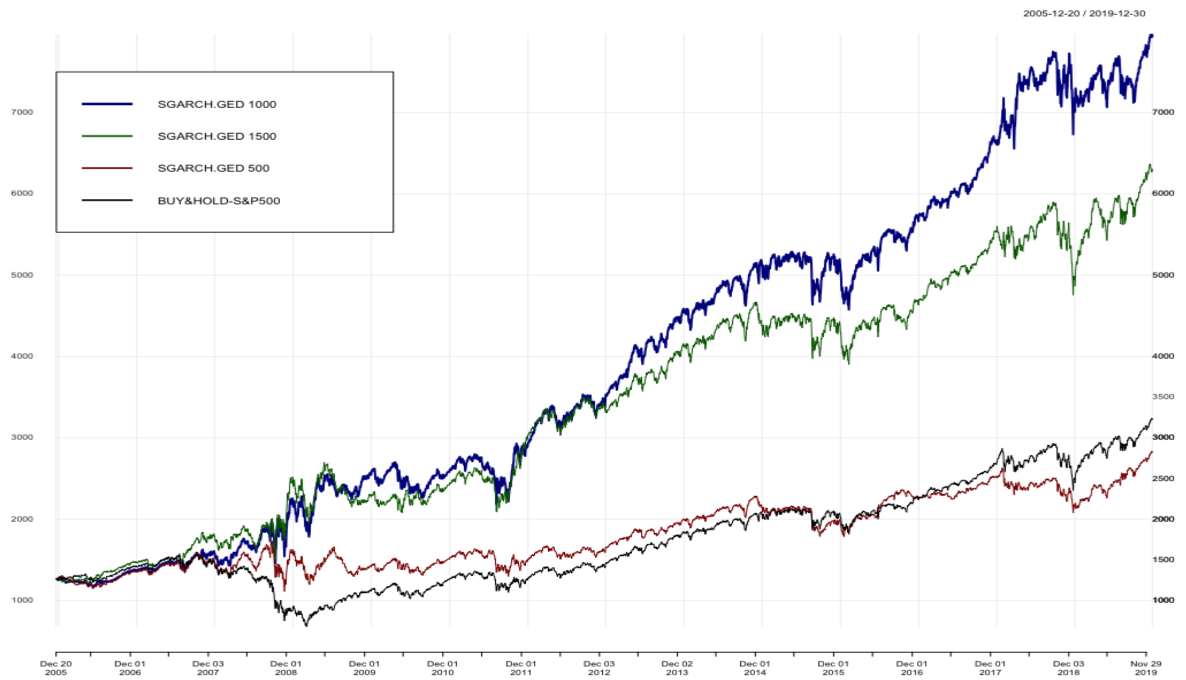


Note: In order to simplify the structure of the legend, SGARCH.GED 1000 / SGARCH.GED 1500 is understood as ARIMA(p,1,q)-SGARCH(1,1) with GED distribution and window size(s) equal to 1000/1500 days.

Figures 9 and 10 illustrate the equity curves of tested strategies and the benchmark with the window sizes of 500 and 1500, respectively. Noticeably, in Figure 9, both ARIMA 500 and SGARCH.GED 500 underperform benchmark at the end. We could see that Buy&Hold strategy is not necessarily the worst for various values. In general, our hybrid models seem to be sensitive to the values of window size. However, we can still conclude that hybrid models outperform ARIMA regardless of the values of window size as an input parameter. To recap, in response to the third research question in this paper, the results obtained from the main test are robust to varying window sizes.

Table 5 shows that with the highest value of ARC and IR* (14.026% and 0.402, respectively), the hybrid models with the window size of 1000 are the best strategy among these three different window sizes. According to Figure 11, SGARCH.GED 1000 beats all mentioned methods. Based on IR* as the main performance indicator in choosing the best model, SGARCH.GED 1000 outperforms the other hybrid models with different values of s = 500 and = 1500, and their differences are rather significant. In conclusion, the reply to the fourth research question in this paper is the hybrid models are sensitive to different window sizes.

Figure 11. Equity curves of ARIMA(p,1,q)-SGARCH(1,1) with different window sizes



Note: In order to simplify the structure of the legend, SGARCH.GED 500 / SGARCH.GED 1000 / SGARCH.GED 1500 is understood as ARIMA(p,1,q)-SGARCH(1,1) with GED distribution and window size(s) equal to 500/ 1000 / 1500 days.

