University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 10/2020 (316)

# Investing in VIX futures based on rolling GARCH models forecasts

Oleh Bilyk
Paweł Sakowski
Robert Ślepaczuk

Warsaw 2020

WORKING PAPERS 10/2020 (316)

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES

WORKING PAPERS

# Investing in VIX futures based on rolling GARCH models forecasts

**Oleh Bilyk, Paweł Sakowski[a], Robert Ślepaczuk[a]***

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

**Abstract:** The aim of this work is to compare the performance of VIX futures trading strategies built across different GARCH model volatility forecasting techniques. Long and short signals for VIX futures are produced by comparing one-day ahead volatility forecasts with current historical volatility. We found out that using the daily data over the seven-year period (2013-2019), strategy based on the fGARCH-TGARCH and GJR-GARCH specifications outperformed those of the GARCH and EGARCH models, and performed slightly below the "buy-and-hold" S&P 500 strategy. For the base GARCH(1,1) model, the training window size and the type gave stable results, whereas the performance across refit frequency, conditional distribution of returns, and historical volatility estimators varies significantly. Despite non-robustness of some investment strategies and some space for improvements, the presented strategies show their potential in competing with the equity and volatility benchmarks.

**Keywords**: GARCH, VIX index, volatility futures, rolling forecasting, volatility, investment strategies, volatility exposure

**JEL codes**: C4, C45, C61, C15, G14, G17

**Note**: The views presented in this text are those of the authors and do not necessarily represent those of Labyrinth HF project

## 1. Introduction

Volatility could be broadly defined as the degree of a time series variation over the specified time period. It could be considered as the new asset class (Jabłecki et al. (2015)) among both new asset types like cryptocurrency (Kość et al., 2019, Zenkova and Ślepaczuk) and classical ones like stocks, bonds, commodities or currencies (Ślepaczuk et al., 2018), given that nowadays market participants are able to have the direct exposure to volatility trading via the volatility-based liquid products. Despite being new and relatively more complex in comparison to the products related to traditional asset classes, the number of volatility products has been steadily growing and they have a great potential in risk management, asset allocation, and trading strategies development areas.

The goal of this paper is to compare the performance of VIX futures trading strategies which were constructed on the basis of different volatility forecasting techniques. To produce those forecasts, various GARCH model specifications have been used. The main hypothesis is as follows: *Using approaches based on rolling GARCH models we are not able to obtain robust abnormal returns compared to the benchmark.* For the purpose of the hypothesis verification the following research questions are formulated:

1. Which rolling GARCH specification produces the most accurate volatility forecast?
2. Does more frequent model refitting improve portfolio Information Ratio?
3. How does the size of training window affect the strategy performance?
4. Is the base model strategy performance stable with regards to different historical volatility estimators?

We use the daily data of S&P 500 index from 2009-01-01 to 2019-10-03 and VIX index futures from 2013-01-02 to 2019-10-03, respectively. The longer period for S&P 500 index was required for GARCH model estimation in order to produce the first volatility forecasts. Four different GARCH model specifications (classical GARCH, EGARCH, GJR-GARCH, fGARCH-TGARCH) are implemented on the basis of the rolling forecast techniques: moving and expanding training windows.

We believe that the value added of this work is the examination of the forecasting accuracy of GARCH models and the development of the empirical methodology of investing in VIX futures. Based on the evidence from the literature, we expect that GARCH models that account for volatility stylized facts with more frequent refitting will perform better.

The remainder of the paper is organized as follows. In Section 1 we provide the definition of volatility, its measurement, and the literature review of GARCH models. Section 2 is devoted to the description of the rolling approaches undertaken, along with the overview of the selected GARCH specifications and measures of forecasting errors. Section 3 contains the description of the data, research procedure and assumptions behinds investment strategies. The most important R functions used in the rolling forecasts are also described here. In Section 4 the outcomes for the base GARCH(1,1) model and the performance analysis of 432 combinations of strategies parameters and GARCH model specifications is provided. Section 5 discusses the outcomes of the sensitivity analysis. The paper ends with conclusions, and the propositions for further research.

## 2. Theoretical background and literature review

### *2.1. Volatility and its "stylized facts"*

Volatility lies at the heart of derivatives pricing making it one of the most important concepts in the whole of modern finance. Not only has it been the subject of many academic research papers, but it is also widely applied concept in risk management, hedging, and portfolio optimization areas.

The task of volatility estimation and forecasting is not trivial as volatility is not directly observable. There are many estimators developed therefore, but probably the most popular one is standard deviation of returns:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{t=1}^{N}(R_t - R_{mean})^2} \tag{1}$$

where:

$\sigma$ – standard deviation,

$R_t$ – asset return at moment *t*,

$R_{mean}$ – the arithmetic average of returns,

$N$ – the number of observations in the sample.

We will refer to this approach as to the "Close-to-Close" estimator, as it involves asset returns measured with respect to its daily closing prices. It is relatively simple and intuitively understandable. It must be pointed out though that there exists an indirect link between risk of downturn moves and volatility measured as in formula (1) because it penalizes all large moves and does not make distinctions between positive and negative ones. Among dispersion-based metrics, there are semi-variance proposed by Markowitz (1952) and inter-quantile range which are useful but difficult in implementation in portfolio construction.

Alternatively, there are realized range estimators which use a wide range of information like "Open", "Close", "High", "Low" prices to improve the estimator quality. Following the comparison of pros and cons of various volatility estimators described in Ślepaczuk and Zakrzewski (2013) we additionally decided to use estimators proposed by Parkinson (1980) and Garman and Klass (1980).

In the approach proposed by Parkinson (1980), the highest and the lowest prices are incorporated, and the volatility ($\sigma_P^2$) is then given by:

$$\sigma_P^2 = \frac{1}{4\ln(2)}\sum_{i=1}^{N}(ln\frac{h_i}{l_i})^2 \tag{2}$$

where:

$h_i$ and $l_i$ – highest and the lowest prices within the $i$-th interval,

$N$ – number of observations.

Garman and Klass (1980) adjust the (2) by adding the close price ($c_i$) in their estimator:

$$\sigma_{GK}^2 = \frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{h_i}{l_i}\right)^2 - (2ln2 - 1)\frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{c_i}{c_{i-1}}\right)^2 \tag{3}$$

Authors argue that their estimator ($\sigma_{GK}^2$) is seven times more efficient in comparison to the "Close-to-Close", when measured by variance ratio. However, there might be not enough time for the 'true' unobservable prices to reach their minimum and maximum, and statistics might be biased, therefore.

Despite being not observable, volatility possesses some common characteristics which have been identified empirically across different asset classes. They are known as "stylized facts" (see Jabłecki et al. 2012 and Masset 2011).

- Volatility clustering (grouping)

Volatility is not constant, and it shows a tendency to form clusters likewise in an earthquake: once abnormal movement takes place it usually generates effects that persist through time.

- Leverage effects (asymmetry)

Volatility exhibits asymmetry to news arriving to the market which is primarily related to stocks and equity indexes. It could be explained in the following way: when price goes up, volatility decreases but not as much as it increases when price goes down by the same amount. It happens because when price goes down, the debt increases in capital structure and the stock becomes more riskier and volatility increases.

- Long memory

There are number of studies (e.g. Ding et a., 1993, Baillie, 1996) which found out that changes in volatility, especially significant ones, have a long-standing effect on the way volatility evolves through time.

- Mean reverting

In the long-time horizon volatility is expected to revert to its historical mean level even when being currently diverged far away from it. In statistical terms volatility varies within some fixed range of values.

Some authors say that volatility forecasting is a challenging task. Nevertheless, volatility is forecastable and can be used for investment purposes as was shown in Jabłecki et al. (2015). The question is how far ahead it could be accurately forecasted and to what extent its changes could be predicted.

*2.2. Literature review of GARCH models*

The primary motivations of using Autoregressive Conditionally Heteroscedastic (ARCH) models) is that they account for heteroscedasticity of returns. There is a strong evidence of successful ARCH volatility fit on US, UK, Swedish stock markets according to Engle and Mustafa (1992), Loudon et al. (2000), and Frennberg and Hanson (1996), respectively. There are also examples of poor ARCH performance. Balaban and Bayar A. (2002) found out ARCH being the worst out-of-sample performer across exponential smoothing and other GARCH approaches for fourteen stock markets study. Generalized ARCH (GARCH) was proposed by Bollerslev (1986) to overcome some ARCH limitations like constraint restrictiveness and slow responsiveness to shocks. GARCH dominance over ARCH is pointed out by Akgiray (1989), or West and Cho (1995) for US stock index volatility and dollar exchange rate volatility, respectively.

Poon and Granger (2003) conduct an extensive review of 93 studies related to volatility forecasting and found out that option implied standard deviation (ISD) models performed relatively better primarily due to the usage of larger, and more relevant information datasets, whereas there is no clear favorite between historical volatility and GARCH models. However, ISD approach is less used in practice given that implied volatility is not available for all instruments. Nelson (1991) proposed EGARCH and GJR-GARCH models to address the leverage effect in volatility. From the theoretical standpoint EGARCH and GJR-GARCH models are more relevant than classical GARCH model. Poon and Granger (2003) summarize that GJR-GARCH, EGARCH models produce more accurate forecasts on the out-of-sample dataset. Ebeid et.al (2004) made same conclusion for Egyptian stock index (2004). Majmudar and Banerjee (2004) found out also that EGARCH outperforms other specifications and suggest respective options trading strategy for yield enhancement based on volatility forecasting using VIX data. In contrary, Ederington and Guan (1999) claim that there is no significant difference in accuracy forecasting between GARCH and EGARCH and more complex models exhibit over-fitting: predict well on training dataset, but – poorly on the test dataset.

There are some papers that shed light on choice of appropriate GARCH model order, training-test data split, distribution choice, and market volume impact in forecasting. GARCH (1,1) is generally enough to capture heteroscedasticity and models with other lags combinations are rarely being tested in academic literature, according to Brooks (1998). However, Casas and Cepeda (2008) conducted such research and found out that returns series on stock market indexes are best explained by EGARCH (2,1). Engle and Patton (2001) obtain results that forecasting quality depends on in-the-sample dataset frequency. Basic training-test split is not efficient, and better techniques exists. For example, Clark and McCracken (2004) prove that a combination of recursive and rolling schemes with a fixed window size overall improves the forecasting accuracy. One of a few papers that study the impact of distribution chosen in GARCH model is conducted by Ebeid et.al (2004). They found out that skewed t-Student is the most appropriate distribution for volatility modelling of Egyptian stock market index. For European stock indexes there is an empirical evidence that non-parametric GARCH models are not beating the standard GARCH models, see Franses and Van Dijk (1996).

One of the fundamental shortcomings of ARCH and GARCH models proposed by Engle (1982) and Bollerslev (1986), respectively, is that these models poorly fit the situations when volatility increases abruptly (“volatility jumps”): conditional variance in GARCH (ARCH) process is slow in reaching the new level and it takes some time to “catch it up” given returns are the only ones used. The induction of realized measures tackles this problem and results in obtaining more statistical and economic gains according to Christoffersen et. al. (2014). The concept of realized GARCH models lies in realized volatility and returns joint modelling basing on measurement equations which tie realized measure to latent conditional variance. Zakoian (1994) also addresses the volatility leverage effect and given its flexibility allows the conditional variance to fall into two different regimes: it could be stationary and non-stationary. Results show that these models improve forecast on stocks and SPY exchange-traded index in comparison to standard GARCH and EGARCH models, moreover, the simpler the more accurate – such conclusion is obtained across realized GARCH models, according to Xie and Yu (2019) and Hansen and Huang (2016). For some asset classes there is no evidence though that classical GARCH model is *consistently* outperformed by these ‘richer’ models that characterize better volatility stylized facts (see Hansen and Lunde (2005) that claim that nothing is beating GARCH (1,1) for exchange rate data, for example)

To sum up, on the basis of the empirical evidence, GARCH dominates ARCH. There is no such clear dominance in terms of performance across different GARCH specifications as well as across distribution chosen, market volume inclusion, or lags order. However, the models that account for volatility stylized facts are more accurate than the ones that do not. Generally, rolling techniques produce better forecasts in comparison to the classical training/validation/test split. However, there are not many respective research papers for GARCH models. Therefore, we fill this gap by implementing different GARCH specifications on a rolling forecast basis.

## 3. Methodology description

### *3.1. GARCH model selection procedure*

The family of Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models tackles the issue of the irregular pattern of heteroscedasticity of returns in the real financial data. They are designed for capturing the autocorrelation patterns in returns and squared returns in dynamics, in particular – distributional parameters of the mean and variance given the variance depends on time.

The common approach in GARCH model selection for the purposes of forecasting is as follows. The very first thing is to identify the autocorrelation patterns, possible ARCH effects, and possible anomalies in asset’s returns by analyzing their respective ACF functions and ACF of squared returns along with returns’ distributional properties (normality, skewness, and kurtosis). The next step is to choose the proper order of $GARCH\ (p,q)$ model and fit it on the training dataset and check the significance of the obtained parameter estimates. Then the normality and autocorrelation of squared standardized residuals are formally tested, and the optimal model is selected on the basis of Information criteria (AIC, BIC) and forecasting errors estimation. An additional diagnostic of asymmetrical conditional variance reaction or the distribution of conditional error could be conducted by applying other GARCH extensions, i.e. EGARCH or GARCH-in-Mean, respectively.

Now we provide the theoretical background of GARCH specifications that are used in an empirical part of our work. Since the time when the first GARCH model was developed, there have been many extensions proposed in order to overcome the basic version's limitations (see comprehensive study by Bollerslev, 1986 for more details).

*3.1.1. GARCH by Bollerslev (1986)*

The $GARCH\ (p,q)$ was proposed by Bollerslev (1986) and it organically evolved from the ARCH model allowing the conditional variance ($\sigma_t^2$) to be dependent upon its own previous lags. Assuming that the mean equation for the series of logarithmic returns $r_t$ is described by the autoregressive moving average (ARMA) process and that the mean-adjusted logarithmic return is $u_t = r_t - \mu_t$, we may say that $u_t$ follows a classical $GARCH\ (p,q)$ process if (see Tsay 2002):

$$\begin{cases} u_t = \sigma_t \epsilon_t \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \end{cases} \quad (4)$$

where:

$\sigma_t^2$ – known as the *conditional* variance as it is a one-period ahead variance estimator which depends on the past information given $p$ and $q$ lags,

$p$ – maximum lag for the conditional variance,

$q$ – maximum lag of the squared mean-adjusted (unexpected) return,

$\epsilon_t$ – random term which is independently and identically distributed with 0 mean and 1 variance.

The simplest and most frequently used specification if (4) is the GARCH (1,1) model, which could be written as follows:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (5)$$

given the constraints: $0 \leq \alpha_1, \beta_1 \leq 1, (\alpha_1 + \beta_1) < 1$. Unlike the conditional variance that changes, the unconditional variance of $\alpha_t$ is constant and for GARCH(1,1) is defined by:

$$var(u_t) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \quad (6)$$

If $(\alpha_1 + \beta_1) < 1$ does not hold, then $var(u_t)$ is not defined in (6) which means that variance is not stationary. It must be pointed out though that non-stationarity in variance has little theoretical motivation for existence (as it would be for the mean non-stationarity in price series, for example). Brooks (2000) claims that GARCH models which coefficients implicit variance non-stationarity would have undesirable properties.

There are two advantages of the GARCH model: it accounts for fatter tails of returns in comparison to the normal distribution and therefore considers the properties of times series under more realistic perspective, and it provides relatively simple parametric function that describes how volatility evolves. Given $u_l, and\ \sigma_l^2$ are known of time for the one step-ahead volatility forecast at the moment of $l$ can be given by:

$$\sigma_{l+1}^2 = \alpha_0 + \alpha_1 u_l^2 + \beta_1 \sigma_l^2 \tag{7}$$

Although GARCH is believed to be a better version of ARCH it still makes an equal response to the positive and negative news impact.

*3.1.2. EGARCH by Nelson (1991)*

EGARCH was proposed by Nelson in 1991 to deal with the volatility leverage effect. The conditional variance ($\sigma_t^2$) is described by the following equation, see Brooks (2008):

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \gamma \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[ \frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] \tag{8}$$

where:

$u_t = \sigma_t \epsilon_t$ – mean-adjusted logarithmic return,

$\sigma_t > 0$,

$\gamma$ – parameter that accounts for asymmetry,

$\epsilon_t$ – sequence of independently and identically distributed variables with 0 mean and 1 variance.

The difference is that in classical GARCH model volatility is an additive function of past standardized innovations ($\epsilon_t$ divided by their conditional standard deviations), whereas in EGARCH model volatility is an explicit multiplicative function of lagged innovations. The non-negativity constraint does not apply in EGARCH because the logarithm of variances is modelled. EGARCH can react asymmetrically to good or bad news: parameter $\gamma$ takes into the account the sign and extent of innovation, for example, if the observed relationship between returns and volatility is negative, $\gamma$ will be negative as well. Given stylized fact is considered, EGARCH is expected to produce better forecasts. Brooks (2008) mentions that in the majority of EGARCH applications conditional normal errors are employed rather than Generalized Error Distribution (GED) due to their relative easiness in computations and intuitive interpretation.

*3.1.3. GJR-GARCH of Glosten et al. (1993)*

Glosten, Jagannathan, and Runkle (1993) proposed an extension which allows the seasonal patterns in volatility and considers the leverage effect in volatility. The negative and positive shocks on the conditional variance are asymmetrically modelled by introducing indicator function *I*. The conditional volatility ($\sigma_t^2$) is then described by, see Ghalanos (2019):

$$\sigma_t^2 = (\omega + \sum_{j=1}^{m} \varsigma_j v_{jt}) + \sum_{j=1}^{q} (\alpha_j \varepsilon_{t-j}^2 + \gamma_j I_{t-j} \varepsilon_{t-j}^2) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad (10)$$

where:

$\omega, \beta, \alpha$ – coefficients,

$v_{jt}$ – optional external regressors passed pre-lagged,

$\gamma_j$ – the 'leverage' term,

*p* – lag for the conditional variance and *q* is lag of the squared error.

The indicator function $I_t$ is described as follows:

$$I_{t-1} = 1 \; if \; \varepsilon_{t-1} < 0 \qquad (11)$$
$$I_{t-1} = 0 \; for \; \varepsilon_{t-1} \geq 0$$

If the leverage effect $\gamma_j$will be higher than 0, then the model allows for after the shock with the same magnitude different volatility value changes, given $\gamma_j$ is significant. Thus, theoretically speaking given the leverage effect is considered GJR model should lead to forecasting performance enhancement. There is an empirical evidence for that: according to Laurent et al. and Brownlees et al. (2012) GJR produces better forecasts in comparison to classical GARCH specification.

*3.1.4. fGARCH – TGARCH (sub-model) by Zakoian (1994)*

The threshold volatility model (TGARCH) by Zakoian (1994) relates to the GARCH family models (fGARCH) of Hentschel (1995) which subsumes some of the most popular GARCH specifications. In regime switching TGARCH the state of the world is governed by an observable threshold variable and known therefore, while the conditional variance follows the

GARCH process within each state. The general formula of the conditional variance ($\sigma_t^\lambda$) under fGARCH family could be written as, see Ghalanos (2019):

$$\sigma_t^\lambda = w + \sum_{j=1}^{q} \alpha_j \sigma_{t-j}^\lambda \left(\left|z_{t-j} - \eta_{2j}\right| - \eta_{1j}\left(z_{t-j} - \eta_{2j}\right)\right)^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\lambda \qquad (12)$$

where:

$\alpha, \beta, \omega$ – model coefficients,

$\lambda, \delta, \eta_1, \eta_2$ – parameters described below.