5.3 Varying distributions

Table 6 presents the results of hybrid models with varying distributions including GED, SNORM, SSTD and SGED. In terms of error metrics, it is quite evident that ARIMA 1000 has the worst performance with the highest values of MAE, MSE, RMSE and MAPE in comparison with all hybrid models. Although ARIMA 1000's KPIs of performance statistics are higher than the benchmark's, these figures are still lower than all hybrid models'. Particularly, in terms of

IR*, SGARCH.SNORM 1000 with 0.129, SGARCH.SSTD 1000 with 0.147 and SGARCH.SGED 1000 with 0.119 beat ARIMA 1000 with 0.069.

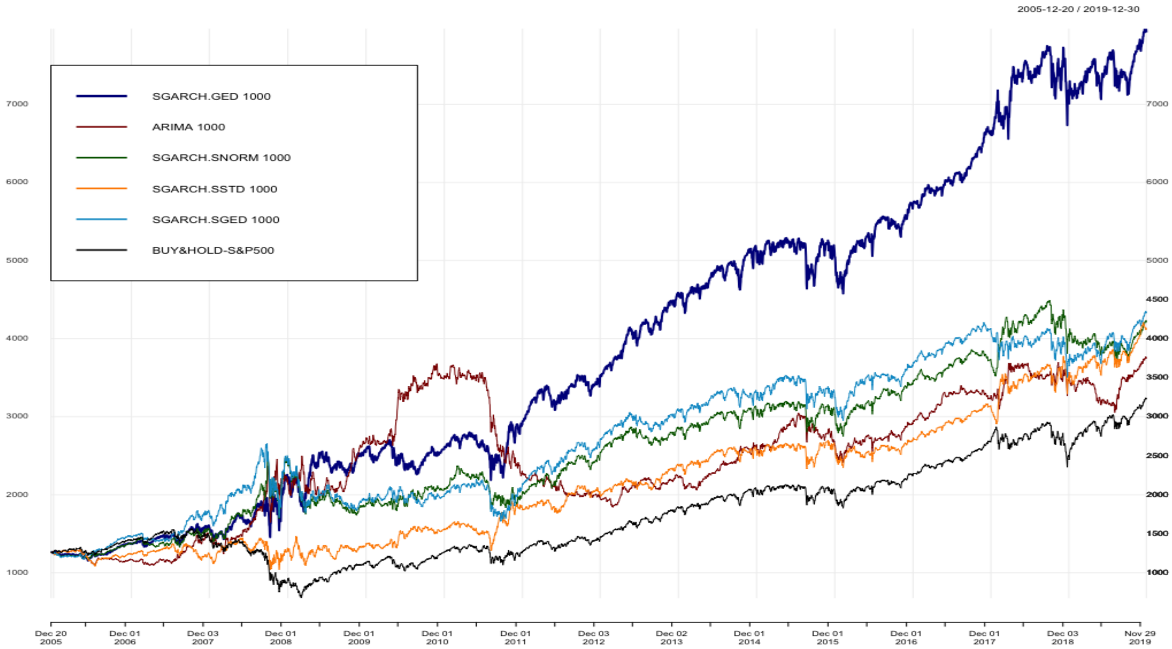
Table 6. Performance of ARIMA(p,1,q) & hybrid models in different distributions

	ERROR METRICS				PERFORMANCE STATISTICS				
METHOD	**MAE**	**MSE**	**RMSE**	**MAPE**	**ARC**	**ASD**	**MD**	**IR**	**IR***
BUY & HOLD S&P 500					6.931%	**18.826%**	56.775%	0.368	0.045
ARIMA 1000	12.122	310.372	17.617	0.00775	8.084%	18.879%	50.007%	0.428	0.069
SGARCH.GED 1000	**11.831**	**303.044**	**17.408**	**0.00754**	**14.026%**	18.893%	**25.885%**	**0.742**	**0.402**
SGARCH.SNORM 1000	11.880	303.151	17.411	0.00758	8.987%	18.890%	33.079%	0.476	0.129
SGARCH.SSTD 1000	11.928	305.642	17.483	0.00762	8.860%	18.881%	28.373%	0.469	0.147
SGARCH.SGED 1000	11.848	302.362	17.389	0.00755	9.201%	18.859%	37.566%	0.488	0.119

Note: In order to simplify the structure of the table, SGARCH.GED 1000 is understood as ARIMA(p,1,q)-SGARCH(1,1) with GED distribution and window size s = 1000 days, similar for SNORM, SGED and SSTD. MAE: Mean Absolute Error; MSE: Mean Squared Error; RMSE: Root Mean Squared Error; MAPE: Mean Absolute Percentage Error; ARC: RC: Annualized Return Compounded; ASD: Annualized Standard Deviation; MD: Maximum Drawdown; IR = ARC/ASD: Information Ratio; IR = (ARC^2 * sign(ARC))/(ASD * MD): adjusted Information Ratio. Figures in bold indicate the best results.*

Figure 12 plots the equity curves of all hybrid models with various distributions as well as ARIMA while window size remains 1000. It can be seen that the cumulative returns of hybrid models with SNORM, SSTD and SGED distributions show no significant differences at the end, but all of them surpass ARIMA's. The results in Figure 12 and Table 6 draw the conclusion that regardless of the distributions, our hybrid models are more profitable than ARIMA. In short, the fifth research question can be answered as follows: the results obtained from the main test are robust to varying distributions. As can be seen in Figure 12, it is obvious that the best model is SGARCH.GED 1000, whose ending point is significantly far away from the rest and performance metrics are much better. As for the last research question of this paper, hybrid models are sensitive to different distribution and this with GED distribution outperforms the ones with other distributions such as SNORM, SSTD and SGED.

Figure 12. Equity curves of all hybrid models with different distributions



Note: In order to simplify the structure of the legend, SGARCH.GED 1000 is understood as ARIMA(p,1,q)-SGARCH(1,1) with GED distribution and window size(s) equal to 1000 days, similar for SNORM, SSTD and SGED.

6. Conclusion

The main hypothesis of this paper is whether the ARIMA(p,1,q)-SGARCH(1,1) (hybrid model) with window size equal to 1000 can generate an algorithmic trading strategy that outperforms ARIMA(p,1,q). Based on this hypothesis, the research questions are constructed as follows: whether the results in main test are robust to (RQ1) varying family of GARCH model; (RQ3) varying window sizes; and (RQ5) varying distributions. Besides, we also evaluate and examine more research questions namely whether the performance of hybrid models in the main test changes with (RQ2) varying family of GARCH model; (RQ4) varying window sizes; and (RQ6) varying distributions.

The dataset used for this research consists of the quotations of S&P500 index. The data are collected on the daily basis over the period from 01/01/2000 to 31/12/2019. Next forecasted value is generated based on the best ARIMA model as a result of the best combination of p from 0 to 5 and q from 0 to 5, which has the lowest value of AIC. Rolling window of 1000 with one day ahead moving is selected for the main test. The vector of forecasted log returns with 3530 elements is generated. Based on these values, we set up the trading signals in which we enter the long position if forecasted log return is positive and enter the short one if forecasted log return is negative. By assuming the initial investment of $1259.92 (the level of S&P500

index at the starting date), we calculate returns from the starting point of our out-of-sample window on 19/12/2005. Similar steps and process are conducted for ARIMA(p,1,q)-SGARCH(1,1) with GED distribution and window size s = 1000. The difference is by combining ARIMA with SGARCH to create hybrid ARIMA-GARCH model, conditional mean and variance can be simultaneously modeled (unlike ARIMA(p,1,q) where only conditional mean is modeled). We then calculate error metrics and performance statistics. Our benchmark is simply Buy&Hold strategy on S&P500 index. For the robustness test analysis, we conduct the same procedure step by step like we do for the main test with changing input parameters such as replacing GARCH with EGARCH, varying window sizes (500 & 1500) as well as distributions (SNORM, SSTD, SGED). In order to evaluate the performance of these models, we compute error metrics (MAE, MSE, MAPE, RMSE), performance statistics (ARC, ASD, MD, IR, IR*), and present the equity curve for each model and also the benchmark.