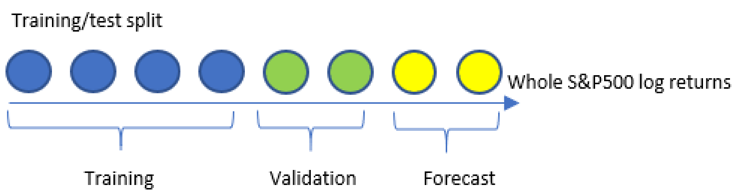
The decomposition of residuals is allowed, and it is driven by the different powers for $z_t$ and $\sigma_t$. Moreover, rotations and shifts are enabled in the news impact curve, namely the shift takes place when the small shocks occur, whereas the rotation is for the large shocks. The shape of the conditional variance is ruled by $\lambda$ parameter. The absolute value function is transformed via $\delta$ parameter, $\eta_1$ and $\eta_2$ are parameters via which rotations and shifts are conducted. One of the advantages of the TGARCH is that linear restriction on the conditional variance dynamics is relaxed.

TGARCH sub-model have the following parameters: $\lambda = \delta = 1, \left|\eta_{1j} \leq 1\right|, \eta_{2j} = 0$ which have to be input in the formula above (12).

*3.2. Rolling-based forecasting approaches*

The basic forecasting approach is to split the time series into : 1) training (in-the-sample), 2) validation, and 3) test (out-of-sample) datasets:

**Figure 1. Classical forecasting approach**

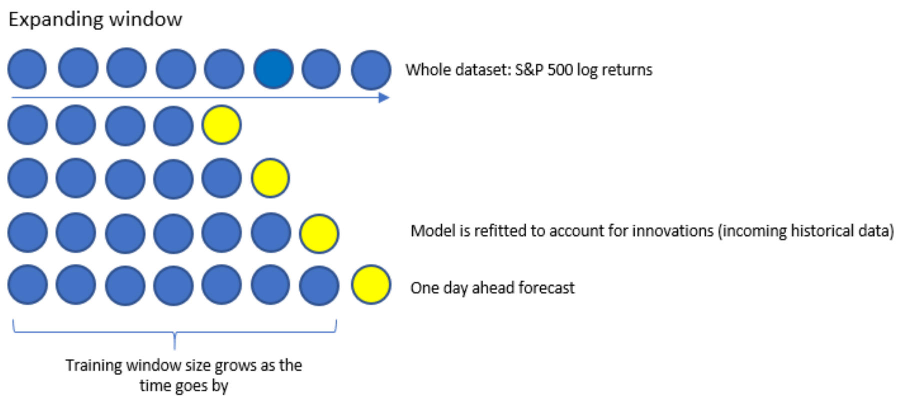



Source: Own example

The model is fitted on the training set and the best one is selected on the basis of the accuracy of forecasts obtained on validation set (pre-test). Then one or multi period ahead forecasts are produced on the test dataset with respect to the fitted parameters estimated from the training dataset. There is still a continuous conceptual debate around choosing the proper training/test split. In practice 80/20 is often used, but there is no right answer and in most of the cases the split depends on the peculiarities of data and the model constructed. The second point is that parameters usually change over time and information used in the initial training dataset does not account for the innovations – incoming data.

Figure 2 and Figure 3 visualize the concept of LOOCV approach for an expanding and moving training windows.

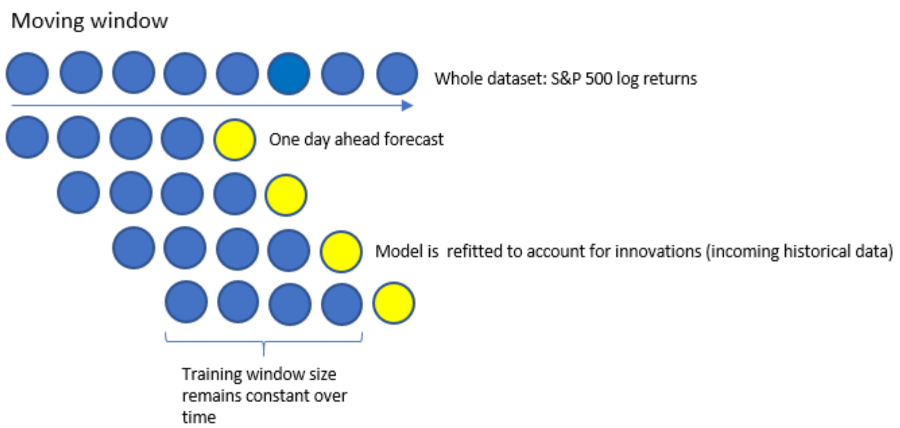
**Figure 2. Rolling forecast technique: Expanding window**




Source: Own example

Unlike the above, the rolling-based technique takes into the account the whole dataset: each data point falls into the training and test datasets depending on the moment of the model estimation. Thus, the time varying factor is considered, and the model is adjusted accordingly. Additionally, it is possible to refit the model, and the frequency of such refits depends on the training window size, distributional properties of time series, asset class type and its peculiar traits identified if any. Therefore, in the back-testing the 'leave one out cross validation' (LOOCV) technique is the powerful one making one day ahead forecast with the model re-estimations across different GARCH specifications.

**Figure 3. Rolling forecast technique: Moving window**




Source: Own example

Expanding window includes all the previous data points and as the time goes by the new observations are included with the old ones remaining. In contrast, there is the moving window which size is fixed at each estimation, and it is shifting by n-days ahead, one-day ahead here.

There are also other forecasting methods based on neural networks and machine learning techniques, but they do require much time and power for computing and are beyond the scope of this paper.

*3.3. Forecasting errors and the economic loss function (ELF)*

The successful volatility model is primarily determined by its ability to accurately forecast on the out-of-sample dataset. Therefore, the prediction accuracy is one of the most important indicators. According to the 'Ex post' forecasting quality estimation the forecasted values are compared with the actual ones on the "post the event" basis.

In the academic literature the metrics proposed are as follows: Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Theil-U statistic and others, for example (see Ederington and Guan 1999 and Brooks 2008).

It must be pointed out that a trading strategy based on the model with lower estimation errors does not necessarily produce higher profits than the one with higher errors, and vice versa – there are many empirical evidences Gerlow et.al (1993), for example. Some authors criticize deviation-based metrics (MAE, RMSE) and argue that models constructed on the percentage of correct sign or direction change lead to relatively higher profits, see Leitch and Tanner (1991). Given there are opposing views we compute forecasting errors based on RMSE and the percentage of the correct sign change prediction.

The RMSE formula is as follows:

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(VolF_t - VolH_t)^2} \tag{13}$$

where:

$T$ - the total number of observations,

$VolF_t$ – the forecasted volatility,

$VolH_t$ – the estimated volatility.

The $VolH_t$ is calculated based on the formula (1) for close-to-close estimator with n=21.

The Economic Loss Function (ELF) will be defined as the percentage of the correct sign change prediction. The formula (Brooks, 2008) is given by:

$$ELF = \frac{1}{T - (T_1 - 1)} \sum\nolimits_{t=T_1}^{T} z_t \quad (14)$$

$$z_t = 1 \; if \; \left(VolF_{t,s} * VolH_{t,s}\right) > 0$$

$$z_{t+s} = 0 \; otherwise$$

where:

$z_t$ – correct sign prediction at *t* moment,

*T* – the total number of observations.

This metric shows how many times correct sign change is predicted to the total number of observations. ELF is more appropriate error metric in comparison to RMSE when we want to focus on the correct prediction of direction only instead of accounting for large values of errors.

## 4. Data and research description

### *4.1. Data*

In our research there are two time series being observed and analyzed: daily OHLC quotes of S&P 500 index and VIX futures in the period between 2009-01-01 and 2019-10-03, and from 2013-01-02 to 2019-10-03, respectively. All historical VIX futures data is downloaded from the CBOE Global Markets, Inc. website: https://markets.cboe.com/us/futures/market_statistics/historical_data/.[1] We developed the R code that download automatically daily VIX futures prices with all available weekly and monthly expirations since January the 2nd, 2013 till October the 3rd, 2019.

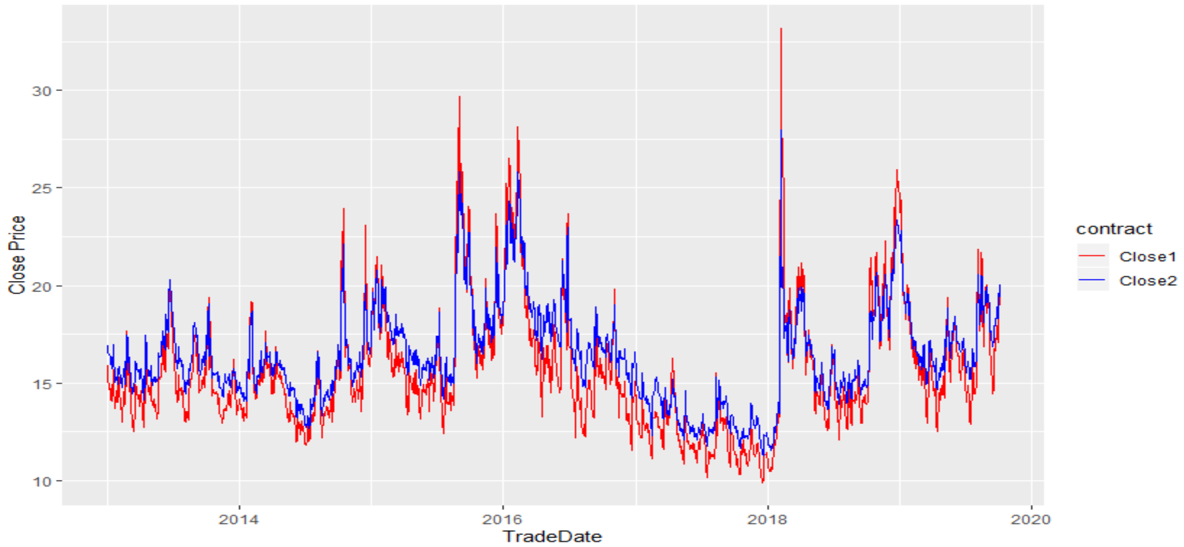
We filtered VIX data by contracts with monthly expirations only as they are the ones used in our investment strategies. In order to calculate profit and loss for investment strategy VIX futures with the shortest maturity were selected. On the trade day there are two contracts that are the closest to expiration among all contracts available (the first to be expired and the consecutive one) with their respective contract names, close prices and time to maturities. Figure 4 shows these VIX futures close prices.

Volatility of VIX futures fluctuations can be very high and heavily affects the results of various investment strategies, see Table 1 and Figure 5. On 2018-02-5, nine days before its

[1] All available data on the website

expiration, the "G (Feb 2018) contract" experienced the largest daily price increase since its introduction. Measured as percentage and point change, it was respectively +112.4 % and +17.57. This day it reached its peak at \$33.2 while on 2017-12-18 it was just traded at \$9.88 close price minimum.

**Figure 4. VIX futures with different maturities**



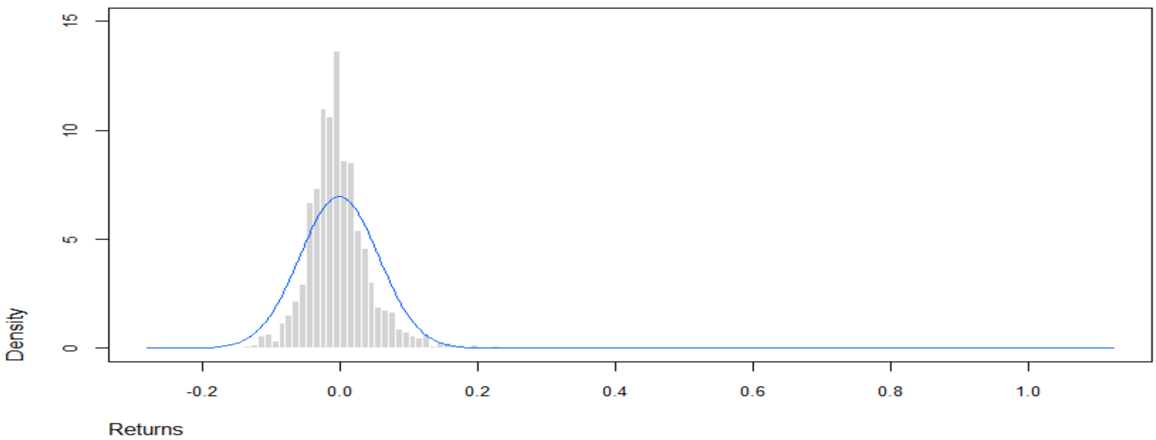

Note: The nearest to expiration (Close 1) and the 2nd nearest to expiration (Close 2) contracts. In most cases Close 2 is (as expected) above the Close 1 on the whole time period observed

**Table 1. Descriptive statistics of nearest to expiration VIX futures returns.**

| *Mean* | *Std. dev.* | *Median* | *Min* | *Max* | *Range* | *Skewness* | *Kurtosis* |
|---|---|---|---|---|---|---|---|
| -0.00151 | 0.05726 | -0.0071 | -0.28012 | 1.12412 | 1.14042 | 5.2866 | 5.3641 |

Note: These descriptive statistics were calculated in the period from 2013-01-02 to 2019-10-03. The range exceeded 114%. Skewness was high due to these extreme movements

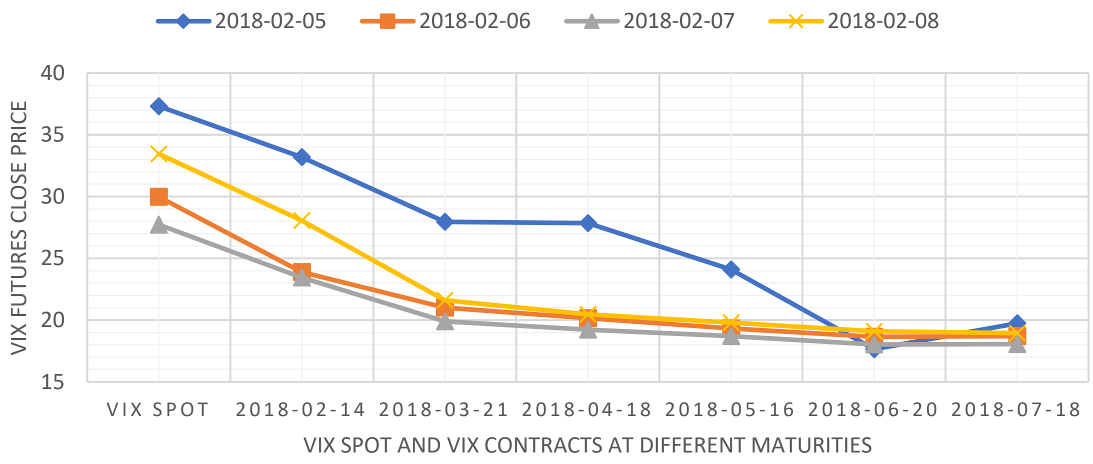
**Figure 5. The returns distribution of the nearest to expiration VIX futures**




Note: The distribution of returns is skewed towards the right and have high kurtosis. On 5th February return was 112% (all-time record), whereas the minimum returns drop was 28%

Such periods, after sharp volatility jump are usually characterized by inverted curve of volatility term structure, see Figure 6, for example. Normal or inverted curves show futures close prices at different contract maturities in the specific point of time, whereas contango and backwardation compare the actual futures prices with the expected futures spot prices. If actual prices are higher than the expected spot ones, then we may say that over time prices' structure is falling to converge to spot and market is in contango state. Vice versa is for normal backwardation given actual prices are lower than the expected futures spots.

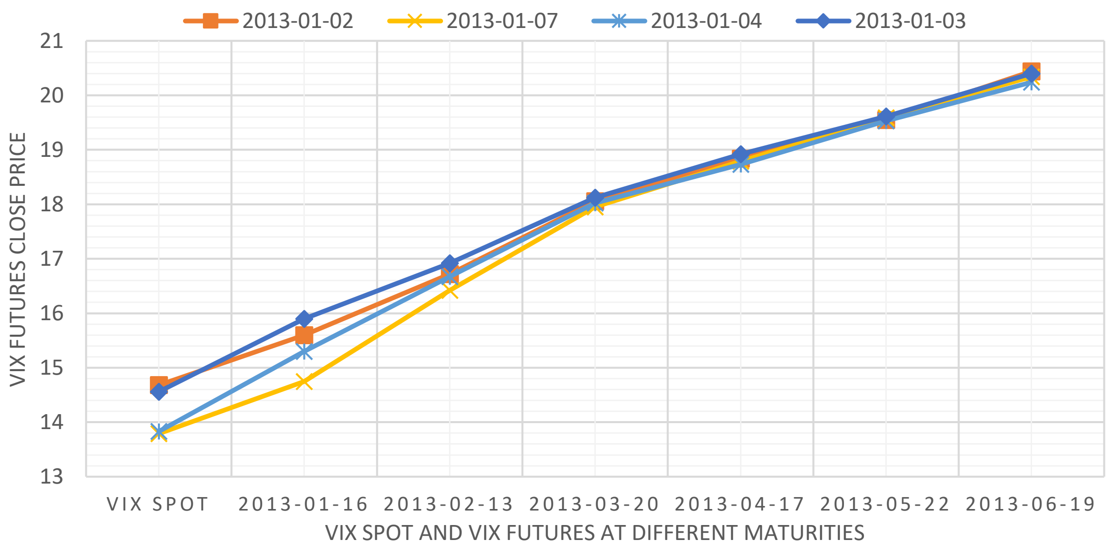
**Figure 6. Term structure of VIX futures and spot prices (Inverted curve)**




Note: In all four trading days observed VIX spot exceeded VIX futures prices at different maturities

In contrary, normal VIX futures curve is observed when there is relative VIX futures standstill, for example, for the very first four contract trading days, it looks like on Figure 7.