Overall, the result shows that hybrid methods can generate a strategy that can outperform ARIMA model even when we change our initial assumptions concerning the family of GARCH (RQ1), window sizes (RQ3) as well as distributions (RQ5). However, they are not always more efficient than our benchmark like it was in the case of hybrid model of SGARCH with GED distribution and window size of s = 500. Besides, the hybrid model of SGARCH with GED distribution and window size equal to 1000 performs the best model in comparison to other hybrid models in terms of changing different GARCH model (RQ2), window sizes (RQ4) and distributions (RQ6). Especially, even being introduced as a model which can mitigate the disadvantages of SGARCH, the EGARCH model cannot beat SGARCH in this research. To conclude, from the obtained results, hybrid ARIMA-GARCH can generate a trading strategy that outperforms ARIMA(p1,q) and should be taken into consideration in predicting returns and building trading strategies instead of applying only ARIMA into the series.

There are some limitations of this paper which can be improved in future works. The first and the biggest is that proposed ugarchroll function in rugarch package (Author: Alexios Ghalanosonly) supports moving one period ahead only. It is time-consuming and reducing efficiency in practice. This issue may be addressed if we incorporate extra wrapped function into the main one by making use of the underlying functions in the package. In addition, we apply fixed window sizes as 500, 1000 and 1500 without checking which exact value of the window size will deliver the best results. To solve this issue, we can try to build a loop function with the input parameter that can be a range of window sizes (for example the range of 500, 501, 503, …, 1499, 1500). In this way, the best value of window size which delivers the best trading strategy can be selected. Moreover, we apply trading strategy based on ideal conditions without any transaction costs or a certain threshold. It means we should take into consideration the magnitude of forecast return value instead of building strategy based on the sign of forecasted values. This issue may be addressed by assigning cost for any transaction and setting a threshold with which we can compare forecasted returns before generating the direction for trading (+1 for entering long or -1 for entering short).

References

Abbasi U., Almarashi A. M., Khan K. and Hanif S., 2017, Sci. Int. (Lahore), 29(5), 1039-1043.

Allison, P. D., 1999. Multiple regression: A primer. Pine Forge Press.

Amin A., Colman A., & Grunske L., 2012, An Approach to Forecasting QoS Attributes of Web Services Based on ARIMA and GARCH Models, IEEE 19th International Conference on Web Services (ICWS).

Ariyo A., Adewumi A. & Ayo C., 2014 UK Sim-AMSS 16th International Conference on Computer Modelling and Simulation, Computer Science, Economics, 105-111.

Azlan A., Yusof Y., Mohamad M., Mohamad F., 2019, Determining the impact of window length on time series forecasting using deep learning. International Journal of Advanced Computer Research, 9 (44), 260-267.

Bollerslev T., 1986, Generalized autoregressive conditional heteroscedasticity, Journal of Econometics 31, 307-327.

Bollerslev T., R. Chou, and K. Kroner, 1992, ARCH modeling in finance, Journal of Econometrics, 52, 5–59.

Bollerslev, T., 1987, A conditionally heteroskedastic time series model for speculative prices and rates of return, The review of economics and statistics, 542-547.

Box G. E. P. et al, Time Series Analysis: Forecasting & Control, 5th, JW: 372-375.

Box G.E.P. and Jenkins G.M., 1976, Time Series Analysis Forecasting and Control, Holden Day, San Francisco.

Brooks C., 2002, Introductory econometrics for finance, Cambridge: Cambridge Uni. Press.

Campbell J. Y., Lo A. W. C., MacKinlay A. C., 1997. The econometrics of financial markets (Vol. 2, pp. 149- 180). Princeton, NJ: princeton University press.

Chen C., Hu J., Meng Q., & Zhang Y., 2011, Short-time traffic flow prediction with ARIMA-GARCH model, IEEE Intelligent Vehicles Symposium (IV).

Chua C. L., Tsiaplias S., 2019, Information flows and stock market volatility. Journal of Econometrics 34, 129-148.

Cryer J. D. & Chan K. S., 2008, Time Series Analysis With Applications in R, Second Edition, 280.

Engle R. F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50, 987-1007.

Engle, R. F, Ng, V.K., Rothschild, 1990, Asset pricing with a factor-Arch covariance structure: Empirical estimates for treasure bills, Journal of Econometrics, 45(1-2), 213-237.