**Figure 7. Term structure of VIX futures and spot prices (Normal curve)**




Note: In all four trading days VIX spot was below VIX futures prices at different maturities

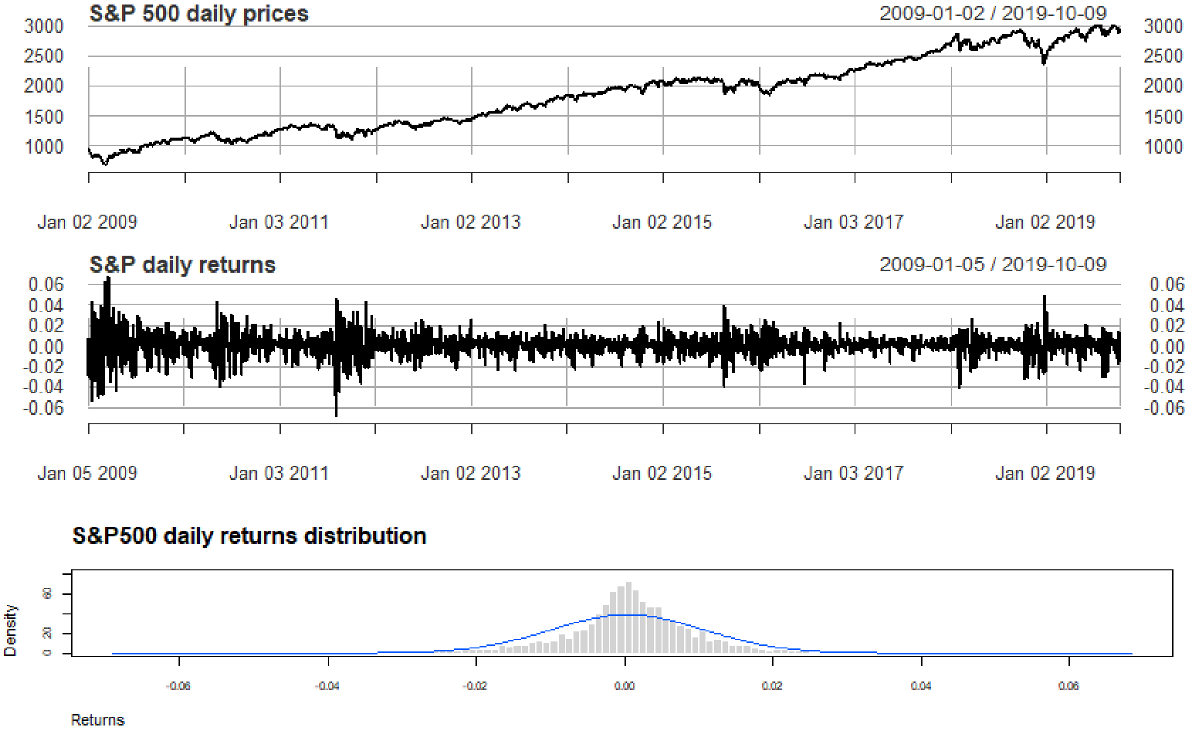
The daily observations of S&P 500 index have been downloaded from Yahoo Finance from 2009.01.03 to 2019.10.10 – 5029 time series data points in total. There are Table 2 with the descriptive statistics and Figure 8 below representing S&P 500 closing prices, its respective logarithmic returns along with the returns' distribution.

**Table 2. Descriptive statistics of S&P500 returns over the period 2009-2019**

| *Mean* | *Std. dev.* | *Median* | *Min* | *Max* | *Range* | *Skewness* | *Kurtosis* |
|---|---|---|---|---|---|---|---|
| *0.00043* | 0.00938 | 0.00056 | -0.06666 | 0.04959 | 0.11622 | -0.3416 | 4.9753 |

Note: The range of S&P500 returns is more than ten times less than VIX futures range

**Figure 8. S&P 500: close price, logarithmic returns, distribution**




Note: Overall eleven-year S&P500 upward trend is observed on the top graph, respective distribution of logarithmic returns (bottom graph) has high kurtosis and is skewed slightly towards the left

It is clearly seen that after financial crisis in 2008 S&P 500 index has been growing steadily dropping only in December 2018 around 10%. In the last five-year period buy-and-hold strategy produced 10.7 % of annualized return. The return distribution is leptokurtic and right-skewed slightly - normality is rejected by Shapiro-Wilk test as well (W=0.94387, p-value $< 0.22 * 10^{-15}$). The volatility of returns was the largest during the financial crisis 2008, the volatility clustering is observed in the middle graph in Figure 8 above.

*4.2. Research description*

The aim of this work is to compare the performance of VIX futures trading strategies across different GARCH specifications. The research stages are as follows.

1. The data is collected and prepared: S&P 500 index and VIX futures daily observations. The performance of the passive long and short volatility strategies in VIX futures along with the buy-and-hold S&P 500 index strategy is computed and referenced as the benchmark. Passive long and short volatility strategies assumes that we are always 100% long or short in VIX futures, in the monthly contract which is the closest to maturity.
2. Volatility of S&P 500 returns is forecasted by univariate GARCH models estimated on a rolling basis and forecasting errors are calculated across different metrics. In our research we exclusively use the rolling approach with one day ahead forecast. Rolling approach caused that we had to estimate each specification of GARCH model around 2500 times on the S&P 500 data since 2009-01-01 to 2019-10-03. Obviously, we are not able to carry out the full diagnostics for each time we fit the model, as described in Section 2.1. Instead of this, we monitor model coefficients while refitting the model. They are presented in Figure 11. For example, for the classical GARCH (1,1) we visualize the sum of alpha and beta under unconditional variance constraint over time which must be less than 1.
3. Trading strategy based on the rolling forecasts and different historical volatility estimators is constructed. Given that a signal might be sensitive to the historical volatility estimator we construct it under the following three volatility estimators: “Close-to-Close”, “Garman-Klass”, and “Parkinson”.
4. The sensitivity analysis with special regards to training window size, refit frequency, model GARCH specifications, rolling parameters, and historical volatility estimators is conducted in the end in order to check whether the obtained strategy performance is robust or not.

### *4.3. Investment strategy*

The key assumptions of investment strategy are the following. In the base case the degree of financial leverage *(DFL)* equals 25%, whereas initial amount of capital *(capital)* invested is constant and set to be 1 million USD. The multiplicator *(mult)* for one-point change of VIX futures is 1000 USD.

We invest in the nearest to expiration monthly VIX futures contract. On the day before expiration of this contract we enter the second closest to expiration contract which becomes the next day the following first to expire, and so on.

The trading signal mechanism is as follows:

$$signal_t = \begin{cases} 1 \text{ if } VolF_{t+1} > VolH_t \\ -1 \; if \;\; VolF_{t+1} < \; VolH_t \\ signal_{t-1} \; otherwise \end{cases} \tag{15}$$

where $VolF_{t+1}$ and $VolH_t$ are volatility forecast and estimated historical volatility, respectively.

If the forecasted volatility on a one day ahead ($t$+1) is higher/lower than the current ($t$) historical volatility ($t$), then long/short position in the nearest to expiration VIX monthly contract is undertaken. If there is no change – then we stay on the market with the previous position. This signal generation is a very important moment as the strategy performance will depend on when and which positions are entered. There are multiple ways of how to estimate historical volatility. Therefore, we test three estimators: “Close-to-Close”, “Garman-Klass”, and “Parkinson” – in order to check stability of signals.

The special attention must be given to the absolute profit and loss *(PnL)* calculation at the trading days when contract rolling takes place. For example, there are two contracts *F (Jan 2013)* and *G (Feb 2013)* and *F (Jan 2013)* expires first, while *G (Feb 2013)* – second. One day before *F (Jan 2013)* expiration (*t*-1) we close position in F (Jan 2013) and according to the current signal we open the position in *G (Feb 2013)*. *PnL* on the following day (*t*) is calculated as the close price difference for *G (Feb 2013)* multiplied by the previous signal.

Post that we calculate the number of positions available ($nop_t$):

$$nop_t = \frac{capital_t * DFL}{close_t * mult} \tag{16}$$

The equity line *(portfolio)* including transactional costs *(TC)* set at $2 level per one contract is given by:

$$portfolio_t = portfolio_{t-1} + nop_{t-1} * PnL_t * mult - TC * |nop_t - nop_{t-1}| \tag{17}$$

The comparison of the performance of investment strategies based on the forecasted values could be conducted basing on the following metrics, (see Ryś and Ślepaczuk, 2019):

1. Annualized rate of return (ARC) – shows annualized percentage return an investment generates each year over a time period specified $(t_1, t_2)$:

$$ARC(V)_{t_1}^{t_2} = \left(\frac{V_{t_2}}{V_{t_1}}\right)^{\frac{1}{D(t_1,t_2)}} - 1 \tag{18}$$

where: asset of value process is $V_t$, $D(t_1, t_2)$ is the time between $t_1$ and $t_2$ in years.

2. Annualized standard deviation (ASD) – the empirical standard deviation normalized, according to time:

$$ASD(V)_{t_1}^{t_n} = \sqrt{\frac{1}{n}\sum_{t=t_1}^{t_n}(R_t - \bar{R})^2 * \frac{1}{D(t_1, t_2)}} \quad (19)$$

where: $R_t$ is series of returns, $\bar{R} = \frac{1}{n}\sum_{t=t_1}^{t_2} R_t$ *and* $D(t_1, t_2)$ is the time between $t_1$ and $t_2$ in years.

3. Maximum Drawdown (MD) is the maximum portfolio percentage loss (from a peak to the bottom before a new peak is attained) observed in equity line over a time period specified $(t_1, t_2)$ and for price series $S_t$ described by:

$$MDD(S)_{t_1}^{t_2} = sup_{(x,y)\in\{[t_1,t_2]^2 : x \leq y\}} \frac{S_x - S_y}{S_x} \quad (20)$$

4. Information Ratio represents how much units of annualized returns are obtained per one unit of annualized standard deviation:

$$IR = \frac{ARC(V)_{t_1}^{t_2}}{ASD(V)_{t_1}^{t_n}} \quad (21)$$

All metrics except IR are expressed as percentage values and were calculated by appropriate functions from the 'Performance Analytics' R package. We assumed 252 trading days for VIX futures annually, hence the scaling parameter is set accordingly to 252.

4.4. Rolling forecast implementation

The key functions used come from the 'rugarch' R package. The GARCH specification is chosen by 'ugarchspec' function. The key arguments to be defined are as follows:

- `variance.model` There is a wide variety of models (GARCH, EGARCH, GJR-GARCH, etc.) and their respective sub-models to be chosen from the proposed list. Additionally, there is an option to include external regressors and variance targeting approach;
- `garchOrder` $p$ *and* $q$ lags to be set for ARCH and GARCH, respectively. We use $p$ *and* $q$ both equal to 1 as it was enough to address heteroscedasticity in the data;
- `mean.model` specification of the expected mean equation. We included the mean parameter in our all four GARCH specifications;
- `distribution.model` type of conditional density to describe mean-corrected returns (normal, skewed-normal, normal inverse gaussian);

Once specification is chosen the rolling GARCH forecast is run by implementing ‘ugarchroll’ function on the whole dataset with the following parameters.

- `spec` the GARCH model specification by ‘ugarchspec’ function described above;
- `refit.window` There are two types: expanding which includes all the previous data points and as the time goes by the new observations are included with the old ones remaining; and moving – when the size of the training window is fixed at each estimation, and it is shifted by n-days ahead;
- `window.size` It determines the size of the training window in the rolling estimation, namely, the starting point of the forecast initialization (for example, window size which is set at 252 means that the first forecast will be produced on 253rd trading day);
- `n.ahead`. one day ahead forecast is set by default;
- `refit.every` This argument defines the frequency of the model refitting: daily, weekly, monthly, etc.;
- `solver`: type of solver to be used. We encountered an issue with the non-converged window estimations which has been successfully resolved by additionally using ‘resume’ function and passing the arguments to ‘gosolnp’ solver which appeared to be much more efficient compared to the other ones.

## 5. Empirical results

### *5.1. The base model*

In this section we compare the strategy based on the GARCH model (referred as the base model) with the three benchmark strategies. The specified characteristics of the base model are as follows:

1. specification: GARCH (1,1) by Bollerslev (1986)
2. training window size: 252 days
3. refit every: 21 trading day
4. window type: moving
5. distribution model: normal
6. historical volatility estimator: "Close-to-Close"

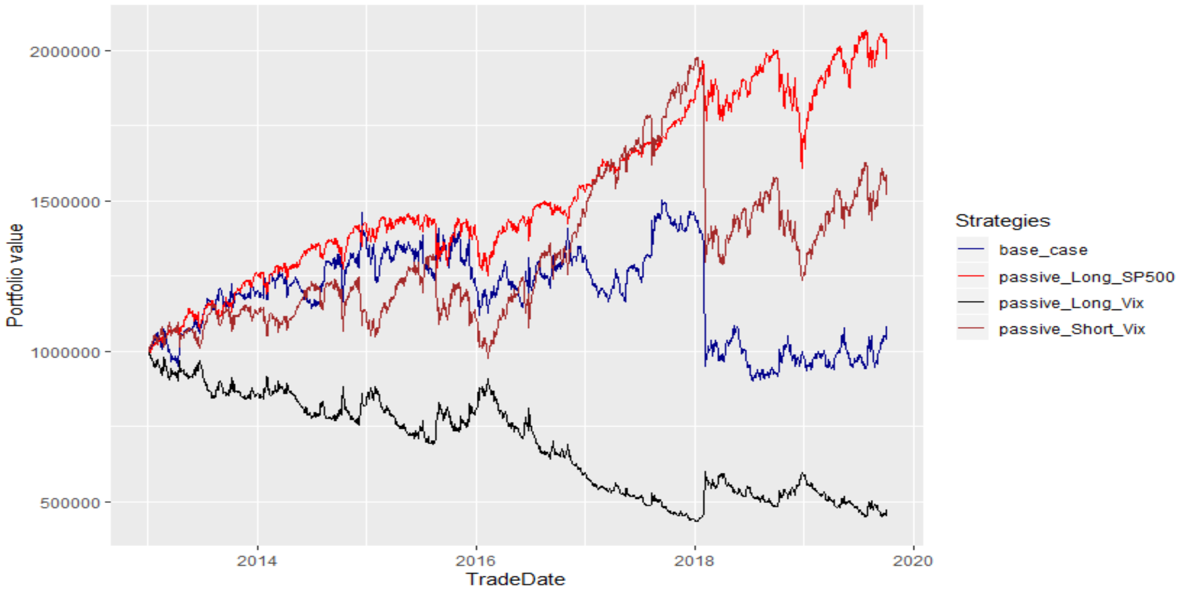
7. degree of financial leverage (DFL): 0.25

The benchmark is being compared with:

- Passive long strategy in S&P 500
- Passive long strategy in VIX futures
- Passive short strategy in VIX futures

Transactional costs are included for all strategies presented here and further. According to the passive strategy trading position (respective long or short) is entered on 2013-01-02 and remained unchanged until 2019-10-03. The performance of the strategies is presented on Figure 9 and Table 3.

**Figure 9. Equity lines of the base model against three benchmark strategies: 1) passive long in VIX futures, 2) passive short in VIX futures and 3) passive long in S&P 500**




Note: Buy-and-hold S&P500 and passive short VIX futures strategies dominate over the rest.

**Table 3. The base model performance against the benchmark**

| *Strategy* | *Annualized Return, in %* | *Annualized SD, in %* | *Information Ratio* | *Maximum Drawdown, in %* | *Total number of trades* | *% of long signal to all signals* | *RMSE* | *Correct sign change prediction, in %* |
|---|---|---|---|---|---|---|---|---|
| *Passive long S&P 500* | 10.7 | 12.9 | **0.83** | 19.7 | 1 | 100 | - | - |
| *Passive long VIX* | -10.6 | 21.4 | **-0.498** | 56.7 | 1 | 100 | - | - |
| *Passive short VIX* | 6.6 | 22.1 | **0.299** | 37.4 | 1 | 0 | - | - |
| *base model* | 1.00 | 22.0 | **0.045** | 39.9 | 159 | 57 | 0.232 | 53.7 |

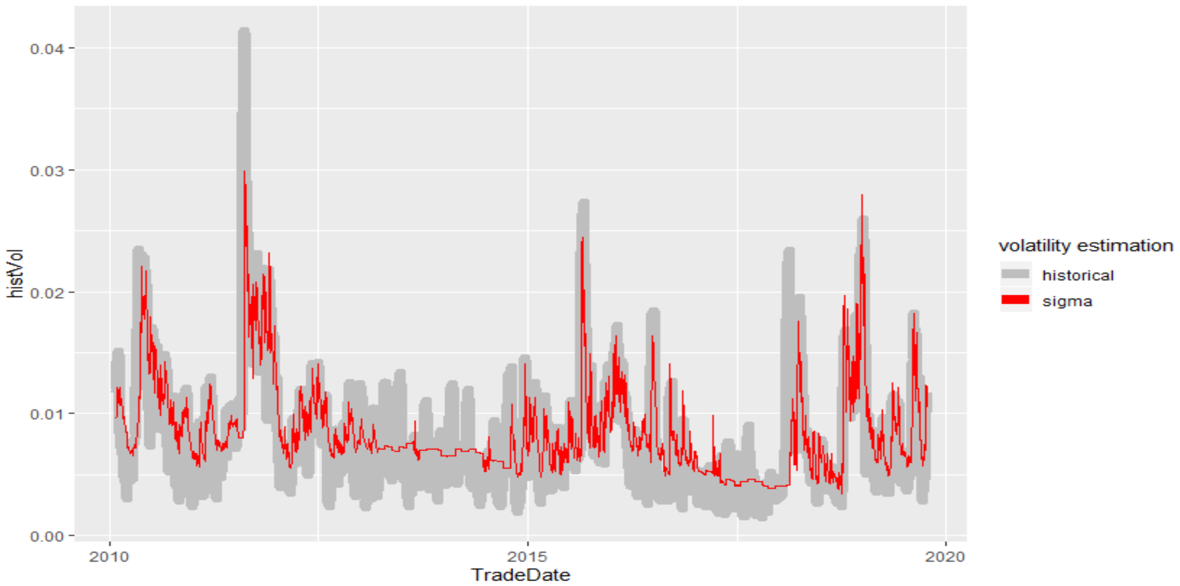
Note: All GARCH models have notoriously large maximum drawdowns and are almost two times riskier than passive long S&P 500 strategy

The portfolio value based on the long VIX strategy steadily declined, whereas - the S&P 500 index grew consistently over the whole period observed. The short VIX exceeded the S&P 500 until the beginning of 2018, and then dropped significantly when VIX futures experienced the all-time percentage close price decline. The base model dominated the passive short VIX, which performed little below the S&P 500 benchmark until the middle of 2017 and then dropped significantly following the same decrease as VIX short strategy in the beginning of the February 2018.

Despite early relative success the strategy based on the base model ended up at around 1% of annualized return and the slightly positive Information Ratio (0.045). Annualized standard deviations for all VIX-based strategies are similar and almost two times higher compared to S&P 500. According to the base model. More than 50% of the sign change has been accurately predicted, whereas RMSE value is 0.232%. Essentially, the investment strategy performance depends on the comparison between the historical volatility and sigma, what is presented in Figure 10.

Over the eleven-year period of 2009-2019 the S&P 500 index showed an upward trend with relatively minor declines, whereas the volatility has been steadily decreasing. Given this the volatility strategies in which more long signals occur are expected to perform worse, and vice versa.

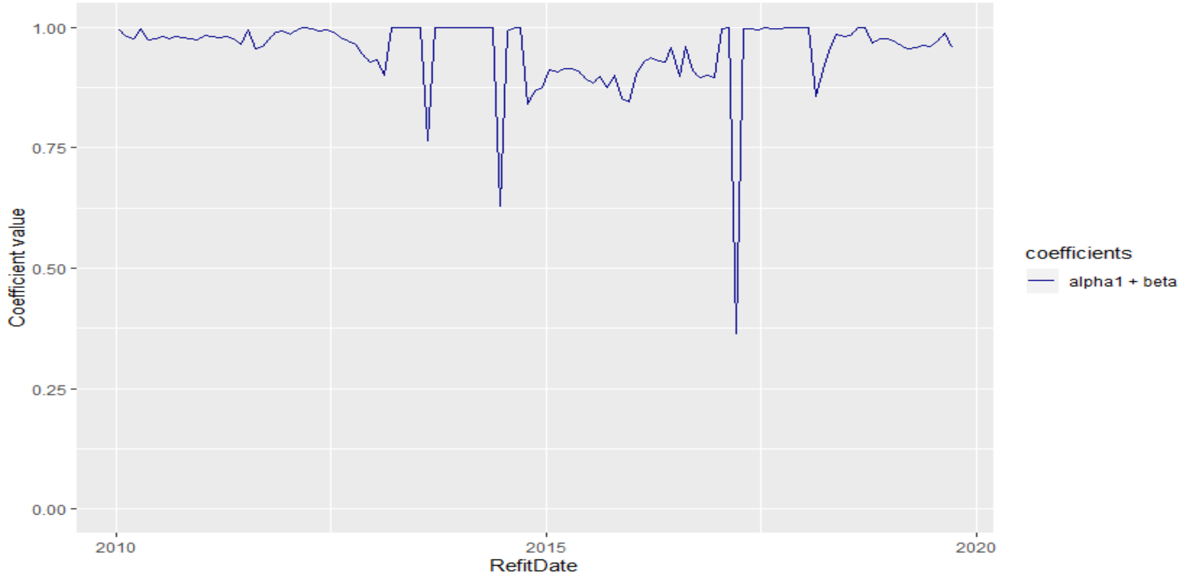
**Figure 10. Daily historical volatility by "Close-to-Close" estimator and volatility forecasted by the GARCH (1,1) model**




Note: Forecasted volatility does not significantly diverge from the historical volatility and significant moves are captured

Fitting coefficients for unconditional variance constraint in GARCH model are acceptable over the whole observed period as the sum of $alpha_1$ and $beta_1$ from the unconditional volatility constraint in equation (6) does not exceed the threshold:

**Figure 11. The sum of the rolling GARCH (1,1) parameter estimates over time.**




Note: We can see that the sum of $\text{alpha}_1$ and $\text{beta}_1$ does not exceed one but sometimes is equal to one. If the sum is equal to 1, then current information remains important when forecasting the volatility for all time horizons and the correct model specification is integrated GARCH (I-GARCH), introduced by Engle and Bollerslev (1986).

## 5.2. Other strategies based on the GARCH volatility forecasts.

In this section we compare performance of numerous trading strategies constructed on conditional variance forecasts from different GARCH models specifications. We also take into account a wide set of strategy assumptions. Our main aim is to identify the robustness of the results to initial assumptions and indirectly assess which model or set of models produces the best investment outcomes.

The set of tested parameters is as follows:

1. Specification: GARCH, EGARCH, GJR-GARCH, fGARCH-TGARCH[2] (all are of 1,1 order)
2. Training window: 126 days, 252 days, 504 days
3. Model refit frequency: 21, 63, 126 trading days
4. Window type: moving, expanding
5. Conditional distribution of returns: normal (norm), skewed-normal (snorm), normal inverse gaussian (nig)

[2] In rugarch package these specifications are denoted as follows: sGARCH, eGARCH, gjrGARCH, and fGARCH, respectively. We use these notations in the text.

6. Historical volatility estimator: "Close-to-Close", "Garman-Klass". The Parkinson estimator was not included in the sensitivity analysis since its volatility forecasts were very close to those produced by the Garman-Klass estimator– respective outcomes do not make any substantial changes to the results.

There are in total 4*3*3*2*3*2=432 combinations of the above parameters. The performance metrics, the forecasting errors, and the percentage of 'long" signals of each model are computed in R via the custom written loop. In total we obtained 430 outcomes, as for two of them there were irresolvable issues with convergence of the optimization algorithm. The outcomes were filtered by "Close-to-Close" historical volatility estimator across all distributions. Given there were no obvious outliers, we computed the average of the performance metrics in each group of tested parameters. The three key observations are presented below. We compared the strategies with respect to the Information Ratio metric *(IR)*, as described in formula (21) in Section 3.3.

### *5.2.1. Model specification*

Table 4 and Figure 9 present results of strategies with respect to different GARCH model specification. The fGARCH-TGARCH and GJR-GARCH performed significantly better that the GARCH and EGARCH.

**Figure 12. Investment outcome across GARCH specifications**


Note: Annualized standard deviation and average of long signals proportion are quite similar across all GARCH specifications

**Table 4. Investment outcome across GARCH specifications**

| *Specification* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of ELF* | *RMSE* |
|---|---|---|---|---|---|---|---|---|
| *eGARCH* | 1.78 | 22.01 | **0.08** | 42.96 | 176.78 | 55.71 | 56.50 | 0.254 |
| *fGARCH* | 9.15 | 22.17 | **0.41** | 31.21 | 196.52 | 56.23 | 58.01 | 0.302 |
| *gjrGARCH* | 8.03 | 22.18 | **0.36** | 37.14 | 166.63 | 56.33 | 57.64 | 0.294 |
| *sGARCH* | -0.22 | 21.63 | **-0.01** | 48.08 | 152.00 | 56.77 | 53.53 | 0.259 |
| *Grand Total* | 4.69 | 22.00 | **0.21** | 39.85 | 172.98 | 56.26 | 56.42 | 0.852 |

Note: The ELF (economic loss function) is the percentage of the correct sign change prediction as described in the formula (14) in Section 2.3. above.

Despite having significantly higher RMSE in comparison to the classical GARCH model, strategies based on the fGARCH-TGARCH and GJR-GARCH model specifications perform better than those based on the GARCH models in terms of Information Ratio metric: 0.41 and 0.36 against -0.01, respectively. The fGARCH-TGARCH model shows the outstanding 9.14% of the annualized return and 31.21% of the maximum drawdown. It has been also quite accurate in terms of the correct sign change prediction (58%), while the GJR-GARCH did slightly worse (57.6%). On average, the fGARCH-TGARCH generates the most signals (196.52), whereas GARCH – the least (152). The detailed results for model specification are summarized in Tables 14-17 in Appendix.

### *5.2.2. Training window size*

Results of the strategies with respect to different training window size are shown in the Table 5. All strategies performed similarly, and we cannot identify any substantial differences among them.

**Table 5. Investment outcomes across trading window size**

| *Size* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of ELF* | *RMSE* |
|---|---|---|---|---|---|---|---|---|
| *126* | 4.56 | 21.78 | **0.21** | 39.55 | 177.31 | 54.46 | 54.75 | 0.32 |
| *252* | 4.39 | 22.11 | **0.20** | 45.03 | 170.75 | 56.39 | 56.76 | 0.29 |
| *504* | 5.11 | 22.10 | **0.23** | 34.95 | 170.89 | 57.92 | 57.75 | 1.9 |
| *Grand Total* | 4.69 | 22.00 | **0.21** | 39.85 | 172.98 | 56.26 | 56.42 | 0.84 |

Note: The longest training window produced the best outcomes under IR and was the most accurate for correct sign change prediction, but the differences are rather small

The detailed results for various window size are presented in Tables 18-20 in Appendix.

### *5.2.3. Refit window*

Table 6 presents results of the strategies with respect to the GARCH model refitting frequency. Again, we do not observe any significant differences among them.

**Table 6. Investment outcomes across GARCH model refitting frequency**

| *Refit frequency* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of elf* | *RMSE* |
|---|---|---|---|---|---|---|---|---|
| *21* | 4.97 | 21.81 | **0.22** | 37.69 | 186.31 | 54.40 | 55.95 | 1.52 |
| *63* | 4.51 | 22.10 | **0.20** | 40.61 | 175.49 | 56.53 | 56.74 | 0.33 |
| *126* | 4.58 | 22.08 | **0.21** | 41.24 | 157.15 | 57.84 | 56.57 | 0.41 |
| *Grand Total* | 4.69 | 22.00 | **0.21** | 39.85 | 172.98 | 56.26 | 56.42 | 0.75 |

The detailed results for different refitting frequencies can be found in Tables 21-24 in Appendix.

## 6. Sensitivity analysis

In this section we check whether the strategy based on the base model yields the robust results. The tested parameters take different values while the others remain unchanged. For each tested parameter we visualize the equity lines and provide the performance tables of the base model against three benchmark strategies. The parameters of the base model are presented in Table 7.

**Table 7. Sensitivity analysis: testing parameters stability**

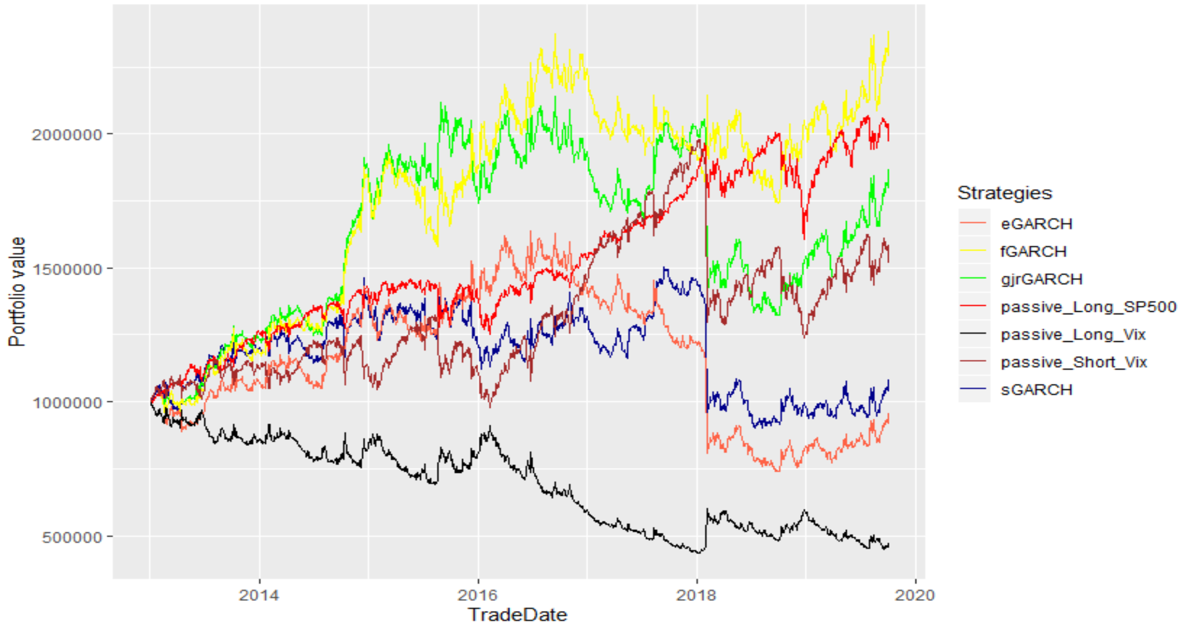
| *Specification* | *sGARCH* | *eGARCH* | *gjrGARCH* | *fGARCH/TGARCH sub-model* |
|---|---|---|---|---|
| *Training window size (days)* | 126 | <u>252</u> | 504 | 756 |
| *Refit every* | <u>21st</u> | 63rd | 126th | No refit |
| *Training window type* | <u>Moving</u> | Expanding | - | - |
| *Distributional model* | <u>Norm</u> | Skewed norm | Normal inverse Gaussian | - |
| *Historical volatility estimator* | <u>Close-to-Close</u> | Garman-Klass | Parkinson | - |

Note: This table summarizes all parameters tested in the sensitivity analysis. We underlined starting parameters for the base model

### *6.1. GARCH specification*

As expected, across different GARCH specifications the base model is not stable at all, and its performance varies significantly, see Table 8. Surprisingly, according to Figure 13 the fGARCH-TGARCH specification outperformed both the passive long S&P 500 and passive short VIX futures strategies and produced outstanding annualized return on the level of 13.53% against 10.74% and 6.63%, respectively. The GJR-GARCH did similarly well: 9.49% of annualized return being the most accurate specification in terms of sign change prediction. Once risk is considered, the passive long S&P 500 is still far the best strategy: 0.83 IR against 0.61 of fGARCH and 0.428 of GJR-GARCH, respectively. The exponential GARCH produced negative annualized return (-0.82%) with the highest percentage of long signals generated. Strategies based on EGARCH and GARCH models produced quite robust results though.

**Figure 13. Sensitivity analysis: equity lines for strategies with different GARCH specifications**




Note: fGARCH and gjrGARCH showed competitive outcomes in comparison to long S&P500 strategy, whereas sGARCH and eGARCH did not produced high risk-adjusted returns

**Table 8. Sensitivity analysis: strategy performance under different GARCH specifications**

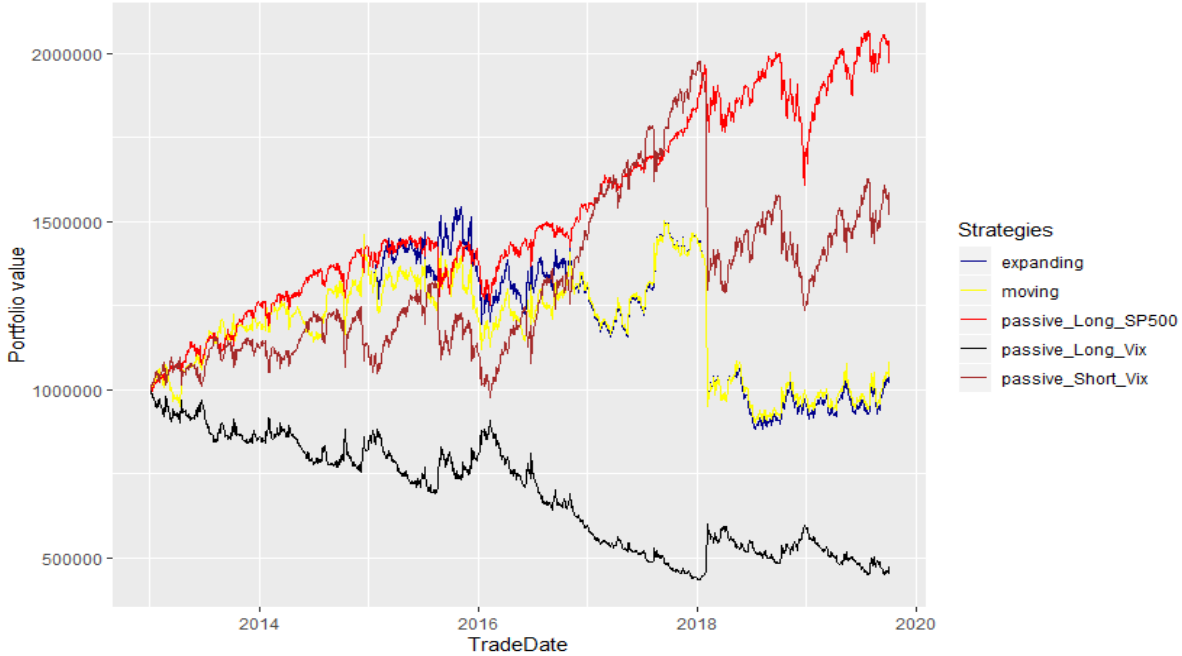
| *Strategy* | *Annualized Return, in %* | *Annualized SD, in %* | *Information Ratio* | *Maximum Drawdown, in %* | *Total number of trades* | *% of long signal to all signals* | *RMSE* | *Correct sign change prediction (ELF)* |
|---|---|---|---|---|---|---|---|---|
| *Passive long S&P 500* | 10.7 | 12.9 | **0.83** | 19.7 | 1 | 100 | - | - |
| *Passive long VIX* | -10.6 | 21.4 | **-0.50** | 56.7 | 1 | 100 | - | - |
| *Passive short VIX* | 6.6 | 22.1 | **0.30** | 37.4 | 1 | 0 | - | - |
| *SGARCH* | 1.0 | 22.0 | **0.04** | 39.9 | 159 | 57 | 0.232 | 53.7 |
| *EGARCH* | -0.8 | 22.0 | **-0.04** | 54.9 | 206 | 53.5 | 0.294 | 56.5 |
| *GJR-GARCH* | 9.4 | 22.1 | **0.49** | 38.2 | 184 | 55.8 | 0.287 | 57.5 |
| *fGARCH-TGARCH* | 13.5 | 22.1 | **0.61** | 28.8 | 214 | 55.1 | 0.308 | 57.5 |

Note: Except EGARCH all GARCH specifications produced positive IR, while passive long S&P500 strategy is the best in terms of IR metric primarily due to relatively low annualized SD and high annualized return

### *6.2. Training window type*

The obtained results let conclude that the strategy is robust to the window type. In the mid of 2016 the strategy based on the expanding window performed slightly better. The similar results were obtained for almost all performance statistics. In terms of annualized returns, Information Ratio and maximum drawdown, the “moving” window did slightly better, though. RMSE is identical for both, but “moving” strategy is little more accurate with respect to the correct sign change prediction. The Strategy with the expanding window is more active in terms of the total number of trades generated: 167 against 159 given the higher long signals proportion. There was a minor difference in the equity lines observed in the middle of 2016, which anyway disappeared later, according to Figure 14.

**Figure 14. Sensitivity analysis: equity lines for expanding and moving rolling windows**




Note: Quite similar equity line pattern with slight divergence in the mid of 2016

**Table 9. Sensitivity analysis: strategy performance under different training window type**

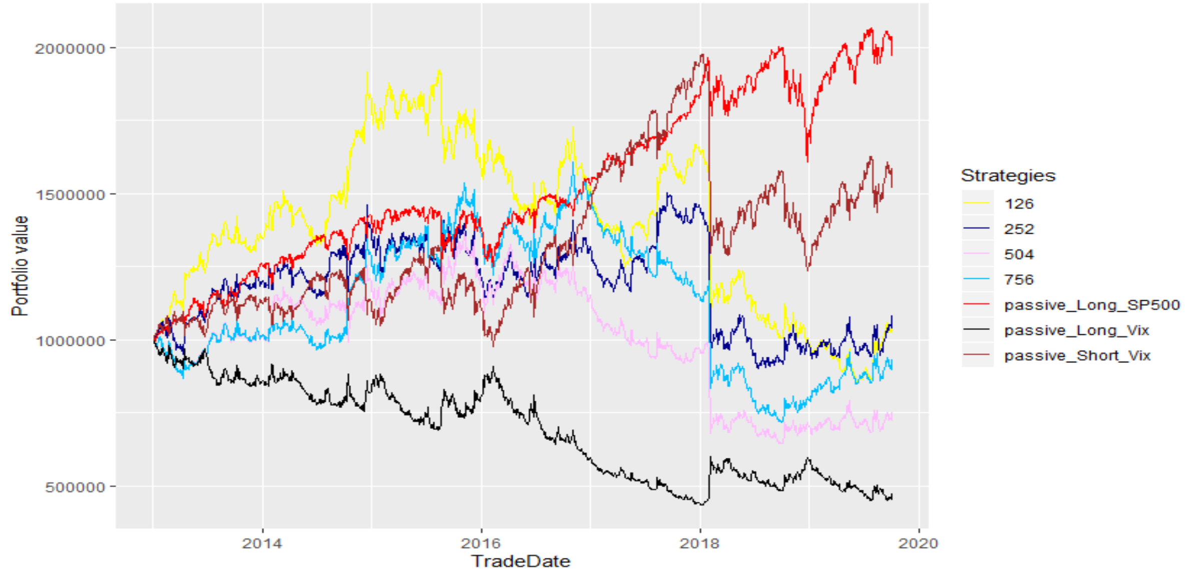
| *Strategy* | *Annualized Return, in %* | *Annualized SD, in %* | *Information Ratio* | *Maximum Drawdown, in %* | *Total number of trades* | *% of long signal to all signals* | *RMSE* | *Correct sign change prediction (ELF)* |
|---|---|---|---|---|---|---|---|---|
| *Passive long S&P 500* | 10.7 | 12.9 | **0.83** | 19.7 | 1 | 100 | - | - |
| *Passive long VIX* | -10.6 | 21.4 | **-0.500** | 56. | 1 | 100 | - | - |
| *Passive short VIX* | 6.6 | 22.2 | **0.30** | 37.4 | 1 | 0 | - | - |
| *Moving* | 1.00 | 22.0 | **0.04** | 39.9 | 159 | 57 | 0.232 | 53.7 |
| *Expanding* | 0.7 | 22.0 | **0.03** | 42.82 | 167 | 57.6 | 0.232 | 53.6 |

Note: Relatively robust training window outcomes according to all performance metrics and forecasting accuracy

### *6.3. Training window size*

Surprisingly, the longer training window resulted in worse annualized returns (which were negative ones but significantly better than the passive long VIX futures though) and larger maximum drawdowns despite being the relatively more accurate under two forecasting metrics, see Table 10 above. At the same time the modification with the smallest training window produced the largest maximum drawdown (56%), whereas one-year training window show near to the passive short VIX strategy maximum drawdown (40%). The trading volume of the strategy with training window of 126 days was the largest: 182 total trades in comparison to 159 by the strategy with 252 training window size. According to Information Ratio, the best trading strategy is the one with 252 training days, while the worst results are obtained for the strategy with 504 training days. We conclude that the strategy is not stable under training window size.

**Figure 15. Sensitivity analysis: equity lines across different training window size**




Note: The base model occurred to be the best across the other training window size strategies, but was still far away from passive long S&P500 and passive short VIX strategies

**Table 10. Sensitivity analysis: strategy performance under different training window size**

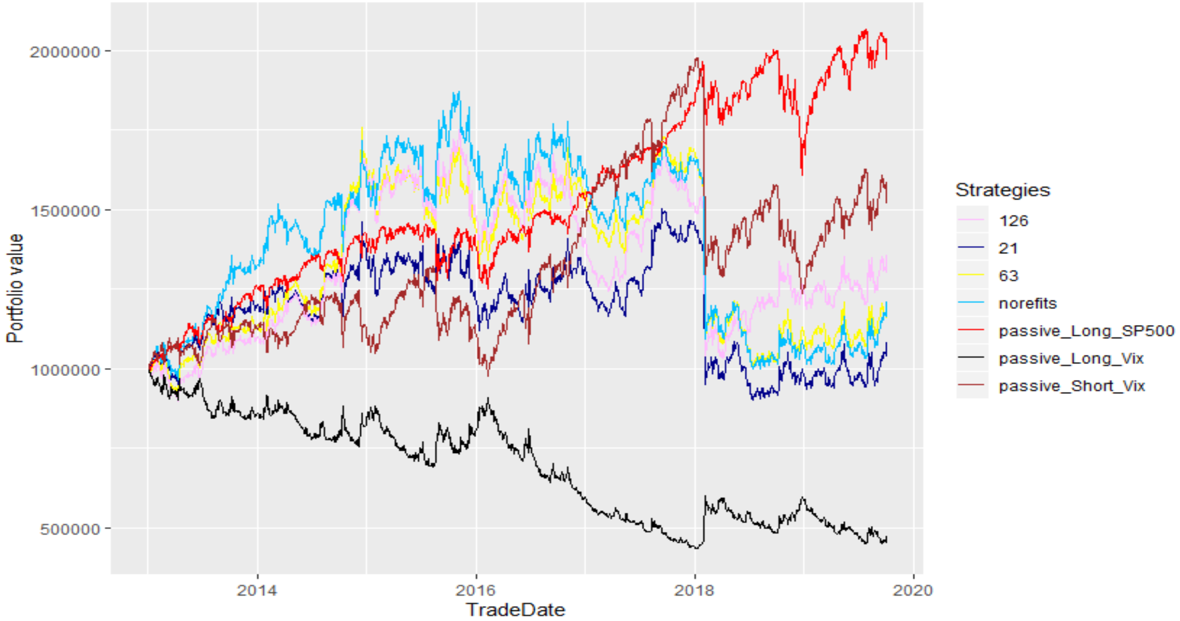
| *Strategy* | *Annualized Return, in %* | *Annualized SD, in %* | *Information Ratio* | *Maximum Drawdown, in %* | *Total number of trades* | *% of long signal to all signals* | *RMSE* | *Correct sign change prediction (ELF)* |
|---|---|---|---|---|---|---|---|---|
| *Passive long S&P 500* | 10.7 | 12.9 | **0.83** | 19.7 | 1 | 100 | - | - |
| *Passive long VIX* | -10.6 | 21.4 | **-0.50** | 56.7 | 1 | 100 | - | - |
| *Passive short VIX* | 6.6 | 22.1 | **0.30** | 37.4 | 1 | 0 | - | - |
| *126* | 0.7 | 22.0 | **0.03** | 55.7 | 182 | 53.1 | 0.263 | 51.9 |
| *252* | 1.00 | 22.0 | **0.04** | 39.9 | 159 | 57 | 0.226 | 54.3 |
| *504* | -4.3 | 21.8 | **-0.19** | 52.7 | 157 | 60.8 | 0.218 | 56.1 |
| *756* | -1.2 | 22.0 | **-0.05** | 55.3 | 143 | 62.8 | 0.207 | 56.8 |

Note: 252-strategy showed only 39.914% of maximum drawdown, which is significantly better than the other GARCH models

### *6.4. Refitting frequency*

Surprisingly, the model that has been refitted the most frequently, ended up as the worst one from the Information Ratio perspective among its less frequently refitted peers despite having the lowest RMSE (0.226). Except annualized returns, the performance metrics are very similar, and none of the strategies outperformed neither the passive long S&P 500 nor the passive short VIX futures strategies. Therefore, we can say that our initial guesses with regards to refitting frequency have not been justified.

**Figure 16. Sensitivity analysis: equity lines across different model refitting frequencies**




Note: The 126-strategy performed better that those with the other GARCH models, but still was far behind passive S&P500 and VIX long and short strategies, respectively.

**Table 11. Sensitivity analysis: investment performance across different refitting frequency**

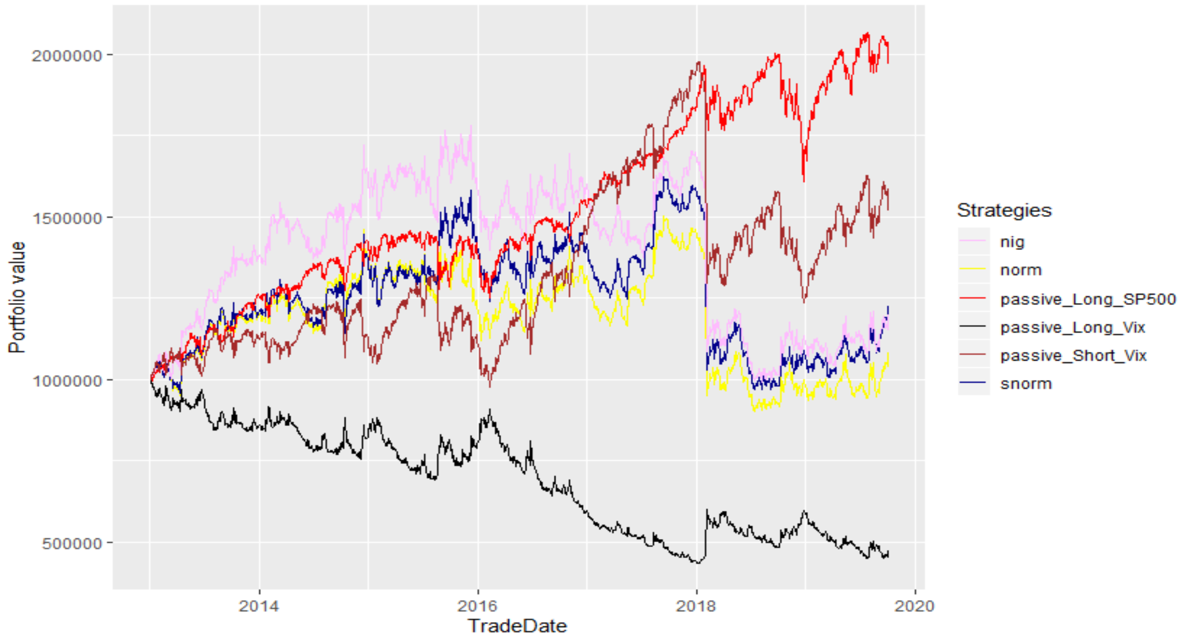
| *Strategy* | *Annualized Return, in %* | *Annualized SD, in %* | *Information Ratio* | *Maximum Drawdown, in %* | *Total number of trades* | *% of long signal to all signals* | *RMSE* | *Correct sign change prediction (ELF)* |
|---|---|---|---|---|---|---|---|---|
| *Passive long S&P 500* | 10.7 | 12.9 | **0.83** | 19.7 | 1 | 100 | - | - |
| *Passive long VIX* | -10.6 | 21.4 | **-0.50** | 56.7 | 1 | 100 | - | - |
| *Passive short VIX* | 6.6 | 22.1 | **0.30** | 37.4 | 1 | 0 | - | - |
| *21* | 1.00 | 22.0 | **0.04** | 39.9 | 159 | 57 | 0.226 | 54.3 |
| *63* | 2.6 | 22.0 | **0.12** | 42.7 | 165 | 58.5 | 0.269 | 55.0 |
| *126* | 4.4 | 22.0 | **0.20** | 41.0 | 153 | 58.1 | 0.278 | 54.8 |
| *no refit* | 2.6 | 22.0 | **0.12** | 46.7 | 170 | 57 | 0.239 | 54.1 |

Note: More often model refitting does not improve Information Ratio

*6.5. Distribution model*

The equity line pattern started to stabilize since the beginning of 2018 just post the significant portfolio value decline, see Figure 17. As expected, the normal inverse gaussian distribution improved the overall forecasting errors, however it performed slightly worse that the model with skewed-normal distribution: 2.5% of annualized returns against 2.86% - still far below the short VIX and long S&P 500 passive strategies. One of the reasons why these two outperformed the model with the normal distribution might be that they take into the account the volatility stylized facts.

**Figure 17. Sensitivity analysis: equity lines across different assumptions about conditional distribution of returns in the model.**




Note: Similar equity line pattern observed across all GARCH model distributions

**Table 12. Sensitivity analysis: investment performance across different assumptions about conditional distribution of returns in the model.**

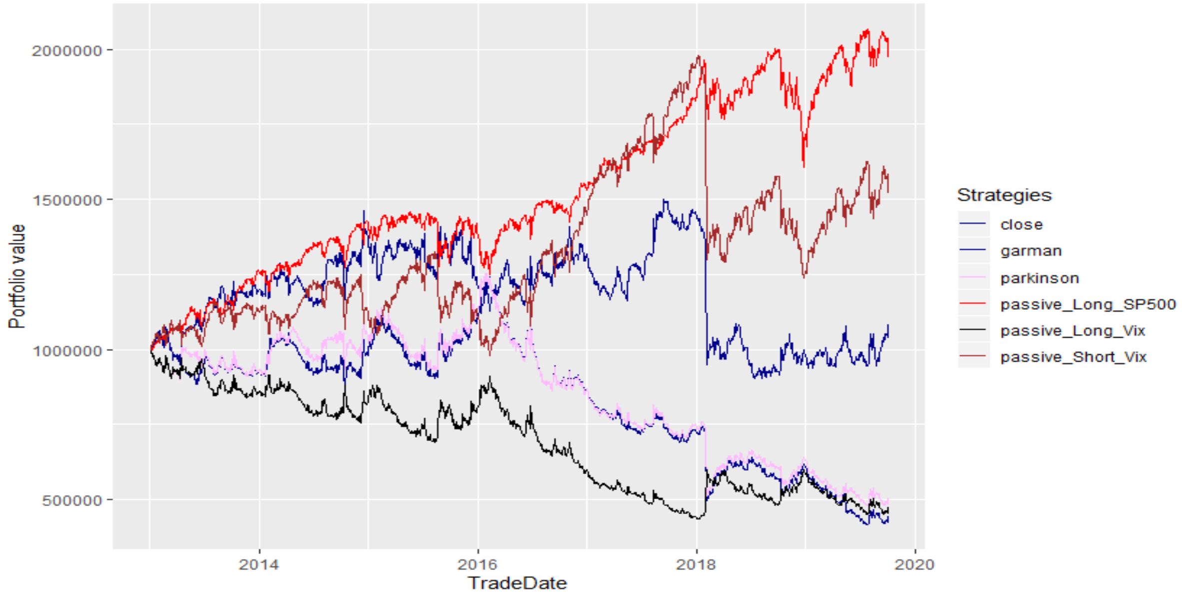
| *Strategy* | *Annualized Return, in %* | *Annualized SD, in %* | *Information Ratio* | *Maximum Drawdown, in %* | *Total number of trades* | *% of long signal to all signals* | *RMSE* | *Correct sign change prediction (ELF)* |
|---|---|---|---|---|---|---|---|---|
| *Passive long S&P 500* | 10.7 | 12.9 | **0.83** | 19.7 | 1 | 100 | - | - |
| *Passive long VIX* | -10.6 | 21.4 | **-0.50** | 56.7 | 1 | 100 | - | - |
| *Passive short VIX* | 6.6 | 22.1 | **0.30** | 37.4 | 1 | 0 | - | - |
| *Norm* | 1.00 | 22.0 | **0.04** | 39.9 | 159 | 57 | 0.226 | 54.3 |
| *Snorm* | 2.8 | 22.1 | **0.13** | 40.4 | 151 | 55.1 | 0.217 | 55.2 |
| *Nig* | 2.5 | 22.3 | **0.11** | 44.2 | 155 | 56.6 | 0.212 | 56.0 |

Note: The strategies with snorm and nig model distributions produced more than two times higher IR outcomes in comparison to the strategy based on normal distribution

### *6.6. Historical volatility estimator*

Performance of the strategies based on the Garman-Klass and the Parkinson estimators are very similar, and the trading results are significantly worse: -11.5% and -9.8% of annualized returns, respectively, in comparison with 1.0% by the "Close-to-Close" estimator. There is not much left from the initial capital invested at the end of the trading period given the notorious drawdowns: 61.7% by Parkinson and 66.5% for Garman. The strategies also did poorly in terms of correct sign change predictions 50.9% against 54.3% of the base case. To investigate it further we plot the daily estimated historical volatilities (Figure 19).

**Figure 18. Sensitivity analysis: equity lines under different historical volatility estimators**




Note: The strategies based on Garman-Klass and Parkinson estimators are the worst and initial capital invested eroded quickly as in a case of passive short VIX strategy

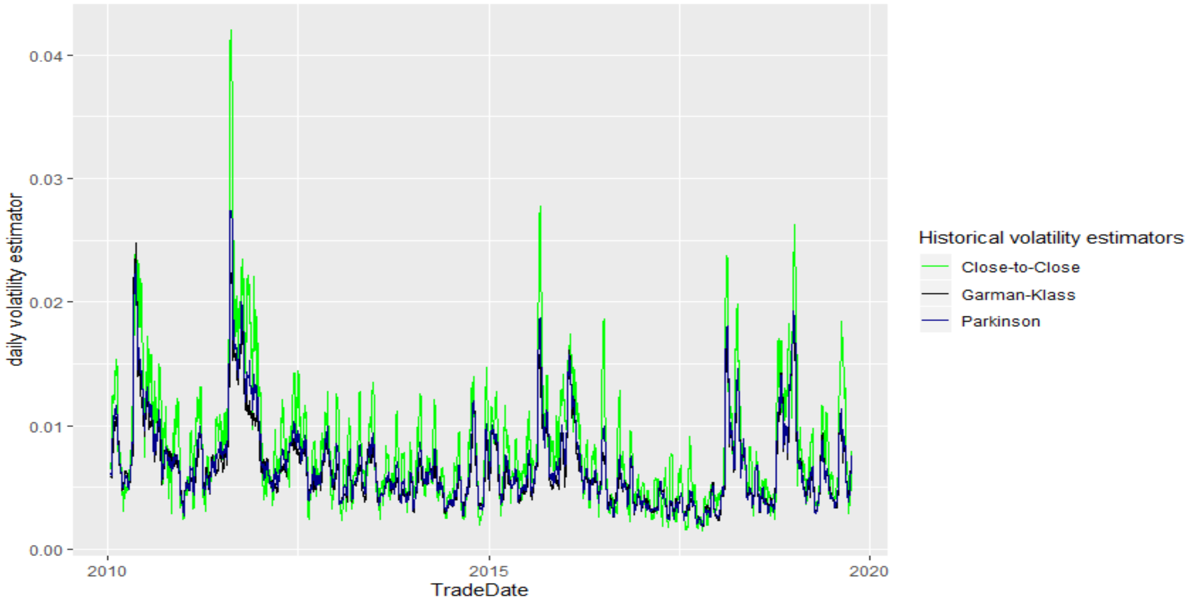
**Table 13. Sensitivity analysis: investment performance across different volatility estimators**

| *Strategy* | *Annualized Return, in %* | *Annualized SD, in %* | *Information Ratio* | *Maximum Drawdown, in %* | *Total number of trades* | *% of long signal to all signals* | *RMSE* | *Correct sign change prediction (ELF)* |
|---|---|---|---|---|---|---|---|---|
| *Passive long S&P 500* | 10.7 | 12.9 | **0.83** | 19.7 | 1 | 100 | - | - |
| *Passive long VIX* | -10.6 | 21.4 | **-0.50** | 56.7 | 1 | 100 | - | - |
| *Passive short VIX* | 6.6 | 22.1 | **0.30** | 37.4 | 1 | 0 | - | - |
| *Close-to-Close* | 1.00 | 22.0 | **0.04** | 39.9 | 159 | 57 | 0.226 | 54.3 |
| *Garman-Klass* | -11.5 | 21.93 | **-0.52** | 66.5 | 89 | 88.3 | 0.288 | 51.2 |
| *Parkinson* | -9.7 | 21.8 | **-0.44** | 61.7 | 91 | 85.2 | 0.257 | 50.9 |

Note: Garman-Klass based strategy is the worst according to IR, whereas Parkinson strategy is slightly better than passive long VIX strategy. Both strategies generated almost two times less signals than the Close-to-Close strategy

The Parkinson and Garman-Klass estimators in 75% of all trading days are lower than the Close-to-Close estimator generating much more long signals (see Figure 19). Thus, the Parkinson and the Garman-Klass estimators generated almost two times less trading signals (most long ones) 91, 89, respectively, in comparison to 159 by the Close-to-Close estimator. This explains the relatively poor performance of strategies based on the Parkinson and the Garman-Klass estimators, the latter falling even below the passive long VIX strategy.

**Figure 19. Daily historical volatility across different volatility estimators**




Note: Garman-Klass and Parkinson generate significantly more buy signals almost replicating passive long VIX strategy

## Conclusions

In the presented study we compared the performance of VIX futures trading strategies based on different GARCH model specifications. We used the daily observations of the S&P 500 index and VIX futures (contracts) from the period of 2009-01-01 to 2019-10-03, and 2013-01-02 to 2019-10-03, respectively. The rolling volatility forecasting techniques were applied across the different GARCH model specifications.

Referring to the main hypothesis we can say that we were not able to obtain robust abnormal returns with comparison to the equity benchmark strategies. Further empirical findings support this statement. We found out that (see Table 8) across four GARCH model specifications considered, the strategy based on the threshold GARCH (fGARCH – TGARCH extension) was the most attractive one producing the highest value of Information Ratio, the highest annualized returns and the lowest maximum drawdown. Interestingly, the classical

GARCH model had the best predicting power under RMSE – much more accurate than fGARCH – TGARCH and GJR-GARCH. The more frequent model refitting did not improve portfolio's Information Ratio – not as it was initially expected. Regarding the size of the training window, we were unable to conclude that the longer or shorter one necessarily improves or diminishes Information Ratio. In our research based on the data used we obtained that there is no direct relationship – and the optimal training window size is within 126 and 252 trading days range. The performance of strategies under different volatility estimators differ considerably. The poor performance of the Garman-Klass and Parkinson estimators might be partially explained by relatively higher number of long signals generated during the overall seven-year downward volatility trend observed.

The further research ideas which could be conducted within the scope of the presented study and result in the overall subject benefit are as follows. It would be interesting to see how performance would change if weekly expiries VIX futures had been also included in the investment strategy. Given there are turbulent periods in which VIX close prices change dramatically and all strategies have maximum drawdown exceeding 30%, implementing the tool (like stop-loss) which prevents such losses seems to be a reasonable idea. In addition to that it would be worth to account for the pattern of VIX volatility term structure and adjust the strategy accordingly, especially when the positions are rolled out. To sum up, although there is a space for improvement for presented GARCH based rolling strategies, they showed their potential in competing with the benchmark.

## References


Akgiray, V., 1989, *Conditional heteroskedasticity in time series of stock returns: evidence and forecasts*, Journal of Business, 62 (1), 55-80.

Baillie, R.T., 1996, *Long memory processes and fractional integration in econometrics*, Journal of Econometrics, Volume 73, Issue 1, July 1996, Pages 5-59.

Balaban E., Bayar A., 2005, *Stock returns and volatility: empirical evidence from fourteen countries*, Applied Economics Letters, Vol. 12, Issue 10.

Bollerslev T., 1986, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics, 31 (3), 307-327.

Brooks C., 2008, *Introductory econometrics for finance*, Cambridge University Press. Second Edition, University of Reading.

Brownlees C., Engle R., Kelly B., 2011, *A Practical guide to volatility forecasting through calm and storm*, Available at SSRN: https://ssrn.com/abstract=1502915.

Casas M., Cepeda E., 2008, *ARCH, GARCH and EGARCH models: Applications to financial series*, Cuadernos de Economia, Vol. 27, No. 48.

Christoffersen P., Feunou B., Jacobs K., Meddahi N., 2014, *The economic value of realized volatility: using high-frequency returns for option valuation*, Journal of Financial and Quantitative Analysis, 663–697.

Clark T.E., McCracken W., 2009, *Improving forecast accuracy by combining recursive and rolling forecasts*, International Economic Review, Vol. 50, Issue 2, 363-395.

Ding Z., Granger C.W.J., Engle R.F., 1993, *A long memory property of stock market returns and a new model*, Journal of Empirical Finance, Vol. 1, Issue 1, June 1993, Pages 83-106.

Ebeid S., Alkholi B., Gamal B. A., 2004, *Volatility modeling and forecasting of the Egyptian stock market index using ARCH models*, Available at SSRN: https://ssrn.com/abstract=631887.

Ederington L., Guan W., 2004, *Forecasting volatility*. Available at SSRN: https://ssrn.com/abstract=165528 or http://dx.doi.org/10.2139/ssrn.165528.

Engle R., Bollerslev T, 1986, *Modelling the persistence of conditional variances*, Econometric Reviews, 5 (1), 1-50.

Engle R.F., Mustafa C., 1992, *Implied ARCH models from option prices*, Journal of Econometrics, Vol. 52, Issues 1-2, April-May, Pages 289-311.

Engle R.F., Patton Andrew J., 2007, *What is a good vitality model? Forecasting volatility in the Financial Markets (Third Edition)*, Quantitative Finance, 47-63.

Franses P.H., Van Dijk D., 1996, *Forecasting stock market volatility using (non-linear) GARCH models*, Journal of Forecasting, Vol. 15, Issue 3, 229-235.

Frennberg P., Hansson B., 1996, *An evaluation of alternative models for predicting stock volatility*, Journal of International Financial Markets, Institutions and Money, 5, 117-134.

Garman, M. B., Klass M.J., 1980, *On the estimation of security price volatilities from historical data*, The Journal of Business, 53 (1), 67.

Geoffrey F. Loudon, Wing H. Watt, Pradeep K. Yadav, 2000. An Empirical analysis of alternative parametric ARCH models. *Journal of Applied Econometrics*, Vol. 15, No. 2, 117-136.

Gerlow M. E., Irwin S. H., Liu T.-R., 1993., *Economic Evaluation of Commodity Price Forecasting Models*, International Journal of Forecasting, 9, 387-97.

Ghalanos A., 2017, *Introduction to the rugarch package*, Version 1.3-8.

Glosten R. L., Jagannathan R., Runkle D., 1993, *On the relation between the expected value and volatility of the nominal excess returns on stocks*, Journal of Finance, Vol. 48(5), 1779-1801.

Hansen P., 2005, *A forecast comparison of volatility models: does anything beat GARCH (1,1)*? Journal of Applied Econometrics, Vol. 20, Issue 7, 873-889.

Hansen P., Huang Z., 2016, *Exponential GARCH modelling with realized measures of volatility*, Journal of Business & Economic Statistics, Vol. 34, Issue 2, 269-287.

Hentschel L., 1995, *All in the family nesting symmetric and asymmetric GARCH models*, Journal of Financial Economics, 39(1), 71-104.

Jabłecki J., Kokoszczyński R., Sakowski P., Ślepaczuk R., Wójcik P., 2012, *Pomiar i modelowanie zmienności – przegląd literatury*, Ekonomia. Rynek, Gospodarka, Społeczeństwo, 31, 22-55.

Jabłecki J., Kokoszczyński R., Sakowski P., Ślepaczuk R., Wójcik P., 2015, *Volatility as an Asset Class, Obvious Benefits and Hidden Risk*, Frankfurt, PeterLang, 1-231.

Kość K., Sakowski P., Ślepaczuk R., 2019, *Momentum and Contrarian effects on the cryptocurrency market*, Physica A. Statistical Mechanics and its Applications, 523, 691-701.

Laurent, S., Rombouts, J. V., & Violante, F., 2012, *On the forecasting accuracy of multivariate GARCH models*, Journal of Applied Econometrics, 27(6), 934-955.

Loudon F., Watt W. H., Yadav P., 2000, *An empirical analysis of alternative parametric ARCH models*, Journal of Applied Econometrics, Vol. 15, Issue 2, 117-136.

Leitch G., Tanner J. E., 1991, *Economic forecast evaluation: profit versus the conventional error measures*, American Economic Review, 81 (3), 580-90.

Majmudar U., Banerjee A., 2004, *VIX Forecasting*. Available at SSRN: https://ssrn.com/abstract=533583 or http://dx.doi.org/10.2139/ssrn.533583.

Markowitz H., 1952. Portfolio Selection. *Journal of Finance*, Vol. 7, No. 1.

Masset P., 2011, *Volatility stylized facts.* Available at SSRN: https://ssrn.com/abstract=1804070 or http://dx.doi.org/10.2139/ssrn.1804070.

Nelson D.B., 1991, *Conditional heteroskedasticity in asset returns: A new approach*, Econometrica, 59(2), 347-70.

Parkinson M., 1980, *The extreme value method for estimating the variance of the rate of return*, Journal of Business, 53 (1), 61-65.

Poon S., Granger C. W.J., 2003, *Forecasting volatility in financial markets: A Review*, Journal of Economic Literature, Vol. 41, No. 2.

Ryś P. Ślepaczuk R., 2019, *Machine Learning in algorithmic trading strategy optimization – implementation and efficiency*, Central European Economic Journal, 5 (1), 206-229.

Tsay R., 2002, *Analysis of financial time series*, John Wiley & Sons, inc. Wiley series in probability and statistics. Financial engineering section.

West K. D., Cho D., 1995, *The predictive ability of several models of exchange rate volatility*, Journal of Econometrics, 69, 367-91.

Xie H., Yu C., 2019, *Realized GARCH models: simpler is better*, Finance Research Letters – Available online 2 July 2019.

Ślepaczuk R., Zakrzewski G., 2013, *High-frequency and model-free volatility estimators*, LAP LAMBERT Academic Publishing.

Ślepaczuk R., Zakrzewski G., Sakowski P., 2018, *Investment strategies that beat the market. What can we squeeze from the market*, e-Finanse 14 (4), 36-55.

Zakoian J.M., 1994. *Threshold heteroskedastic models*, Journal of Economic Dynamics and Control, 18 (5), 931-955.

Zenkova M., Ślepaczuk R., *Robustness of Support Vector Machines in Algorithmic Trading on Cryptocurrency Market*, Central European Economic Journal 5 (1), 186-205.

window with 126 days

# APPENDIX

## List of tables

Table 14. The analysis of the performance of 432 GARCH model described in detail in Section 4.2: Model specification – GARCH model

| *Strategy* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of elf* | *Average of rmse* |
|---|---|---|---|---|---|---|---|---|
| ***126*** | **-1.28** | **20.82** | **-0.06** | **52.84** | **147.78** | **52.09** | **49.21** | **0.31** |
| ***expanding*** | **-1.81** | **19.57** | **-0.08** | **50.09** | **134.22** | **48.61** | **46.44** | **0.29** |
| *21* | -0.89 | 14.74 | -0.04 | 36.19 | 116.00 | 34.90 | 34.41 | 0.18 |
| *63* | -3.54 | 21.99 | -0.16 | 60.55 | 162.33 | 54.33 | 52.20 | 0.34 |
| *126* | -1.01 | 21.98 | -0.05 | 53.53 | 124.33 | 56.60 | 52.70 | 0.37 |
| ***moving*** | **-0.75** | **22.07** | **-0.03** | **55.59** | **161.33** | **55.57** | **51.99** | **0.32** |
| *21* | -0.32 | 22.16 | -0.01 | 54.91 | 184.00 | 53.93 | 51.72 | 0.27 |
| *63* | -0.94 | 22.14 | -0.04 | 56.07 | 176.33 | 54.93 | 52.18 | 0.34 |
| *126* | -0.97 | 21.92 | -0.05 | 55.80 | 123.67 | 57.83 | 52.08 | 0.36 |
| ***252*** | **1.18** | **22.08** | **0.05** | **45.47** | **153.00** | **57.94** | **54.81** | **0.26** |
| ***expanding*** | **1.29** | **22.08** | **0.06** | **44.67** | **154.56** | **57.49** | **54.72** | **0.26** |
| *21* | 1.05 | 22.06 | 0.05 | 41.97 | 160.33 | 56.43 | 54.56 | 0.22 |
| *63* | -0.76 | 22.02 | -0.03 | 48.91 | 160.33 | 57.83 | 54.78 | 0.27 |
| *126* | 3.58 | 22.16 | 0.16 | 43.13 | 143.00 | 58.20 | 54.81 | 0.29 |
| ***moving*** | **1.07** | **22.07** | **0.05** | **46.27** | **151.44** | **58.39** | **54.90** | **0.26** |
| *21* | 2.12 | 22.15 | 0.10 | 41.51 | 155.00 | 56.23 | 54.78 | 0.22 |
| *63* | -0.18 | 22.06 | -0.01 | 49.88 | 157.00 | 58.50 | 54.90 | 0.27 |
| *126* | 1.25 | 22.01 | 0.06 | 47.44 | 142.33 | 60.43 | 55.01 | 0.28 |
| ***504*** | **-0.55** | **22.00** | **-0.03** | **45.92** | **155.22** | **60.27** | **56.58** | **0.21** |
| ***expanding*** | **0.09** | **21.99** | **0.00** | **45.74** | **153.89** | **60.11** | **56.56** | **0.21** |
| *21* | 0.65 | 22.08 | 0.03 | 43.85 | 162.33 | 59.93 | 56.85 | 0.21 |
| *63* | -0.58 | 21.93 | -0.03 | 47.63 | 156.33 | 59.93 | 56.72 | 0.21 |
| *126* | 0.19 | 21.97 | 0.01 | 45.75 | 143.00 | 60.47 | 56.10 | 0.21 |
| ***moving*** | **-1.19** | **22.01** | **-0.05** | **46.11** | **156.56** | **60.43** | **56.60** | **0.21** |
| *21* | -1.29 | 21.99 | -0.06 | 44.65 | 163.67 | 60.07 | 56.85 | 0.21 |
| *63* | -2.16 | 22.08 | -0.10 | 50.06 | 159.67 | 60.30 | 56.78 | 0.21 |
| *126* | -0.11 | 21.97 | -0.01 | 43.63 | 146.33 | 60.93 | 56.19 | 0.22 |
| ***Grand Total*** | **-0.22** | **21.63** | **-0.01** | **48.08** | **152.00** | **56.77** | **53.53** | **0.26** |

Note: For GARCH model on average the strategies with 252 training window days (expanding) performed relatively better (IR = 0.06) than the other strategies

Table 15. The analysis of the performance of 432 GARCH model described in detail in Section 4.2: Model specification – fGARCH (TGARCH extension) model

| *Strategy* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of elf* | *Average of rmse* |
|---|---|---|---|---|---|---|---|---|
| ***126*** | **7.67** | **22.14** | **0.35** | **31.85** | **214.22** | **56.75** | **57.62** | **0.33** |
| ***expanding*** | **9.03** | **22.17** | **0.41** | **29.49** | **213.56** | **57.32** | **57.76** | **0.32** |
| *21* | 8.59 | 22.17 | 0.39 | 30.15 | 238.00 | 55.53 | 58.14 | 0.29 |
| *63* | 9.78 | 22.14 | 0.44 | 26.39 | 218.00 | 56.43 | 57.90 | 0.32 |
| *126* | 8.71 | 22.20 | 0.39 | 31.92 | 184.67 | 60.00 | 57.24 | 0.35 |
| ***moving*** | **6.31** | **22.12** | **0.28** | **34.21** | **214.89** | **56.18** | **57.48** | **0.33** |
| *21* | 5.89 | 22.10 | 0.27 | 30.40 | 238.67 | 55.70 | 57.62 | 0.28 |
| *63* | 8.00 | 22.11 | 0.36 | 29.20 | 220.67 | 55.07 | 57.56 | 0.36 |
| *126* | 5.03 | 22.14 | 0.23 | 43.03 | 185.33 | 57.77 | 57.26 | 0.34 |
| ***252*** | **11.52** | **22.20** | **0.52** | **32.25** | **190.67** | **55.12** | **58.03** | **0.31** |
| ***expanding*** | **11.50** | **22.19** | **0.52** | **31.96** | **190.11** | **55.23** | **58.05** | **0.32** |
| *21* | 14.73 | 22.27 | 0.66 | 27.76 | 201.33 | 54.87 | 58.37 | 0.30 |
| *63* | 10.28 | 22.17 | 0.46 | 32.44 | 195.00 | 55.30 | 58.02 | 0.32 |
| *126* | 9.49 | 22.12 | 0.43 | 35.69 | 174.00 | 55.53 | 57.75 | 0.33 |
| ***moving*** | **11.53** | **22.21** | **0.52** | **32.54** | **191.22** | **55.01** | **58.01** | **0.31** |
| *21* | 14.35 | 22.22 | 0.65 | 29.21 | 201.33 | 54.83 | 58.31 | 0.30 |
| *63* | 11.54 | 22.23 | 0.52 | 31.60 | 190.33 | 54.73 | 58.21 | 0.31 |
| *126* | 8.72 | 22.19 | 0.39 | 36.82 | 182.00 | 55.47 | 57.51 | 0.32 |
| ***504*** | **8.27** | **22.16** | **0.37** | **29.52** | **184.67** | **56.83** | **58.39** | **0.27** |
| ***expanding*** | **8.04** | **22.16** | **0.36** | **29.48** | **184.00** | **56.87** | **58.31** | **0.27** |
| *21* | 7.21 | 22.16 | 0.33 | 29.19 | 193.33 | 56.83 | 58.09 | 0.27 |
| *63* | 9.24 | 22.16 | 0.42 | 28.91 | 176.33 | 57.20 | 58.09 | 0.27 |
| *126* | 7.66 | 22.17 | 0.35 | 30.33 | 182.33 | 56.57 | 58.75 | 0.27 |
| ***moving*** | **8.50** | **22.16** | **0.38** | **29.57** | **185.33** | **56.79** | **58.46** | **0.27** |
| *21* | 8.72 | 22.14 | 0.39 | 29.13 | 193.33 | 56.77 | 58.39 | 0.26 |
| *63* | 9.13 | 22.21 | 0.41 | 29.15 | 181.00 | 56.73 | 58.48 | 0.27 |
| *126* | 7.64 | 22.13 | 0.35 | 30.43 | 181.67 | 56.87 | 58.53 | 0.27 |
| ***Grand Total*** | **9.15** | **22.17** | **0.41** | **31.21** | **196.52** | **56.23** | **58.01** | **0.30** |

Note: For fGARCH – TGARCH the best IR (0.52) is obtained on 252 window training size (similarly for expanding and moving training window types)

Table 16. The analysis of the performance of 432 GARCH model described in detail in Section 4.2: Model specifications – EGARCH model

| *Strategy* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of elf* | *Average of rmse* |
|---|---|---|---|---|---|---|---|---|
| ***126*** | **4.93** | **22.02** | **0.22** | **36.98** | **171.33** | **55.11** | **55.21** | **0.31** |
| ***expanding*** | **3.15** | **22.01** | **0.14** | **39.89** | **175.78** | **56.21** | **55.20** | **0.31** |
| *21* | 2.99 | 22.00 | 0.13 | 46.95 | 212.67 | 50.57 | 55.20 | 0.26 |
| *63* | 4.40 | 22.11 | 0.20 | 33.87 | 164.33 | 58.30 | 54.95 | 0.31 |
| *126* | 2.07 | 21.90 | 0.09 | 38.84 | 150.33 | 59.77 | 55.44 | 0.38 |
| ***moving*** | **6.70** | **22.04** | **0.30** | **34.07** | **166.89** | **54.01** | **55.23** | **0.32** |
| *21* | 6.50 | 22.05 | 0.29 | 31.94 | 200.67 | 50.20 | 55.46 | 0.22 |
| *63* | 8.87 | 22.11 | 0.40 | 31.47 | 162.67 | 54.30 | 54.73 | 0.37 |
| *126* | 4.75 | 21.96 | 0.21 | 38.80 | 137.33 | 57.53 | 55.50 | 0.36 |
| ***252*** | **-3.98** | **21.94** | **-0.18** | **58.10** | **173.56** | **57.28** | **56.36** | **0.46** |
| ***expanding*** | **-4.61** | **21.90** | **-0.21** | **60.39** | **175.78** | **57.29** | **56.40** | **0.46** |
| *21* | -1.86 | 22.01 | -0.08 | 56.35 | 194.67 | 53.93 | 56.31 | 0.39 |
| *63* | -5.97 | 21.84 | -0.27 | 62.12 | 182.00 | 57.63 | 56.79 | 0.46 |
| *126* | -6.01 | 21.86 | -0.28 | 62.68 | 150.67 | 60.30 | 56.10 | 0.1 |
| ***moving*** | **-3.35** | **21.97** | **-0.15** | **55.81** | **171.33** | **57.28** | **56.33** | **0.46** |
| *21* | -2.40 | 21.94 | -0.11 | 54.90 | 200.67 | 54.10 | 56.85 | 0.39 |
| *63* | -5.27 | 21.91 | -0.24 | 58.83 | 175.33 | 57.87 | 56.42 | 0.46 |
| *126* | -2.39 | 22.05 | -0.11 | 53.68 | 138.00 | 59.87 | 55.71 | 0.52 |
| ***504*** | **4.41** | **22.06** | **0.20** | **33.80** | **185.44** | **54.72** | **57.92** | **6.84** |
| ***expanding*** | **4.49** | **22.07** | **0.20** | **33.27** | **180.89** | **54.56** | **57.83** | **1.23** |
| *21* | 6.14 | 22.09 | 0.28 | 27.14 | 186.00 | 54.13 | 57.68 | 0.28 |
| *63* | 3.34 | 22.01 | 0.15 | 37.53 | 178.00 | 55.17 | 57.94 | 0.8 |
| *126* | 3.98 | 22.11 | 0.18 | 35.12 | 178.67 | 54.37 | 57.89 | 2.61 |
| ***moving*** | **4.33** | **22.04** | **0.20** | **34.34** | **190.00** | **54.89** | **58.00** | **12.45** |
| *21* | 5.58 | 22.14 | 0.25 | 31.20 | 193.33 | 55.50 | 57.87 | 36.6 |
| *63* | 2.62 | 21.97 | 0.12 | 39.38 | 193.33 | 54.80 | 58.22 | 0.35 |
| *126* | 4.79 | 22.01 | 0.22 | 32.45 | 183.33 | 54.37 | 57.90 | 0.4 |
| ***Grand Total*** | **1.78** | **22.01** | **0.08** | **42.96** | **176.78** | **55.71** | **56.50** | **2.54** |

Note: For EGARCH the strategy based on moving training window with 126 training window days produced on average the best IR (0.3)

Table 17. The analysis of the performance of 432 GARCH model combinations described in detail in Section 4.2: Model specifications – GJR-GARCH model

| *Strategy* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of elf* | *Average of rmse* |
|---|---|---|---|---|---|---|---|---|
| ***126*** | **6.92** | **22.14** | **0.31** | **36.54** | **175.89** | **53.91** | **56.95** | **0.33** |
| ***expanding*** | **6.87** | **22.16** | **0.31** | **37.26** | **172.22** | **54.06** | **56.94** | **0.32** |
| *21* | 4.85 | 22.15 | 0.22 | 37.24 | 198.67 | 51.27 | 56.86 | 0.28 |
| *63* | 7.50 | 22.16 | 0.34 | 38.74 | 168.67 | 54.27 | 57.36 | 0.32 |
| *126* | 8.26 | 22.15 | 0.37 | 35.81 | 149.33 | 56.63 | 56.60 | 0.37 |
| ***moving*** | **6.98** | **22.12** | **0.31** | **35.81** | **179.56** | **53.76** | **56.96** | **0.33** |
| *21* | 1.33 | 22.06 | 0.06 | 39.37 | 206.67 | 50.53 | 56.42 | 0.29 |
| *63* | 11.52 | 22.21 | 0.52 | 31.83 | 176.00 | 53.53 | 57.39 | 0.36 |
| *126* | 8.08 | 22.09 | 0.37 | 36.22 | 156.00 | 57.20 | 57.07 | 0.35 |
| ***252*** | **8.85** | **22.24** | **0.40** | **44.32** | **165.78** | **55.23** | **57.84** | **0.28** |
| ***expanding*** | **8.56** | **22.22** | **0.38** | **44.44** | **164.67** | **55.11** | **57.77** | **0.29** |
| *21* | 10.56 | 22.27 | 0.47 | 40.49 | 172.00 | 55.17 | 58.06 | 0.28 |
| *63* | 7.69 | 22.24 | 0.35 | 43.68 | 169.33 | 54.83 | 57.77 | 0.28 |
| *126* | 7.43 | 22.16 | 0.33 | 49.16 | 152.67 | 55.33 | 57.48 | 0.30 |
| ***moving*** | **9.14** | **22.26** | **0.41** | **44.19** | **166.89** | **55.36** | **57.91** | **0.28** |
| *21* | 9.87 | 22.26 | 0.44 | 40.07 | 176.00 | 55.03 | 58.00 | 0.28 |
| *63* | 8.71 | 22.28 | 0.39 | 44.61 | 170.00 | 55.53 | 57.89 | 0.28 |
| *126* | 8.83 | 22.23 | 0.40 | 47.90 | 154.67 | 55.50 | 57.86 | 0.28 |
| ***504*** | **8.32** | **22.16** | **0.38** | **30.56** | **158.22** | **59.86** | **58.12** | **0.27** |
| ***expanding*** | **9.36** | **22.15** | **0.42** | **26.71** | **158.89** | **59.70** | **58.23** | **0.27** |
| *21* | 8.83 | 22.12 | 0.40 | 25.38 | 160.67 | 59.37 | 58.16 | 0.27 |
| *63* | 8.09 | 22.11 | 0.37 | 27.69 | 160.00 | 59.73 | 58.29 | 0.27 |
| *126* | 11.17 | 22.22 | 0.50 | 27.05 | 156.00 | 60.00 | 58.23 | 0.27 |
| ***moving*** | **7.28** | **22.18** | **0.33** | **34.42** | **157.56** | **60.01** | **58.02** | **0.27** |
| *21* | 6.07 | 22.15 | 0.27 | 34.56 | 162.00 | 59.83 | 57.83 | 0.27 |
| *63* | 7.03 | 22.19 | 0.32 | 34.05 | 158.67 | 59.53 | 58.25 | 0.27 |
| *126* | 8.74 | 22.19 | 0.39 | 34.64 | 152.00 | 60.67 | 57.97 | 0.27 |
| ***Grand Total*** | **8.03** | **22.18** | **0.36** | **37.14** | **166.63** | **56.33** | **57.64** | **0.29** |

Note: For GJR-GARCH the strategy based on expanding training window with 504 training window days produced on average the best IR (0.42)

Table 18. The analysis of the performance of 432 GARCH model combinations described in detail in Section 4.2: Training window size – 126

| Strategy | Average of ARC | Average of ASD | Average of IR | Average of MD | Average of trades | Average of signal_long | Average of elf | Average of rmse |
|---|---|---|---|---|---|---|---|---|
| ***eGARCH*** | **4.93** | **22.02** | **0.22** | **36.98** | **171.33** | **55.11** | **55.21** | **0.31** |
| *expanding* | 3.15 | 22.01 | 0.14 | 39.89 | 175.78 | 56.21 | 55.20 | 0.31 |
| *21* | 2.99 | 22.00 | 0.13 | 46.95 | 212.67 | 50.57 | 55.20 | 0.26 |
| *63* | 4.40 | 22.11 | 0.20 | 33.87 | 164.33 | 58.30 | 54.95 | 0.31 |
| *126* | 2.07 | 21.90 | 0.09 | 38.84 | 150.33 | 59.77 | 55.44 | 0.38 |
| *moving* | 6.70 | 22.04 | 0.30 | 34.07 | 166.89 | 54.01 | 55.23 | 0.32 |
| *21* | 6.50 | 22.05 | 0.29 | 31.94 | 200.67 | 50.20 | 55.46 | 0.22 |
| *63* | 8.87 | 22.11 | 0.40 | 31.47 | 162.67 | 54.30 | 54.73 | 0.37 |
| *126* | 4.75 | 21.96 | 0.21 | 38.80 | 137.33 | 57.53 | 55.50 | 0.36 |
| ***fGARCH*** | **7.67** | **22.14** | **0.35** | **31.85** | **214.22** | **56.75** | **57.62** | **0.33** |
| *expanding* | 9.03 | 22.17 | 0.41 | 29.49 | 213.56 | 57.32 | 57.76 | 0.32 |
| *21* | 8.59 | 22.17 | 0.39 | 30.15 | 238.00 | 55.53 | 58.14 | 0.29 |
| *63* | 9.78 | 22.14 | 0.44 | 26.39 | 218.00 | 56.43 | 57.90 | 0.32 |
| *126* | 8.71 | 22.20 | 0.39 | 31.92 | 184.67 | 60.00 | 57.24 | 0.35 |
| *moving* | 6.31 | 22.12 | 0.28 | 34.21 | 214.89 | 56.18 | 57.48 | 0.33 |
| *21* | 5.89 | 22.10 | 0.27 | 30.40 | 238.67 | 55.70 | 57.62 | 0.28 |
| *63* | 8.00 | 22.11 | 0.36 | 29.20 | 220.67 | 55.07 | 57.56 | 0.36 |
| *126* | 5.03 | 22.14 | 0.23 | 43.03 | 185.33 | 57.77 | 57.26 | 0.34 |
| ***gjrGARCH*** | **6.92** | **22.14** | **0.31** | **36.54** | **175.89** | **53.91** | **56.95** | **0.33** |
| *expanding* | 6.87 | 22.16 | 0.31 | 37.26 | 172.22 | 54.06 | 56.94 | 0.32 |
| *21* | 4.85 | 22.15 | 0.22 | 37.24 | 198.67 | 51.27 | 56.86 | 0.28 |
| *63* | 7.50 | 22.16 | 0.34 | 38.74 | 168.67 | 54.27 | 57.36 | 0.32 |
| *126* | 8.26 | 22.15 | 0.37 | 35.81 | 149.33 | 56.63 | 56.60 | 0.37 |
| *moving* | 6.98 | 22.12 | 0.31 | 35.81 | 179.56 | 53.76 | 56.96 | 0.33 |
| *21* | 1.33 | 22.06 | 0.06 | 39.37 | 206.67 | 50.53 | 56.42 | 0.29 |
| *63* | 11.52 | 22.21 | 0.52 | 31.83 | 176.00 | 53.53 | 57.39 | 0.36 |
| *126* | 8.08 | 22.09 | 0.37 | 36.22 | 156.00 | 57.20 | 57.07 | 0.35 |
| ***sGARCH*** | **-1.28** | **20.82** | **-0.06** | **52.84** | **147.78** | **52.09** | **49.21** | **0.31** |
| *expanding* | -1.81 | 19.57 | -0.08 | 50.09 | 134.22 | 48.61 | 46.44 | 0.29 |
| *21* | -0.89 | 14.74 | -0.04 | 36.19 | 116.00 | 34.90 | 34.41 | 0.18 |
| *63* | -3.54 | 21.99 | -0.16 | 60.55 | 162.33 | 54.33 | 52.20 | 0.34 |
| *126* | -1.01 | 21.98 | -0.05 | 53.53 | 124.33 | 56.60 | 52.70 | 0.37 |
| *moving* | -0.75 | 22.07 | -0.03 | 55.59 | 161.33 | 55.57 | 51.99 | 0.32 |
| *21* | -0.32 | 22.16 | -0.01 | 54.91 | 184.00 | 53.93 | 51.72 | 0.27 |
| *63* | -0.94 | 22.14 | -0.04 | 56.07 | 176.33 | 54.93 | 52.18 | 0.34 |
| *126* | -0.97 | 21.92 | -0.05 | 55.80 | 123.67 | 57.83 | 52.08 | 0.36 |
| ***Grand Total*** | **4.56** | **21.78** | **0.21** | **39.55** | **177.31** | **54.46** | **54.75** | **0.32** |

Note: The strategy with 126 training days is the most dominant on fGARCH – TGARCH specification: IR (0.35)

Table 19. The analysis of the performance of 432 GARCH model combinations described in detail in Section 4.2: Training window size – 252

| *Strategy* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of elf* | *Average of rmse* |
|---|---|---|---|---|---|---|---|---|
| ***eGARCH*** | **-3.98** | **21.94** | **-0.18** | **58.10** | **173.56** | **57.28** | **56.36** | **0.46** |
| *expanding* | -4.61 | 21.90 | -0.21 | 60.39 | 175.78 | 57.29 | 56.40 | 0.46 |
| *21* | -1.86 | 22.01 | -0.08 | 56.35 | 194.67 | 53.93 | 56.31 | 0.41 |
| *63* | -5.97 | 21.84 | -0.27 | 62.12 | 182.00 | 57.63 | 56.79 | 0.46 |
| *126* | -6.01 | 21.86 | -0.28 | 62.68 | 150.67 | 60.30 | 56.10 | 0.51 |
| *moving* | -3.35 | 21.97 | -0.15 | 55.81 | 171.33 | 57.28 | 56.33 | 0.46 |
| *21* | -2.40 | 21.94 | -0.11 | 54.90 | 200.67 | 54.10 | 56.85 | 0.39 |
| *63* | -5.27 | 21.91 | -0.24 | 58.83 | 175.33 | 57.87 | 56.42 | 0.46 |
| *126* | -2.39 | 22.05 | -0.11 | 53.68 | 138.00 | 59.87 | 55.71 | 0.52 |
| ***fGARCH*** | **11.52** | **22.20** | **0.52** | **32.25** | **190.67** | **55.12** | **58.03** | **0.31** |
| *expanding* | 11.50 | 22.19 | 0.52 | 31.96 | 190.11 | 55.23 | 58.05 | 0.32 |
| *21* | 14.73 | 22.27 | 0.66 | 27.76 | 201.33 | 54.87 | 58.37 | 0.30 |
| *63* | 10.28 | 22.17 | 0.46 | 32.44 | 195.00 | 55.30 | 58.02 | 0.32 |
| *126* | 9.49 | 22.12 | 0.43 | 35.69 | 174.00 | 55.53 | 57.75 | 0.33 |
| *moving* | 11.53 | 22.21 | 0.52 | 32.54 | 191.22 | 55.01 | 58.01 | 0.31 |
| *21* | 14.35 | 22.22 | 0.65 | 29.21 | 201.33 | 54.83 | 58.31 | 0.30 |
| *63* | 11.54 | 22.23 | 0.52 | 31.60 | 190.33 | 54.73 | 58.21 | 0.31 |
| *126* | 8.72 | 22.19 | 0.39 | 36.82 | 182.00 | 55.47 | 57.51 | 0.32 |
| ***gjrGARCH*** | **8.85** | **22.24** | **0.40** | **44.32** | **165.78** | **55.23** | **57.84** | **0.28** |
| *expanding* | 8.56 | 22.22 | 0.38 | 44.44 | 164.67 | 55.11 | 57.77 | 0.29 |
| *21* | 10.56 | 22.27 | 0.47 | 40.49 | 172.00 | 55.17 | 58.06 | 0.28 |
| *63* | 7.69 | 22.24 | 0.35 | 43.68 | 169.33 | 54.83 | 57.77 | 0.28 |
| *126* | 7.43 | 22.16 | 0.33 | 49.16 | 152.67 | 55.33 | 57.48 | 0.30 |
| *moving* | 9.14 | 22.26 | 0.41 | 44.19 | 166.89 | 55.36 | 57.91 | 0.28 |
| *21* | 9.87 | 22.26 | 0.44 | 40.07 | 176.00 | 55.03 | 58.00 | 0.28 |
| *63* | 8.71 | 22.28 | 0.39 | 44.61 | 170.00 | 55.53 | 57.89 | 0.28 |
| *126* | 8.83 | 22.23 | 0.40 | 47.90 | 154.67 | 55.50 | 57.86 | 0.28 |
| ***sGARCH*** | **1.18** | **22.08** | **0.05** | **45.47** | **153.00** | **57.94** | **54.81** | **0.26** |
| *expanding* | 1.29 | 22.08 | 0.06 | 44.67 | 154.56 | 57.49 | 54.72 | 0.26 |
| *21* | 1.05 | 22.06 | 0.05 | 41.97 | 160.33 | 56.43 | 54.56 | 0.22 |
| *63* | -0.76 | 22.02 | -0.03 | 48.91 | 160.33 | 57.83 | 54.78 | 0.27 |
| *126* | 3.58 | 22.16 | 0.16 | 43.13 | 143.00 | 58.20 | 54.81 | 0.29 |
| *moving* | 1.07 | 22.07 | 0.05 | 46.27 | 151.44 | 58.39 | 54.90 | 0.26 |
| *21* | 2.12 | 22.15 | 0.10 | 41.51 | 155.00 | 56.23 | 54.78 | 0.22 |
| *63* | -0.18 | 22.06 | -0.01 | 49.88 | 157.00 | 58.50 | 54.90 | 0.27 |
| *126* | 1.25 | 22.01 | 0.06 | 47.44 | 142.33 | 60.43 | 55.01 | 0.28 |
| ***Grand Total*** | **4.39** | **22.11** | **0.20** | **45.03** | **170.75** | **56.39** | **56.76** | **0.29** |

Note: Same with 252-days training window – strategy based on fGARCH – TGARCH produced IR (0.52) which is far above the outcomes of remaining specifications

Table 20. The analysis of the performance of 432 GARCH model combinations described in detail in Section 4.2: Training window size – 504

| *Strategy* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of elf* | *Average of rmse* |
|---:|---:|---:|---:|---:|---:|---:|---:|---:|
| ***eGARCH*** | **4.41** | **22.06** | **0.20** | **33.80** | **185.44** | **54.72** | **57.92** | **6.84** |
| *expanding* | 4.49 | 22.07 | 0.20 | 33.27 | 180.89 | 54.56 | 57.83 | 1.23 |
| *21* | 6.14 | 22.09 | 0.28 | 27.14 | 186.00 | 54.13 | 57.68 | 0.28 |
| *63* | 3.34 | 22.01 | 0.15 | 37.53 | 178.00 | 55.17 | 57.94 | 0.8 |
| *126* | 3.98 | 22.11 | 0.18 | 35.12 | 178.67 | 54.37 | 57.89 | 2.61 |
| *moving* | 4.33 | 22.04 | 0.20 | 34.34 | 190.00 | 54.89 | 58.00 | 12.45 |
| *21* | 5.58 | 22.14 | 0.25 | 31.20 | 193.33 | 55.50 | 57.87 | 36.6 |
| *63* | 2.62 | 21.97 | 0.12 | 39.38 | 193.33 | 54.80 | 58.22 | 0.35 |
| *126* | 4.79 | 22.01 | 0.22 | 32.45 | 183.33 | 54.37 | 57.90 | 0.4 |
| ***fGARCH*** | **8.27** | **22.16** | **0.37** | **29.52** | **184.67** | **56.83** | **58.39** | **0.27** |
| *expanding* | 8.04 | 22.16 | 0.36 | 29.48 | 184.00 | 56.87 | 58.31 | 0.27 |
| *21* | 7.21 | 22.16 | 0.33 | 29.19 | 193.33 | 56.83 | 58.09 | 0.27 |
| *63* | 9.24 | 22.16 | 0.42 | 28.91 | 176.33 | 57.20 | 58.09 | 0.27 |
| *126* | 7.66 | 22.17 | 0.35 | 30.33 | 182.33 | 56.57 | 58.75 | 0.27 |
| *moving* | 8.50 | 22.16 | 0.38 | 29.57 | 185.33 | 56.79 | 58.46 | 0.27 |
| *21* | 8.72 | 22.14 | 0.39 | 29.13 | 193.33 | 56.77 | 58.39 | 0.26 |
| *63* | 9.13 | 22.21 | 0.41 | 29.15 | 181.00 | 56.73 | 58.48 | 0.27 |
| *126* | 7.64 | 22.13 | 0.35 | 30.43 | 181.67 | 56.87 | 58.53 | 0.27 |
| ***gjrGARCH*** | **8.32** | **22.16** | **0.38** | **30.56** | **158.22** | **59.86** | **58.12** | **0.27** |
| *expanding* | 9.36 | 22.15 | 0.42 | 26.71 | 158.89 | 59.70 | 58.23 | 0.27 |
| *21* | 8.83 | 22.12 | 0.40 | 25.38 | 160.67 | 59.37 | 58.16 | 0.27 |
| *63* | 8.09 | 22.11 | 0.37 | 27.69 | 160.00 | 59.73 | 58.29 | 0.27 |
| *126* | 11.17 | 22.22 | 0.50 | 27.05 | 156.00 | 60.00 | 58.23 | 0.27 |
| *moving* | 7.28 | 22.18 | 0.33 | 34.42 | 157.56 | 60.01 | 58.02 | 0.27 |
| *21* | 6.07 | 22.15 | 0.27 | 34.56 | 162.00 | 59.83 | 57.83 | 0.27 |
| *63* | 7.03 | 22.19 | 0.32 | 34.05 | 158.67 | 59.53 | 58.25 | 0.27 |
| *126* | 8.74 | 22.19 | 0.39 | 34.64 | 152.00 | 60.67 | 57.97 | 0.27 |
| ***sGARCH*** | **-0.55** | **22.00** | **-0.03** | **45.92** | **155.22** | **60.27** | **56.58** | **0.21** |
| *expanding* | 0.09 | 21.99 | 0.00 | 45.74 | 153.89 | 60.11 | 56.56 | 0.21 |
| *21* | 0.65 | 22.08 | 0.03 | 43.85 | 162.33 | 59.93 | 56.85 | 0.21 |
| *63* | -0.58 | 21.93 | -0.03 | 47.63 | 156.33 | 59.93 | 56.72 | 0.21 |
| *126* | 0.19 | 21.97 | 0.01 | 45.75 | 143.00 | 60.47 | 56.10 | 0.21 |
| *moving* | -1.19 | 22.01 | -0.05 | 46.11 | 156.56 | 60.43 | 56.60 | 0.21 |
| *21* | -1.29 | 21.99 | -0.06 | 44.65 | 163.67 | 60.07 | 56.85 | 0.21 |
| *63* | -2.16 | 22.08 | -0.10 | 50.06 | 159.67 | 60.30 | 56.78 | 0.21 |
| *126* | -0.11 | 21.97 | -0.01 | 43.63 | 146.33 | 60.93 | 56.19 | 0.22 |
| ***Grand Total*** | **5.11** | **22.10** | **0.23** | **34.95** | **170.89** | **57.92** | **57.75** | **1.9** |

Notes: On the longer training window (504) the strategy based on GJR-GARCH model outperformed on the edge the fGARCH – TGARCH: 0.38 against 0.37, respectively

Table 21. The analysis of the performance of 432 GARCH model combinations described in detail in Section 4.2: Refit window – 21

| *Strategy* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of elf* | *Average of rmse* |
|---|---|---|---|---|---|---|---|---|
| ***126*** | **3.62** | **21.18** | **0.16** | **38.39** | **199.42** | **50.33** | **53.23** | **0.27** |
| ***eGARCH*** | 4.74 | 22.03 | 0.21 | 39.45 | 206.67 | 50.38 | 55.33 | 0.24 |
| *expanding* | 2.99 | 22.00 | 0.13 | 46.95 | 212.67 | 50.57 | 55.20 | 0.26 |
| *moving* | 6.50 | 22.05 | 0.29 | 31.94 | 200.67 | 50.20 | 55.46 | 0.22 |
| ***fGARCH*** | 7.24 | 22.14 | 0.33 | 30.27 | 238.33 | 55.62 | 57.88 | 0.29 |
| *expanding* | 8.59 | 22.17 | 0.39 | 30.15 | 238.00 | 55.53 | 58.14 | 0.29 |
| *moving* | 5.89 | 22.10 | 0.27 | 30.40 | 238.67 | 55.70 | 57.62 | 0.28 |
| ***gjrGARCH*** | 3.09 | 22.11 | 0.14 | 38.31 | 202.67 | 50.90 | 56.64 | 0.29 |
| *expanding* | 4.85 | 22.15 | 0.22 | 37.24 | 198.67 | 51.27 | 56.86 | 0.28 |
| *moving* | 1.33 | 22.06 | 0.06 | 39.37 | 206.67 | 50.53 | 56.42 | 0.29 |
| ***sGARCH*** | -0.61 | 18.45 | -0.03 | 45.55 | 150.00 | 44.42 | 43.06 | 0.22 |
| *expanding* | -0.89 | 14.74 | -0.04 | 36.19 | 116.00 | 34.90 | 34.41 | 0.18 |
| *moving* | -0.32 | 22.16 | -0.01 | 54.91 | 184.00 | 53.93 | 51.72 | 0.27 |
| ***252*** | **6.05** | **22.15** | **0.27** | **41.53** | **182.67** | **55.08** | **56.91** | **0.27** |
| ***eGARCH*** | -2.13 | 21.97 | -0.10 | 55.63 | 197.67 | 54.02 | 56.58 | 0.39 |
| *expanding* | -1.86 | 22.01 | -0.08 | 56.35 | 194.67 | 53.93 | 56.31 | 0.39 |
| *moving* | -2.40 | 21.94 | -0.11 | 54.90 | 200.67 | 54.10 | 56.85 | 0.39 |
| ***fGARCH*** | 14.54 | 22.25 | 0.65 | 28.48 | 201.33 | 54.85 | 58.34 | 0.30 |
| *expanding* | 14.73 | 22.27 | 0.66 | 27.76 | 201.33 | 54.87 | 58.37 | 0.30 |
| *moving* | 14.35 | 22.22 | 0.65 | 29.21 | 201.33 | 54.83 | 58.31 | 0.30 |
| ***gjrGARCH*** | 10.21 | 22.26 | 0.46 | 40.28 | 174.00 | 55.10 | 58.03 | 0.28 |
| *expanding* | 10.56 | 22.27 | 0.47 | 40.49 | 172.00 | 55.17 | 58.06 | 0.28 |
| *moving* | 9.87 | 22.26 | 0.44 | 40.07 | 176.00 | 55.03 | 58.00 | 0.28 |
| ***sGARCH*** | 1.59 | 22.10 | 0.07 | 41.74 | 157.67 | 56.33 | 54.67 | 0.22 |
| *expanding* | 1.05 | 22.06 | 0.05 | 41.97 | 160.33 | 56.43 | 54.56 | 0.22 |
| *moving* | 2.12 | 22.15 | 0.10 | 41.51 | 155.00 | 56.23 | 54.78 | 0.22 |
| ***504*** | **5.24** | **22.11** | **0.24** | **33.14** | **176.83** | **57.80** | **57.72** | **3.71** |
| ***eGARCH*** | 5.86 | 22.12 | 0.26 | 29.17 | 189.67 | 54.82 | 57.77 | 24.49 |
| *expanding* | 6.14 | 22.09 | 0.28 | 27.14 | 186.00 | 54.13 | 57.68 | 0.28 |
| *moving* | 5.58 | 22.14 | 0.25 | 31.20 | 193.33 | 55.50 | 57.87 | 36.60 |
| ***fGARCH*** | 7.96 | 22.15 | 0.36 | 29.16 | 193.33 | 56.80 | 58.24 | 0.26 |
| *expanding* | 7.21 | 22.16 | 0.33 | 29.19 | 193.33 | 56.83 | 58.09 | 0.27 |
| *moving* | 8.72 | 22.14 | 0.39 | 29.13 | 193.33 | 56.77 | 58.39 | 0.26 |
| ***gjrGARCH*** | 7.45 | 22.13 | 0.34 | 29.97 | 161.33 | 59.60 | 58.00 | 0.27 |
| *expanding* | 8.83 | 22.12 | 0.40 | 25.38 | 160.67 | 59.37 | 58.16 | 0.27 |
| *moving* | 6.07 | 22.15 | 0.27 | 34.56 | 162.00 | 59.83 | 57.83 | 0.27 |
| ***sGARCH*** | -0.32 | 22.03 | -0.02 | 44.25 | 163.00 | 60.00 | 56.85 | 0.21 |
| *expanding* | 0.65 | 22.08 | 0.03 | 43.85 | 162.33 | 59.93 | 56.85 | 0.21 |
| *moving* | -1.29 | 21.99 | -0.06 | 44.65 | 163.67 | 60.07 | 56.85 | 0.21 |
| ***Grand Total*** | **4.97** | **21.81** | **0.22** | **37.69** | **186.31** | **54.40** | **55.95** | **1.52** |

Notes: The strategy refitted each 21st day produced the best investment outcomes on 252 days training window – IR (0.27)

Table 22. The analysis of the performance of 432 GARCH model combinations described in detail in Section 4.2: Refit window – 63

| *Strategy* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of elf* | *Average of rmse* |
|---|---|---|---|---|---|---|---|---|
| ***126*** | **5.70** | **22.12** | **0.26** | **38.52** | **181.13** | **55.15** | **55.53** | **0.34** |
| ***eGARCH*** | **6.64** | **22.11** | **0.30** | **32.67** | **163.50** | **56.30** | **54.84** | **0.34** |
| *expanding* | 4.40 | 22.11 | 0.20 | 33.87 | 164.33 | 58.30 | 54.95 | 0.31 |
| *moving* | 8.87 | 22.11 | 0.40 | 31.47 | 162.67 | 54.30 | 54.73 | 0.37 |
| ***fGARCH*** | **8.89** | **22.12** | **0.40** | **27.79** | **219.33** | **55.75** | **57.73** | **0.34** |
| *expanding* | 9.78 | 22.14 | 0.44 | 26.39 | 218.00 | 56.43 | 57.90 | 0.32 |
| *moving* | 8.00 | 22.11 | 0.36 | 29.20 | 220.67 | 55.07 | 57.56 | 0.36 |
| ***gjrGARCH*** | **9.51** | **22.18** | **0.43** | **35.29** | **172.33** | **53.90** | **57.37** | **0.34** |
| *expanding* | 7.50 | 22.16 | 0.34 | 38.74 | 168.67 | 54.27 | 57.36 | 0.32 |
| *moving* | 11.52 | 22.21 | 0.52 | 31.83 | 176.00 | 53.53 | 57.39 | 0.36 |
| ***sGARCH*** | **-2.24** | **22.07** | **-0.10** | **58.31** | **169.33** | **54.63** | **52.19** | **0.34** |
| *expanding* | -3.54 | 21.99 | -0.16 | 60.55 | 162.33 | 54.33 | 52.20 | 0.34 |
| *moving* | -0.94 | 22.14 | -0.04 | 56.07 | 176.33 | 54.93 | 52.18 | 0.34 |
| ***252*** | **3.26** | **22.09** | **0.15** | **46.51** | **174.92** | **56.53** | **56.85** | **0.30** |
| ***eGARCH*** | **-5.62** | **21.88** | **-0.26** | **60.48** | **178.67** | **57.75** | **56.60** | **0.46** |
| *expanding* | -5.97 | 21.84 | -0.27 | 62.12 | 182.00 | 57.63 | 56.79 | 0.46 |
| *moving* | -5.27 | 21.91 | -0.24 | 58.83 | 175.33 | 57.87 | 56.42 | 0.46 |
| ***fGARCH*** | **10.91** | **22.20** | **0.49** | **32.02** | **192.67** | **55.02** | **58.11** | **0.32** |
| *expanding* | 10.28 | 22.17 | 0.46 | 32.44 | 195.00 | 55.30 | 58.02 | 0.32 |
| *moving* | 11.54 | 22.23 | 0.52 | 31.60 | 190.33 | 54.73 | 58.21 | 0.31 |
| ***gjrGARCH*** | **8.20** | **22.26** | **0.37** | **44.15** | **169.67** | **55.18** | **57.83** | **0.28** |
| *expanding* | 7.69 | 22.24 | 0.35 | 43.68 | 169.33 | 54.83 | 57.77 | 0.28 |
| *moving* | 8.71 | 22.28 | 0.39 | 44.61 | 170.00 | 55.53 | 57.89 | 0.28 |
| ***sGARCH*** | **-0.47** | **22.04** | **-0.02** | **49.40** | **158.67** | **58.17** | **54.84** | **0.27** |
| *expanding* | -0.76 | 22.02 | -0.03 | 48.91 | 160.33 | 57.83 | 54.78 | 0.27 |
| *moving* | -0.18 | 22.06 | -0.01 | 49.88 | 157.00 | 58.50 | 54.90 | 0.27 |
| ***504*** | **4.59** | **22.08** | **0.21** | **36.80** | **170.42** | **57.93** | **57.85** | **0.34** |
| ***eGARCH*** | **2.98** | **21.99** | **0.14** | **38.46** | **185.67** | **54.98** | **58.08** | **0.62** |
| *expanding* | 3.34 | 22.01 | 0.15 | 37.53 | 178.00 | 55.17 | 57.94 | 0.8 |
| *moving* | 2.62 | 21.97 | 0.12 | 39.38 | 193.33 | 54.80 | 58.22 | 0.35 |
| ***fGARCH*** | **9.19** | **22.19** | **0.41** | **29.03** | **178.67** | **56.97** | **58.29** | **0.27** |
| *expanding* | 9.24 | 22.16 | 0.42 | 28.91 | 176.33 | 57.20 | 58.09 | 0.27 |
| *moving* | 9.13 | 22.21 | 0.41 | 29.15 | 181.00 | 56.73 | 58.48 | 0.27 |
| ***gjrGARCH*** | **7.56** | **22.15** | **0.34** | **30.87** | **159.33** | **59.63** | **58.27** | **0.27** |
| *expanding* | 8.09 | 22.11 | 0.37 | 27.69 | 160.00 | 59.73 | 58.29 | 0.27 |
| *moving* | 7.03 | 22.19 | 0.32 | 34.05 | 158.67 | 59.53 | 58.25 | 0.27 |
| ***sGARCH*** | **-1.37** | **22.00** | **-0.06** | **48.84** | **158.00** | **60.12** | **56.75** | **0.21** |
| *expanding* | -0.58 | 21.93 | -0.03 | 47.63 | 156.33 | 59.93 | 56.72 | 0.21 |
| *moving* | -2.16 | 22.08 | -0.10 | 50.06 | 159.67 | 60.30 | 56.78 | 0.21 |
| ***Grand Total*** | **4.51** | **22.10** | **0.20** | **40.61** | **175.49** | **56.53** | **56.74** | **0.33** |

Notes: For the strategy refitted on 63$^{rd}$ day, the best investment outcomes were obtained for the shortest training window with 126 days

Table 23. The analysis of the performance of 432 GARCH model combinations described in detail in Section 4.2: Refit window – 126

| *Strategy* | *Average of ARC* | *Average of ASD* | *Average of IR* | *Average of MD* | *Average of trades* | *Average of signal_long* | *Average of elf* | *Average of rmse* |
|---|---|---|---|---|---|---|---|---|
| ***126*** | **4.36** | **22.04** | **0.20** | **41.75** | **151.38** | **57.92** | **55.49** | **0.36** |
| ***eGARCH*** | **3.41** | **21.93** | **0.15** | **38.82** | **143.83** | **58.65** | **55.47** | **0.37** |
| *expanding* | 2.07 | 21.90 | 0.09 | 38.84 | 150.33 | 59.77 | 55.44 | 0.38 |
| *moving* | 4.75 | 21.96 | 0.21 | 38.80 | 137.33 | 57.53 | 55.50 | 0.36 |
| ***fGARCH*** | **6.87** | **22.17** | **0.31** | **37.48** | **185.00** | **58.88** | **57.25** | **0.35** |
| *expanding* | 8.71 | 22.20 | 0.39 | 31.92 | 184.67 | 60.00 | 57.24 | 0.35 |
| *moving* | 5.03 | 22.14 | 0.23 | 43.03 | 185.33 | 57.77 | 57.26 | 0.34 |
| ***gjrGARCH*** | **8.17** | **22.12** | **0.37** | **36.02** | **152.67** | **56.92** | **56.84** | **0.36** |
| *expanding* | 8.26 | 22.15 | 0.37 | 35.81 | 149.33 | 56.63 | 56.60 | 0.37 |
| *moving* | 8.08 | 22.09 | 0.37 | 36.22 | 156.00 | 57.20 | 57.07 | 0.35 |
| ***sGARCH*** | **-0.99** | **21.95** | **-0.05** | **54.67** | **124.00** | **57.22** | **52.39** | **0.36** |
| *expanding* | -1.01 | 21.98 | -0.05 | 53.53 | 124.33 | 56.60 | 52.70 | 0.37 |
| *moving* | -0.97 | 21.92 | -0.05 | 55.80 | 123.67 | 57.83 | 52.08 | 0.36 |
| ***252*** | **3.86** | **22.10** | **0.17** | **47.06** | **154.67** | **57.58** | **56.53** | **0.31** |
| ***eGARCH*** | **-4.20** | **21.96** | **-0.19** | **58.18** | **144.33** | **60.08** | **55.91** | **0.52** |
| *expanding* | -6.01 | 21.86 | -0.28 | 62.68 | 150.67 | 60.30 | 56.10 | 0.1 |
| *moving* | -2.39 | 22.05 | -0.11 | 53.68 | 138.00 | 59.87 | 55.71 | 0.52 |
| ***fGARCH*** | **9.10** | **22.16** | **0.41** | **36.25** | **178.00** | **55.50** | **57.63** | **0.33** |
| *expanding* | 9.49 | 22.12 | 0.43 | 35.69 | 174.00 | 55.53 | 57.75 | 0.33 |
| *moving* | 8.72 | 22.19 | 0.39 | 36.82 | 182.00 | 55.47 | 57.51 | 0.32 |
| ***gjrGARCH*** | **8.13** | **22.19** | **0.37** | **48.53** | **153.67** | **55.42** | **57.67** | **0.29** |
| *expanding* | 7.43 | 22.16 | 0.33 | 49.16 | 152.67 | 55.33 | 57.48 | 0.30 |
| *moving* | 8.83 | 22.23 | 0.40 | 47.90 | 154.67 | 55.50 | 57.86 | 0.28 |
| ***sGARCH*** | **2.42** | **22.09** | **0.11** | **45.28** | **142.67** | **59.32** | **54.91** | **0.29** |
| *expanding* | 3.58 | 22.16 | 0.16 | 43.13 | 143.00 | 58.20 | 54.81 | 0.29 |
| *moving* | 1.25 | 22.01 | 0.06 | 47.44 | 142.33 | 60.43 | 55.01 | 0.28 |
| ***504*** | **5.51** | **22.10** | **0.25** | **34.92** | **165.42** | **58.03** | **57.70** | **0.56** |
| ***eGARCH*** | **4.39** | **22.06** | **0.20** | **33.79** | **181.00** | **54.37** | **57.90** | **1.5** |
| *expanding* | 3.98 | 22.11 | 0.18 | 35.12 | 178.67 | 54.37 | 57.89 | 2.61 |
| *moving* | 4.79 | 22.01 | 0.22 | 32.45 | 183.33 | 54.37 | 57.90 | 0.4 |
| ***fGARCH*** | **7.65** | **22.15** | **0.35** | **30.38** | **182.00** | **56.72** | **58.64** | **0.27** |
| *expanding* | 7.66 | 22.17 | 0.35 | 30.33 | 182.33 | 56.57 | 58.75 | 0.27 |
| *moving* | 7.64 | 22.13 | 0.35 | 30.43 | 181.67 | 56.87 | 58.53 | 0.27 |
| ***gjrGARCH*** | **9.96** | **22.21** | **0.45** | **30.84** | **154.00** | **60.33** | **58.10** | **0.27** |
| *expanding* | 11.17 | 22.22 | 0.50 | 27.05 | 156.00 | 60.00 | 58.23 | 0.27 |
| *moving* | 8.74 | 22.19 | 0.39 | 34.64 | 152.00 | 60.67 | 57.97 | 0.27 |
| ***sGARCH*** | **0.04** | **21.97** | **0.00** | **44.69** | **144.67** | **60.70** | **56.14** | **0.21** |
| *expanding* | 0.19 | 21.97 | 0.01 | 45.75 | 143.00 | 60.47 | 56.10 | 0.21 |
| *moving* | -0.11 | 21.97 | -0.01 | 43.63 | 146.33 | 60.93 | 56.19 | 0.22 |
| ***Grand Total*** | **4.58** | **22.08** | **0.21** | **41.24** | **157.15** | **57.84** | **56.57** | **0.41** |

Note: For the strategy refitted on 126$^{th}$ day, the best investment outcomes were obtained for the longest training window – 504 (IR = 0.25)

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