Ghani I. M. Md. and Rahim H. A., 2019, Modeling and Forecasting of Volatility using ARIMA-GARCH: Case Study on Malaysia Natural Rubber Prices. IOP Conf. Ser.: Mater. Sci. Eng. 548 012023.

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.

Hauser M., Kunst R., 1998, Fractionally Integrated Mode is with ARCH errors: With an Application to the Swiss 1-Month Euro market Interest Rate. Review of Quantitative Finance and Accounting 10, 95–113.

Higgins M. L., Bera A. K., 1992, A class of nonlinear ARCH models, International Economic Review, 137-158.

Horvath Z. and Johnston R., 2008, GARCH Time Series Process, Econometrics 7590.

Jabłecki J., Kokoszczyński R., Sakowski P., Ślepaczuk R., Wójcik P., Volatility as an Asset Class. Obvious Benefits and Hidden Risks, Peter Lang, Frankfurt am Main, 2015. ISBN: 978-3-653-97884-1, DOI: *10.3726/978-3-653-04787-5*

Jansen, S., 2020, Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with Python, Packt Publishing Ltd.

Kamruzzaman, Khudri M. & Rahman M., 2017, Modeling and Predicting Stock Market Returns: A Case Study on Dhaka Stock Exchange of Bangladesh, Dhaka University Journal of Science 65(2), 97-101

Kijewski M., Ślepaczuk R., 2020, Predicting prices of S&P500 index using classical methods and recurrent neural networks, Working Papers of Faculty of Economic Sciences, University of Warsaw, WP 27/2020 (333), https://www.wne.uw.edu.pl/index.php/download_file/5769/4919/

Kość K., Sakowski P., Ślepaczuk R., 2019, Momentum and Contrarian Effects on the Cryptocurrency Market, Physica A 523, pp. 691-701, *https://www.sciencedirect.com/science/article/pii/S037843711930216X?dgcid=author*

Mandelbrot B. B., 1963, the variation of certain speculative prices, Journal of Business, XXXVI (1963), 392–417.

Mandelbrot B., 1963, The variation of certain speculative prices, The Journal of Business, Vol. 36, No. 4, 394-419.

Matías J. M., Reboredo J. C., 2012. Forecasting performance of nonlinear models for intraday stock returns. Journal of Forecasting, 31(2), 172-188.

Mohammadi, H., & Su L., 2010, International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models, Energy Economics, 32(5), 1001-1008.

Mustapa F. H. and Ismail M. T., 2019, Modelling and forecasting S&P 500 stock prices using hybrid ARIMA-GARCH Model, Journal of Physics: Conference Series, 1366 (2019) 012130.

Nelson D. B, 1991, Conditional heteroskedasticity in asset returns: A new approach, Econometrica: Journal of the Econometric Society, 347-370.

Sentana, E., 1995, Quadratic ARCH Models, The Review of Economic Studies, 62(4), 639-661. doi:10.2307/2298081.

Strickland J., 2014, Predictive Analytics Using R, ISBN 978-1-312-84101-7, Published by Lulu, Inc.

Sun K., 2017, Equity Return Modeling and Prediction Using Hybrid ARIMA-GARCH Model, International Journal of Financial Research Vol. 8(3), 154-161.

Tsay, R. S., 2014, An introduction to analysis of financial data with R. John Wiley & Sons.

Weiss Andrew A., 1984, ARIMA Models with ARCH errors, Journal of Time Series Analysis, Wiley Blackwell, vol. 5(2), pages 129-143.

Wilhelmsson A., 2006, Garch forecasting performance under different distribution assumptions. Journal of Forecasting, 25, 561-578.

Yaziz, S. R., Azizan, N. A., Zakaria, R., & Ahmad, M. H., 2013, The performance of hybrid ARIMA-GARCH modeling in forecasting gold price, In 20th international congress on modelling and simulation, adelaide, 1-6.

Zenkova M., Ślepaczuk R., 2018, Robustness of Support Vector Machines in Algorithmic Trading on Cryptocurrency Market, Central European Economic Journal, 5(1), pp. 186-205, *https://content.sciendo.com/view/journals/ceej/5/52/article-p186.xml*,

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl