Multifractal volatility predictors for enhanced portfolio strategies

Master Thesis

Selena Pepic

May 22, 2021

Supervisors:
Prof. Didier Sornette
Prof. Thomas Hofmann

Department of Computer Science, ETH Zürich

Acknowledgements

Firstly, I would like to express my gratitude to Prof. Didier Sornette for the opportunity to write my Master thesis under his guidance and for pointing me into the right direction when needed. Not many would give a chance to someone with non-financial background to join this world and this thesis was a great opportunity to get an insight in the research on financial markets volatility.

I would like to thank professor Thomas Hoffman for agreeing to be my official supervisor and ETH Zürich for giving me scholarship and therefore enabling me to attend this great university that helped me improve my knowledge and understanding of science. I am also grateful to Rebecca Westphal and Qian Wang who agreed to e-meet with me and discuss the questions I had. I would also like to thank Nikola, since without him I would never consider changing profession.

I would like to express my gratitude to Anna, Jovan, Jasmina, Milan, Nemanja, Stefan, bridge and chess for the, sometimes, much needed distraction. Most of all, I would like to thank Kosta for being an example and support in the last six months.

Lastly, thank you mom, dad and sister for being the best family in the world. Nothing would have been possible without you.

Abstract

The return for given security or market index is very stochastic and non-stationary with many regime shifts and it seems to be unpredictable or very hard to predict. This is not the case with its second moment - volatility. Volatility is a measure of dispersion of returns and, therefore, it quantifies the level of uncertainty.

(G)ARCH is the most popular model for forecasting volatility and a well-accepted benchmark solution to this problem. There are many variants of ARCH model and the most popular of them are explored and compared in this thesis. However, issue with ARCH is that it does not explain nor takes into account volatility at different frequencies. Another part of this thesis deals with multifractal models which are built in order to account for this downside of (G)ARCH.

Another type of models that proved to be good are regime-switching models. These models deal with modeling underlying process as a Markov process and assume that volatility is determined by components that have different degrees of persistence and that switch over time. Some results based on regime-switching models are outperforming (G)ARCH and therefore are definitely one of the topics of most interest when it comes to volatility predictors.

In this thesis, we show that none of the volatility predictors captured well the pandemic influence on market volatility and propose including exogenous variables to the model. We also show that MSM outperforms other models in the low volatility state, but does not do so when markets are unstable. The performance of mentioned models depends highly on the underlying asset and we evaluate them on four different assets.

Lastly, after building volatility predictors, momentum investment strategies and volatility-based risk hedging is discussed.

Contents

Chapter 1

Introduction

Financial markets are very complicated systems influenced by many people and companies, by various timings across many, more or less, correlated assets. They are naturally driven by news which can range from weather news [46], up to the presidential tweets such as the ones from the former president of the US, Donald Trump [31], or technological innovations that might not come to production in years [22].

Arguably, financial markets are most interesting when they are volatile, i.e. unstable. Volatility, sometimes also called fear index, since it is often associated with risk, measures the magnitude of the change in asset prices. Therefore, volatility forecasting plays an important role for almost everyone involved in the financial markets from options pricing, investment, portfolio optimization, risk management, and more concretely Value-at-Risk to the purely impressive scientific challenge. Moreover, being in accordance with Efficient Market Hypothesis, it has been shown that returns themselves are difficult to forecast. On the other hand, volatility is, to a certain extent, predictable. The main topic of this thesis is the comparison of the various volatility forecasting models.

It comes as no surprise that forecasting financial market volatility has received extensive attention, especially with the rise of popularity of derivatives - main instruments used for hedging. It is the only not directly observable variable in the Black-Scholes formula for options pricing and option values are, therefore, mostly determined by market forecasts of future volatility. Volatility is particularly important for options traders since higher volatility makes options more valuable. Except for this obvious reason for interest, it also represents a challenge to the researchers and is thus a very popular topic in academic circles. When Black-Scholes formula was invented, only a few types of options were traded and even they had short maturity.

Market participants use diverse strategies and investment horizons depending on beliefs or their core business. While there are participants such as high-frequency speculators, arbitrageurs and day traders who exploit opportunities over the very short time periods, there are also participants such as insurance companies or pension funds that look for opportunities that can span over several decades. Of course, with the rise of technology and machine learning, algorithmic trading becomes increasingly popular so that even long-run investors engage in high-frequency trading.

This thesis is based on the two main volatility forecasting models and their alterations. First one is Robert Engle's ARCH and its famous configuration GARCH and the second one is motivated by the research by Calvet and Fisher in which they claim that Markov-switching Multifractal (MSM) can outperform some of the most reliable forecasting models, including previously mentioned (G)ARCH. Markov-switching multifractal, or shortly MSM, is considered a bridge between two important features of volatility - multifractality and regime-switching. It takes regime-switching idea, following the work of Hamilton, and intuition of multifractality which is in essence idea that there are multiple shocks of different degrees of persistence. The models were evaluated and tested on various datasets of different liquidity and characteristics. Findings and conclusions are reported in the chapter 4.

At the end, based on the findings in this thesis, momentum-like trading strategy is proposed, evaluated on multiple datasets and compared to the benchmark models. Momentum is also known as a trend-following strategy and it is one of the oldest and most well-researched phenomena in the finance. Existence of momentum is considered a market anomaly and financial theory struggles to explain it. Big factor in momentum strategy is precisely volatility and while momentum is overall well-performing strategy, momentum crashes are not an unknown term in financial world. Momentum crash refers to worse performing stocks in the market rebounding more quickly than winners, a famous example being Spring 2009.

Chapter 2

Volatility

2.1 Definition

Volatility is a statistical measure of the dispersion of returns. It has no exact or fixed definition but is often referred to as variance, usually denoted by σ^2 or a standard standard deviation of logarithmic returns from expectation and it represents the degree of variation of a price over time.

The logarithmic return of an asset over a time interval Δt is defined as

$$r_{\Delta t} = \ln p_t - \ln p_{t-\Delta t}$$

where p_t is the price of an asset at time t.

In the literature $r_{\Delta t}$ is usually written shorter as r_t and it represents logarithmic return at time t after some assumed Δt. Volatility is, therefore, defined as

$$\sigma^2 = \sqrt{\frac{1}{N-1} \sum_{t=1}^{N} (r_t - \bar{r})}$$

where $\bar{r}$ is the average return over this sample period and N is the number of days. Note also that volatility is sometimes computed as the standard deviation of price returns rather than logarithmic price returns.

2.2 Types of volatility and terminology

There are many possible meanings of word volatility depending on the (missing) adjective in front of it.

Realized volatility, sometimes also called statistical volatility, is, as the name says, the volatility realized in the market, i.e. the one that has already been measured. It is empirical unconditional variance over a given time period and can be calculated as the sum of squared returns, most commonly intraday ones.

Conditional volatility is the volatility of a random variable given some additional information such as the past values of itself or model errors in the past.

Implied volatility comes from the price of an option and represents its volatility in the near future. Option pricing models such as the most popular Black-Scholes model calculate option price based on option's strike price, time to expiration, underlying price and volatility. Since the first three variables in the formula are observable variables, volatility is the main unknown. On the other hand, based on option prices in the market, one can calculate its implied volatility, or, in other words, reverse the equation acting like the option price is the known variable.

One should note that while realized volatility can be calculated even for securities without any options on them, implied volatility does not exist without options. Among many differences of implied and realized volatility, main is that implied volatility is always forward-looking, while realized volatility can relate to the past when it is often called **historical volatility**, or to the future when it is called future realized volatility.

Volatility can also vary depending on a time horizon. **Annualized volatility** is the standard deviation of yearly logarithmic returns. The volatility for time horizon T in years can be calculated as

$$\sigma_T = \sigma_{annually}\sqrt{T}$$

Using the same analogy, if by σ_{daily} we denote the daily volatility, then we can calculate volatility for time horizon of T trading days as

$$\sigma_T = \sigma_{daily}\sqrt{T}$$

In the special case of converting daily volatility to annualized, we can set the value of T to 252 in the previous formula since there are approximately 252 trading days in a year.

2.3 Stylized facts about volatility

There are some well-known regularities when it comes to the asset returns and volatility [37].

1. **Horizontal volatility dependence.** Volatility is **mean-reverting** and tends to **cluster**. Except from that, unlike returns, volatility has **long-memory**, i.e. it exhibits autocorrelation.
2. **Extreme events.** The returns distribution has **fat-tails**. It is noticeable that returns tend to stay close to mean or very far from it.
3. **Leverage effect.** Leverage effect is the effect of volatility having a negative correlation with returns. In other words, it is more common that the volatility increases after negative returns than after the positive ones. This is also known as asymmetric volatility phenomenon.
4. **(Asymmetric) vertical dependence.** Volatility acts differently depending on the time interval we are looking at.

2.3.1 Horizontal volatility dependence

Given Efficient Market Hypothesis, first moment of the asset returns is less predictable than volatility - its second moment[1]. In other words, while return time series exhibit absence of linear autocorrelation[2] and very short memory, financial volatility has long-range linear correlation which is known as clustering. That means that large or small price changes are likely to be followed by respectively large or small price changes which is illustrated in the figure 2.1. While in the figure 2.1 we can clearly see the clustering of the returns data, figure 2.2 highlights the absence of clustering in the standardized residuals after fitting the model to the data.

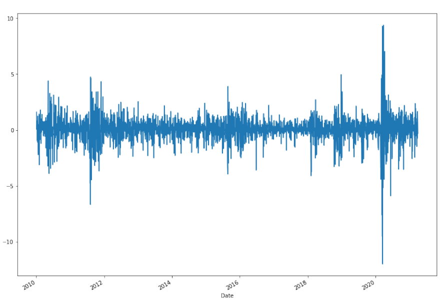

Figure 2.1: Returns of S&P500 data from January 2010 to March 2021

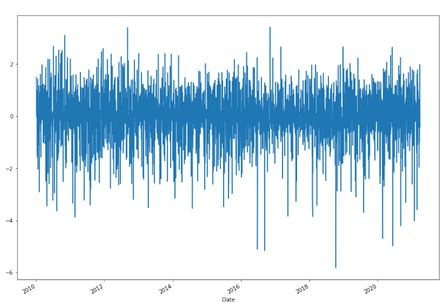

Figure 2.2: Standardized residuals of S&P500 data from January 2010 to March 2021

The autocorrelation coefficient measures the unconditional correlations of two series while partial autocorrelation measures the relationship of two

[1]For the distribution, mean would be the first moment, volatility the second, skewness the third and kurtosis the second moment.

[2]This is actually not entirely true. It is shown that return series contain significant autocorrelations during the early part of the day.

series taking into account the relationships of all previous lags. The difference is, therefore, that autocorrelation of lag k is the correlation between X_t and X_{t+k} while the partial autocorrelation of lag k is the conditional correlation of X_t and X_{t+k} given the values of $X_{t+1}, X_{t+2}, ..., X_{t+k-1}$.

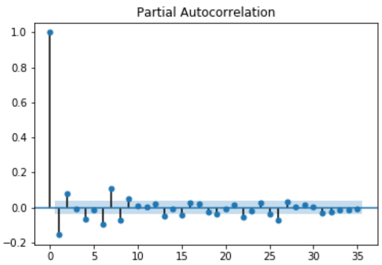

Figure 2.3: Partial Autocorrelation of returns series for S&P500 data from January 2010 to March 2021

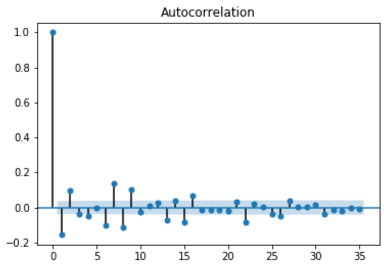

Figure 2.4: Autocorrelation of returns series for S&P500 data from January 2010 to March 2021

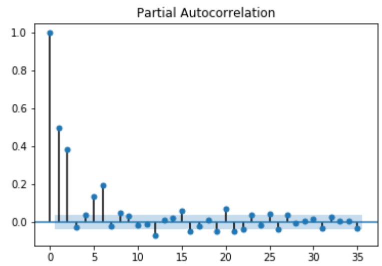

Figure 2.5: Partial Autocorrelation of squared returns series for S&P500 data from January 2010 to March 2021

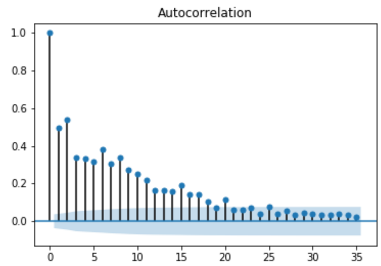

Figure 2.6: Autocorrelation of squared returns series for S&P500 data from January 2010 to March 2021

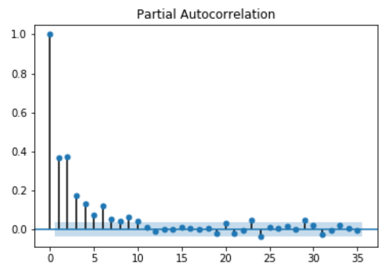

Figure 2.7: Partial autocorrelation of absolute returns series for S&P500 data from January 2010 to March 2021

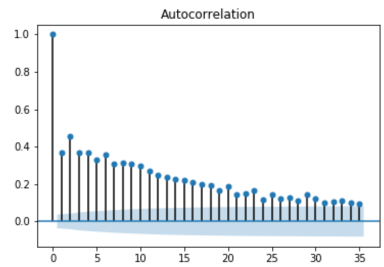

Figure 2.8: Autocorrelation of absolute returns series for S&P500 data from January 2010 to March 2021

On the figures 2.4-2.8 we can see correlograms of S&P500 series both for returns series, squared returns series and absolute returns series. While, for example, partial autocorrelation graph of returns series appears to be random and to be equally likely positive or negative from one observation to the next one, the partial autocorrelation of squared returns and absolute returns series shows significant correlation for the shorter lags. This tells us that while the sign of the value of observations is independent of the past, the magnitude of the change in observations may show correlation. As can be seen in all of the plots, at lag 0, the correlation is 1 as the data is perfectly correlated with itself. The correlograms are done on the daily returns data even though similar plots can be obtained by checking cumulative intraday returns (see for example [21]).

As already mentioned, when returns are largely negative (or positive) on a given day, we would expect large movements on the following days as well. This is one of the main characteristics of volatility and it is incorporated in the ARCH model that will be explored in the next chapter. Persistency of a volatility is present both within small and large intervals of time, namely both within the trading day and across many trading decades.

2.3.2 Extreme events

Distribution of asset returns is much further from normal distribution as was once thought. There exists excess kurtosis and therefore asset returns follow leptokurtic distribution.

An easy way to check for normality is to look at the QQ plot where deviations from the red line represent differences from normal distribution. On the figure 2.9, one can see that distribution of returns of S&P500 data deviates from normal distribution on both tails while having bigger deviations on the left one. While the mean of the plotted distribution is around 0.00 and standard deviation is around 1.09, we have negative skew of -1.11 and kurtosis of 24.20.

There are multiple ways to deal with this characteristic of the distribution of asset returns. This ranges from incorporating this stylized fact into the model used to explain the returns to looking at other distributions, most notably student's t-distribution which will be introduced in the next chapter.

What is also common is that markets experience shift in the volatility level. This usually happens after some big market event such as Asian crisis in 1997 impacting South Korean Stock Exchange Composite Index (KOSPI). Another common reason for the shift is policy changes [21].

2.3.3 Leverage effect

Leverage effect or volatility asymmetry corresponds to a negative correlation between past returns and future volatility, namely, volatility increases

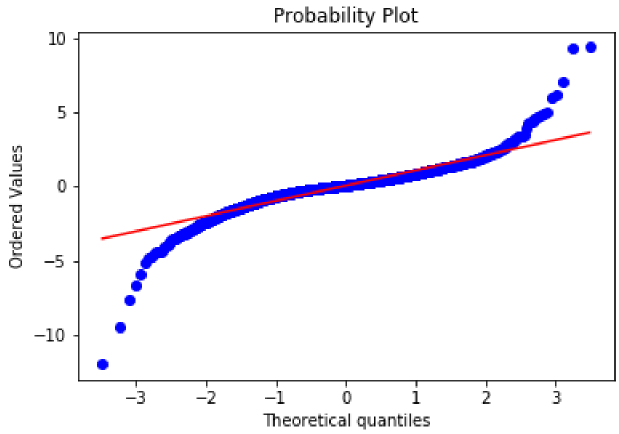

Figure 2.9: QQ plot for S&P500 data from January 2010 to March 2021

when the stock price falls [19]. The unconditional distribution of returns is negatively skewed since large negative returns are more common then the large positive returns. While for individual stocks leverage effect is small and decays over 50 days, for stock indices this effect decays over 10 days but is also much bigger.

Leverage effect was first discussed by Black and this effect is very important for options prices. One of the unanswered questions in this area is also whether volatility increases after prices drop or prices drop because of volatility increase. While Black claims that price drop is the cause of volatility increase since price dropping makes bankruptcy of company more probable and therefore stock more volatile, there is also another direction claiming that it makes more sense that increase of volatility makes the stock less attractive which decreases its price [13].

Some of the possible explanations include volatility feedback effect which explains increase of future volatility by market operators triggering sell orders which makes more shares available and therefore decreases the price of the asset.

While some consider volatility feedback effect and leverage effect two explanations of the same underlying phenomena, others claim that volatility feedback effect explains why an increase in the volatility results in a negative return and leverage effect explains negative return causing an increase in the volatility. Lastly, Bekaert and Wu [11] proposed alternative explanation for volatility causing instability in financial markets. Namely, they proposed that those two effects may be interacting and together producing excess volatility, volatility persistence and volatility smile.

There are numerous observations related to the leverage effect most of which are collected in [19] to which I refer the interested reader.

2.3.4 (Asymmetric) vertical dependence

Data appears to have different characteristics depending on the frequency of intervals over which it was observed. For example, the leptokurtic distribution is more observable for longer time intervals. This also leads to different preference when it comes to models since while some models perform better on the shorter horizons, others do so for the bigger ones. High-frequency trading data has become widely available thus leading to the recent increase in the models based on the intraday data. There are many studies that came to the conclusion that high-frequency volatility models outperform (G)ARCH-like models [34]. This is, as expected, especially true for short horizons. However, usually best-performing models are those that combine low frequency and high-frequency data.

What is also true is that high-frequency data results in the better estimates of the actual volatility. The best possible frequency of the data depends, however, on the market. The data should not be too frequent since this leads to noisy estimates. In the developed and highly liquid market such as the US one, one usually takes 5-minute intervals [21]. Another method frequently applied to the high-frequency estimates is to take the mean of a longer time period such as daily or weekly one since for a small sample, the sample mean is very noisy estimate of the true mean. This thesis, however, does not consider such models and evaluations because of the lack of the access to such data.

Chapter 3

Volatility forecasting

3.1 Types of volatility models

Poon and Granger compared 93 studies [21] on volatility forecasting tested out-of-sample, looking at various assets among many time horizons. Based on those results, it is no surprise that option-implied volatility is ”a must” in industry since their conclusion was that its forecasts are more accurate then those of time series models. Even though among time series models none appeared to be a lot better than others, according to them, historical volatility appears to perform slightly better than generalized autoregressive conditional heteroscedasticity. Those are followed by, at the last place, stochastic volatility models.

As noted in the paper, there are four dominating types of volatility models:

1. **Historical volatility models.** As the name says, historical volatility considers changes in the previous time periods when forecasting future volatility. The calculation itself may be based on intraday changes or changes from one closing price to the next one.

   Some of the historical volatility models are simple historical averages, random walk models, autoregressive moving average models as well as exponential smoothing models.

   The general formula for historical volatility models is:

   $$\hat{\sigma}_t = \varphi_1 \sigma_{t-1} + \varphi_2 \sigma_{t-2} + ... + \varphi_\tau \sigma_{t-\tau}$$

   where $\hat{\sigma}_t$ is the expected standard deviation at time t, φ the weight parameter and σ is the historical standard deviation for period indicated in the subscript.

Most of the historical volatility models are considered naïve since they do not directly model most of the stylized facts about the volatility.

Another disadvantage of historical volatility models is that they are largely affected by large outliers in the past. Namely, the volatility they forecast will be affected by the outlier as long as it is in the volatility estimation period. One of the ways to avoid this is to truncate the outliers by imposing a cap on the largest values. This is of course done if the outliers really represent the outliers[1] or, in other words, exception that is not likely to happen again anytime soon [21].

While most of the historical models are very simple, they can also be quite sophisticated like realized volatility model of Andersen, Bollerslev, Diebold and Labys [3].

2. **(G)ARCH-like models.** As previously mentioned, Engle introduced autoregressive conditional heteroscedasticity (ARCH) model in 1982 for which he received a Nobel prize in 2003. Its more popular generalization, generalized autoregressive conditional heteroscedasticity model, or shortly GARCH, was introduced by his PhD student Bollerslev in 1986 and, since then, many extensions of the (G)ARCH have been proposed that are meant to account for various stylized facts of volatility.

   The meaning behind the name (G)ARCH is exactly one of its stylized facts. Conditional heteroscedasticity means that data has time-dependent varying characteristic, while autoregressive means that the model is estimating volatility at time t based on the information known up to the time point $t-1$.

   Word heteroskedasticity comes from the ancient Greek language and it means "different dispersion". It refers to the variance of residual or error term that varies. One of the main problems in calculating volatility using Black and Scholes formula briefly explained later is that it is based on an assumption of the constant volatility. Today, it is widely accepted that volatility is not constant. While Black-Scholes formula leads to elegant closed-form formula for the volatility, it makes an assumption which is in general not true.

3. **Stochastic volatility models.** Stochastic volatility models have attracted interest recently as one of the alternatives to the popular GARCH model. Stochastic volatility models, similarly to GARCH models, aim to correct constant volatility assumption of implied volatility models, by considering volatility a random process governed by state variables. For example, changes in price level of the underlying

[1] Frequent large numbers should not be called outliers.

security, tendency of volatility to revert to some long-run variable and variance of the volatility process itself have an effect on the volatility forecasting. Stochastic volatility models are explained with following formulas:

$$r_t = \mu + \varepsilon_t$$
$$\varepsilon_t = z_t \exp(h_t/2)$$
$$h_t = \omega + \sum_{j=1}^{p} \beta_j h_{t-j} + v_t$$

where return r_t is defined as usually and where v_t is an innovation term and variables v_t and z_t can be correlated.

Another advantage of stochastic volatility models is that they have a noise term and are therefore less affected by large outliers than the ARCH models. On the other side, ARCH models are less affected by large outliers than historical volatility models [21].

The survey among papers done by Poon and Granger in [21] also leads to the conclusion that despite added complexity, stochastic models do not provide better volatility forecasts. These models are not further discussed in the thesis.

4. **Option-implied models.** Option-implied models are based on the implied volatility which is briefly described in the chapter 2. This is a very popular approach because of the belief that option traders, and therefore option prices, contain in itself all the information including the information extracted from past information. Implied volatility is, therefore, future volatility expected by the options market. When there are multiple options, having the same expiration, listed on the same asset, each can have different implied volatility which is known as volatility skew or smile.

   If by g we denote the model for pricing options, by c the price of the option, by S the price of the underlying asset, by X the exercise price, by σ the volatility, by R a risk-free interest rate and by T the time to option maturity Black-Scholes formula claims the following relation

$$c = g(S, X, \sigma, R, T)$$

   which we can revert to get the volatility.

   Black-Scholes implied that standard deviation tends to be higher than actual volatility which is usually linked to volatility risk premium leading to historical volatility sometimes being used for calibration. One additional insight is that low volatility is often underestimated and high volatility overestimated. Another fact about implied volatility forecasts is that they work worse for smaller markets such as the one in Sweden [21].

3.2 (G)ARCH models and alterations

3.2.1 Some notation

A **white noise** process consists of random variables that cannot be predicted. Therefore, a time series is white noise if the variables it consists of are independent and identically distributed with a mean of zero.

A **residual** is defined as the difference between the observed value of a variable and its predicted value. Ideally, successive residuals are uncorrelated with each other and, therefore, they constitute a white noise time series. If that is the case, the model has predicted all the possible components and what is left is the unpredictable white noise.

A **stationary time series** is time series whose statistical properties are constant over time. These properties can include mean, variance, autocorrelation etc.

Skewness is a measure of the lack of symmetry in the data.

Kurtosis is a measure of heavy-tailedness of the data compared to the normal distribution.

3.2.1.1 Distributions

Even though, as already mentioned, asset returns do not follow **normal distribution**, it still has an important spot in finance and is basis for the other distributions. The probability density function of the standard normal distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where μ is the mean or expectation of the distribution, while σ is its standard deviation.

Student's t-distribution is a distribution that is more representative of the real financial data. It is symmetric and bell-shaped like the normal distribution but has heavier tails and therefore is more likely to produce results that fall far from its mean. Student's t-distribution has parameter ν

which indicates the shape of a curve. The larger the ν is, the more peaked the curve becomes, which implies higher kurtosis or in other words heavier tails.

The probability density function of the **student's t-distribution** is:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where ν is, as mentioned, the number of degrees of freedom and Γ is the gamma function. In this thesis the constraint $\nu > 2$ is usually assumed to ensure that the second order moment exists.

While considering student's t-distribution gives us needed heavier tails, it does not account for negative skewness. If we want to model asymmetry in the distribution, we can represent data with **skewed student's t-distribution** which has additional parameter λ that indicates skewness. λ is between -1 and 1 and $\lambda < 0$ indicates negative skewness. There are various ways in which one might introduce skewness to the distribution and I would refer the interested reader for this specific case to the [17].

Generalized error distribution, also known as the exponential power distribution, is a parametric family of symmetric distributions. The probability density function of the standardized **generalized error distribution** is described with:

$$f(\eta;\nu) = \frac{\nu e^{-\frac{1}{2}|\eta/\lambda|^{\nu}}}{\lambda 2^{1+\frac{1}{\nu}}\Gamma\left(\frac{1}{\nu}\right)}$$

where $\nu > 0$ is the shape parameter and

$$\lambda = \left(\frac{\Gamma\left(\frac{1}{\nu}\right)}{4^{\frac{1}{\nu}}\Gamma\left(\frac{3}{\nu}\right)}\right)^{\frac{1}{2}}$$

3.2.2 (G)ARCH model

Assume that we want to predict the return r_t of an asset at the time t having all of the information up to that time point available. Denote the predicted mean return at time t with r_t. The prediction will most likely not be correct and we will have some remaining residual ε_t.

$$r_t = \mu_t + \varepsilon_t$$

Similarly, we model the volatility as the expected variance given all the information up to the point t. Volatility is not directly observable but is related to the prediction error since if the prediction works well, the residual in the volatility is just the volatility multiplied with some white noise.

$$\varepsilon_t = \sigma_t \cdot z_t$$

where z_t is a strong white noise process and σ_t is a time-dependent standard deviation. The series σ_t^2 is then modelled as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i > 0$. In the literature α_0 is often named ω and those two notations will be used in the thesis interchangeably.

The lag length $p \geq 0$ is part of the model specification. It can be decided using the Box-Pierce or its more popular modification Ljung-Box test for autocorrelation significance.

In practice, the generalized version of ARCH, GARCH, is more commonly used. GARCH model is ARCH that incorporates a moving average component together with the autoregressive component. As already mentioned, Tim Bollerslev, the student of Robert Engle himself, extended the ARCH model to allow volatility to have an additional autoregressive element within itself. The GARCH(p,q) model equation is as follows

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

Besides p-period lags of residuals, GARCH adds q-period lags of variances for predicting the current variance. Its most basic example is GARCH (1,1) model. In the formula above, ω is a constant term, β is an autoregressive parameter and α is a moving-average parameter. Therefore we could say that complete GARCH model consists of three components: a mean model, a volatility process and a distribution of the standardized residuals.

3.2.3 GARCH(1,1)

A very basic but very effective and widely used model is GARCH(1,1) described with the following equations

$$r_t = \mu + \varepsilon_t$$

$$\varepsilon_t = \sigma_t e_t$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

As before, some conditions need to be met, namely all of its parameters ω, α, and β need to be non-negative. Secondly, in order to ensure mean-reverting to the long-run variance, we need $\alpha + \beta < 1$. The long-run variance is equal to $\frac{\omega}{1-\alpha-\beta}$. As follows from the definition, the larger the α, the bigger the immediate impact of residuals also called prediction errors or shocks. For a fixed α, bigger β means longer duration of the impact or in other words higher persistence for both low and high volatility periods.

3.2.3.1 Parameter estimation

Parameter estimation of the model GARCH(1,1) is a simple optimization process and is done via maximum likelihood estimation in a usual way. Namely, to perform the maximum-likelihood estimation, one makes distributional assumptions on e_t.

If e_t follows a standardized Gaussian distribution, we have:

$$LLF = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{N}\log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{N}\frac{\varepsilon_t^2}{\sigma_t^2}$$

If e_t follows a student's t-distribution, we have:

$$LLF = N\log\left(\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)\Gamma(\frac{\nu}{2})}}\right) - \frac{1}{2}\sum_{t=1}^{N}\log\sigma_t^2 - \frac{\nu+1}{2}\sum_{t=1}^{N}\log\left(1+\frac{\varepsilon_t^2}{\sigma_t^2(\nu-2)}\right)$$

3.2.4 RiskMetrics model

In 1996, JP Morgan suggested special case of exponentially weighted moving average (EWMA) method for forecasting volatility. Their idea of estimating value at risk, or shortly, VaR has, since then, become very popular. The usual EWMA formula is

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2$$

In general, EWMA has some positive properties such as a greater weight upon more recent observations, but it also has drawbacks such as the fixed decay factor that introduces subjectivity into the estimation. The decay factor is λ and is sometimes also known as the smoothing constant. This factor determines the exponentially declining weighting scheme of the observations giving bigger weight to the more recent information instead of as in the more naive models giving the, for example, same weight to all the previous information. A high λ naturally indicates slower decay, or, in other words, longer persistence. Even though the optimal λ varies by the asset class, the overall optimal parameter proposed by RiskMetrics uses a lambda of 0.94 which is appropriate for analysing daily data and λ of 0.97 for monthly data. This value was found to minimize the mean squared error of volatility forecasts for asset prices [38].

EWMA is actually a special case of GARCH(1,1) where $\omega = 0$ and $\alpha + \beta = 1$. The difference between EWMA and GARCH is that GARCH includes the additional term for mean reversion.

3.2.5 EGARCH

Exponential GARCH, or shortly EGARCH, was introduced by Nelson in 1991 [39]. This model is based on the log-variance instead of the variance itself and is introduced in order to address the asymmetric shock effect by adding a conditional component. Another advantage of EGARCH is that, since it is based on the log-variance, it does not need non-negativity constraints on α and β. This means that maximum likelihood optimization is faster during model fitting.

As with the GARCH, we have parameter p which is the order of the symmetric innovation, q which is the order of the lagged (transformed) conditional variance, but now we also have parameter o which is the order of asymmetric innovation and, as already mentioned, this parameter models one of the stylized facts about volatility - leverage effect.

EGARCH is explained with the following equation:

$$\ln \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i(|e_{t-i}| - \sqrt{2/\pi}) + \sum_{j=1}^{o} \gamma_j e_{t-j} + \sum_{k=1}^{q} \beta_k \ln \sigma_{t-k}^2$$

where $e_t = \varepsilon_t/\sigma_t$. If $\gamma < 0$, then negative shocks have a bigger impact on future volatility than the positive shocks.

3.2.6 (F)IGARCH

One big downside of the GARCH model is its short memory. Another way to see this downside of GARCH is to consider one of the best performing GARCH models - GARCH(1,1). One forward iteration of GARCH(1,1) gives

$$\begin{aligned}
\sigma_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\
&= \omega + \alpha\varepsilon_{t-1}^2 + \beta(\omega + \alpha\varepsilon_{t-2}^2 + \beta\sigma_{t-2}^2) \\
&= \omega + \beta\omega + \beta^2\omega + ... + \alpha \sum_{i=1}^{\infty} \beta^{i-1} r_{t-i}^2 \\
&= \frac{\omega}{1-\beta} + \alpha \sum_{i=1}^{\infty} \beta^{i-1} r_{t-i}^2
\end{aligned} \tag{3.1}$$

This means that volatility shock declines at a single exponential rate β which implies picking up short-run autocorrelation but not so easily longer cycles.

Fractionally Integrated GARCH, or shortly, FIGARCH was proposed by Baillie, Bollerslev and Mikelsen in 1996 [9] and it is meant to capture long-run dynamic dependencies in the conditional variance. Having its basis on ARFIMA type representation, FIGARCH is just an extension of the IGARCH model that allows fractional orders of integration in the autoregressive polynomial. One of the stylized facts of financial time series is long-memory which can be seen in autocorrelation plots. While most of the models capture persistence, models like GARCH and GJRGARCH have memories that are too short to model financial time series [21].

The GARCH(p,q) model can be rewritten to the equivalent ARMA-type representation [30] which is :

$$[1-\alpha(L)-\beta(L)]\varepsilon_t^2 = \alpha_0 + [1-\beta(L)](\varepsilon_t^2-\sigma_t^2)$$

where $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + ... + \alpha_q L^q$ and $\beta(L) = \beta_1 L + \beta_2 L^2 + ... + \beta_p L^p$ are lag operators. The integrated GARCH(p,q) process is then

$$[1-\alpha(L)-\beta(L)](1-L)\varepsilon_t^2 = \alpha_0 + [1-\beta(L)](\varepsilon_t^2-\sigma_t^2)$$

and fractionally integrated GARCH is

$$[1-\alpha(L)-\beta(L)](1-L)^d\varepsilon_t^2 = \alpha_0 + [1-\beta(L)](\varepsilon_t^2-\sigma_t^2)$$

where $(1-L)^d$ is the fractional differencing operator.

3.2.7 Other famous (G)ARCH-like models

3.2.7.1 HARCH

Heterogeneous ARCH process, or shortly, HARCH describes the conditional variance as a function of the square of the sum of lagged innovations, or the squared lagged returns, over different horizons.

Variance is described with

$$\sigma_t^2 = \omega + \sum_{i=1}^{m} \alpha_{l_i} \left(l_i^{-1} \sum_{j=1}^{l_i} \varepsilon_{t-j}^2 \right)$$

Note that lag-1 HARCH is identical to ARCH(1), but, for example if lags are equal to [1,5,22] the model is

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_5 \left(\frac{1}{5} \sum_{j=1}^{5} \varepsilon_{t-j}^2 \right) + \alpha_{22} \left(\frac{1}{22} \sum_{j=1}^{2} 2\varepsilon_{t-j}^2 \right)$$

3.2.7.2 TARCH

TARCH, sometimes also called ZARCH is the model that uses absolute values.

The volatility is described by the following equation:

$$\sigma_t = \omega + \alpha|\varepsilon_{t-1}| + \gamma|\varepsilon_{t-1}|I_{[\varepsilon_{t-1}<0]} + \beta\sigma_{t-1}$$

Using the similar principle we can build model with other powers for which the volatility equation would look like:

$$\sigma_t^k = \omega + \alpha|\varepsilon_{t-1}|^k + \gamma|\varepsilon_{t-1}|^k I_{[\varepsilon_{t-1}<0]} + \beta\sigma_{t-1}^l$$

where the conditional variance is $(\sigma_t^k)^{2/k}$.

3.2.7.3 GJRGARCH

The GJR-GARCH named after its innovators Glosten, Jagannathan and Runkle [20] allows the conditional variance to respond differently to the past innovations depending on whether they are positive or negative. This model is sometimes called Sign-GARCH and similarly to EGARCH, model is developed to address the asymmetric shock effect on volatility. γ parameter is introduced to account for the leverage effect and the model is described by

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 I_{[\varepsilon_{t-1}<0]} + \beta\sigma_{t-1}^2$$

where I is just an indicator function that takes the value 1 when its argument is true

$$I_{[\varepsilon_{t-1}<0]} = \begin{cases} 0, & \text{if } r_{t-1} \geq \mu. \\ 1, & \text{if } r_{t-1} < \mu. \end{cases} \tag{3.2}$$

3.3 Regime-switching models

Almost any financial time series experiences dramatic breaks over a sufficiently long period of time and there are many possible reasons for that such as wars, financial panics, changes in government policies, or most recently big pandemics. Many economic variables behave differently during economic downturns and regime-switching models are designed to account for that.

Hamilton proposed in [27] that model that describes volatility of a certain asset should, therefore, change accordingly. On the 9th of March 2020, volatility index hit the highest level since 2008 financial crisis due to the COVID-19 pandemic and possible influence of Russia-Saudi Arabia oil price war. Indeed market volatility has changed and one possible way of modeling volatility is that for data prior to 9th of March we use a model such as

$$y_t - \mu_1 = \varphi(y_{t-1} - \mu) + \varepsilon_t$$

and for data after 9th of March 2020 to use

$$y_t - \mu_2 = \varphi(y_{t-1} - \mu_2) + \varepsilon_t$$

where $\mu_2 < \mu_1$.

Forecasting the model in this way has to take into account that if the process has changed in the past, it might change in the future. Since the change in regime is not a foreseeable event, it should be seen as a random variable [27]. To model this, one needs the way to describe the probability of the changes between regimes.

Denote with s_t^* regime or state at the time t and let the model have two regimes, namely $s_t^* \in \{1, 2\}$. We can then group two of the equations above into

$$y_t - \mu_{s_t^*} = \varphi(y_{t-1} - \mu_{s_{t-1}^*}) + \varepsilon_t$$

We can then build a transition matrix where each p_{ij} denotes the probability that the regime j will be followed by the regime i.

In the following section we will discuss one such model and explain the details of regime-switching through a transition matrix.

3.3.1 MSGARCH

GARCH model was introduced to account for volatility clustering but it comes with a significant drawback. Volatility might have structural breaks, namely, GARCH does not change parameters if a big event that influences asset volatility happens. In other words, the pure GARCH model may fail to capture changes in the volatility dynamics. The very basic idea behind the Markov-switching GARCH is to decrease the long GARCH persistence by switching from one variance structure to another.

In the figure 3.1 we can see two GARCH model simulations with different parameters connected. On the left side we have a low volatility state and on the right side we have a high volatility state. If we try to estimate this model as a single-parameters model using GARCH we get a model which has alpha+beta¿1 which means that as time goes, volatility goes to infinity. This problem clearly illustrates the issue with structural breaks and a big downside of GARCH and in general single-regime models. Model that naturally arises is Markov-switching GARCH, or shortly, MSGARCH. Markov-switching GARCH allows parameters of GARCH to change over time according to a latent discrete Markov process.

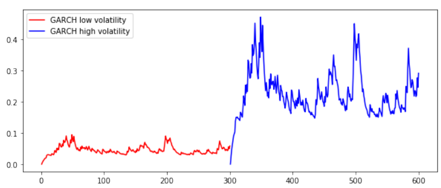

Figure 3.1: Structural break in volatility - simulation

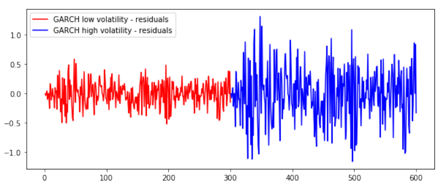

Figure 3.2: Structural break in residuals - simulation

Even though Quandt and Goldfeld were the first ones to introduce Markov-switching regression model, Hamilton's work [27] is better known. He extended Markov-switching regressions for AR processes and provided a nonlinear estimation filter.

Markov-switching models are applied in many social sciences. Hamilton [27] was the first one to model GDP with a switching process. What followed was modeling interest and exchange rates, monthly stock returns and even modeling US states as Democratic or Republican [45]. In health sciences, models are used to model rapid cycling bipolar disorder as well as incidence rate of infectious disease in epidemic and non-epidemic states.

Markov-Switching GARCH model consists of K regimes where each is a (G)ARCH-like model of unique parameters. Namely,

$$
\begin{aligned}
h_{1,t} &= \omega_1 + \alpha_1 y_{t-1}^2 + \beta h_{1,t-1} \\
h_{2,t} &= \omega_2 + \alpha_2 y_{t-1}^2 + \beta h_{2,t-1} \\
&\vdots \\
h_{K,t} &= \omega_K + \alpha_K y_{t-1}^2 + \beta_K h_{K,t-1}
\end{aligned}
$$

As already mentioned, the idea of MSGARCH is to allow parameters of the GARCH model to vary over time according to a latent dicrete Markov process instead of being fixed. Well-known problem of GARCH model is that it adapts slowly and MSGARCH is an attempt to solve this downside.

There are K separate GARCH processes, each being for one regime of the unobserved Markov chain. Let y_t be variable that is observed at time t. We assume $\mathbb{E}[y_t] = 0$ and $\mathbb{E}[y_t y_{t-l}] = 0$ for all $l \neq 0$ or in other words y_t is zero-mean and is not serially correlated.

If we denote by $\mathcal{I}_{t-1}$ the information available as of time $t-1$, then the general Markov-switching GARCH specification can be expressed as:

$$y_t | (s_t = k, \mathcal{I}_{t-1}) \sim \mathcal{D}(0, h_{k,t}, \xi_k)$$

where $\mathcal{D}(0, h_{k,t}, \xi_k)$ is a continuous distribution with zero mean, time varying variance $h_{k,t}$ and additional shape parameters ξ_t.

There are two popular aproaches of the dynamics of the variable s_t which represents the regime at time t.

1. First-order ergodic homogeneous Markov chain model [24]
2. Mixture of GARCH models [25]

3.3.2 First-order Markov chain

Following this approach, it is assumed that s_t evolved according to transition probability matrix P

$$P = \begin{bmatrix} p_{1,1} & \cdots & p_{1,K} \\ \vdots & & \vdots \\ p_{K,1} & \cdots & p_{K,K} \end{bmatrix}$$

where $p_{i,j} = P[s_t = j|s_{t-1} = i]$ is the probability of transition from state $s_{t-1} = i$ to state $s_t = j$. Each of the entries in the matrix need to satisfy $0 < p_{i,j} < 1$, $\forall i, j \in \{1, ..., K\}$ and $\sum_{j=1}^{K} p_{i,j} = 1$, $\forall i \in \{1, ..., K\}$. We also have $\mathbb{E}[y_t^2|s_t = k, \mathcal{I}_{t-1}] = h_{k,t}$, meaning that $h_{k,t}$ is the variance of y_t conditional on $s_t = k$.

3.3.3 Independent states

In this case s_t is sampled independently over time from a Multinomial distribution with vector of probabilities $\omega = (\omega_1, ..., \omega_K)^T$, or in other words $P[s_t = k] = \omega_k$.

Estimation of MSGARCH as well as mixture of GARCH can be done with regular maximum likelihood estimation. Another possible way is to use Bayesian Markov chain Monte Carlo (MCMC) technique.

3.4 Multifractal volatility models

Besides the mentioned slow adaption to new regimes, one big downside of the ARCH model is that it does not explain return phenomena at different frequencies. Multifractal models attempt to overcome this drawback by assuming that volatility is determined by components that have different degrees of persistence which randomly switch over time.

Fractals are patterns that are self-similar across different scales and they are created through a repetition of a single, simpler process. A fractal is basically a shape that can be separated into parts, each of which is the smaller version of the whole. The famous examples include Cantor set where one removes the middle third of the interval or the Koch flake. Mandelbrot [35] contains a great introduction to fractal theory and its applications in the natural sciences. The key step between fractals and multifractals is lengthening or shortening the horizontal time axis so that the pieces are stretched or squeezed. The simplest multifractal is Mandelbrot's binomial measure on $[0, 1]$ and it can be derived as the limit of a multiplicative cascade as explained as follows. Let μ_0 be the uniform probability measure on the unit interval and let m_0 and m_1 be two positive real numbers adding up to 1. In the first part of the cascade, we define a new measure denoted by μ_1 whose density is a step function and which uniformly spreads the mass m_0 on the left half of the subinterval and m_1 on the right half of the subinterval. Following the cascade in the similar way, we split the left interval into two subintervals of the equal length and allocate to the left subinterval, namely $[1, 1/4]$ a fraction m_0 of $\mu_1[0, 1/2]$ while the right subinterval gets allocated m_1. Continuing this process further, we get an infinite sequence of measures that weakly converges to the binomial measure. Note that the mass is preserved since $m_0 + m_1 = 1$. This measure can be generalized by splitting interval into arbitrary many subintervals b at each step of the process and by randomizing the allocation of mass between subintervals.

Multifractality has been observed in many systems, one important being financial markets. The ease of access and availability of huge amounts of data at different frequencies has lead to the recent increase in the number of proposed methods. Some of the most popular models summarized in [30] are the Multiplicative cascade models whose main idea was already briefly explained; Multifractal model of asset returns, or shortly MMAR, that incorporates two important elements - long tails and long-dependence; Multifractal random walk, or shortly MRW, that models volatility such that it can be reduced as an exponential of a long memory process; Exponentials of long memory processes described in [50]; Agent based models and one of the most popular ones - Markov-switching Multifractal which will be explored further.

Another emerging idea of interest which considers multifractality is based on defining multifractal volatility and including it in the already existing models. One of the recommended proxies for volatility [28] is

$$\sigma_t = \left[\frac{1}{n} \sum_{j=1}^{n} \frac{r_j^2}{RV_j^2} \right] RV_t, \tag{3.3}$$

where n is the number of days in the sample and RV_t is the realized variance calculated by taking the sum of squared intraday returns.

Wei and Wang constructed the formula for multifractal volatility [52]

$$\sigma_t = \left[\frac{\sum_{j=1}^{n} r_j^2}{\sum_{\ell=1}^{n} \Delta\alpha_\ell} \right] \Delta\alpha_t, \tag{3.4}$$

where $\Delta\alpha_t$ is the singularity width from the intraday data on the t-th day.

One can then construct analogous definitions to the one widely accepted such as

$$\sigma_t = \left[\frac{1}{n} \sum_{j=1}^{n} \frac{r_j^2}{\Delta\alpha_j} \right] \Delta\alpha_t \tag{3.5}$$

and

$$\sigma_t = \left[\frac{\sum_{j=1}^{n} RV_j^2}{\sum_{\ell=1}^{n} \Delta\alpha_\ell} \right] \Delta\alpha_t \tag{3.6}$$

$\Delta\alpha_t$ can then be viewed as a measure of the asset risk [30].

This multifractal volatility definition can be incorporated into existing volatility models for volatility forecasting. The ARFIMA(1, d, 1) model is defined as

$$(1 - \varphi L)(1 - L)^d [RV_t - \mu] = (1 - \theta L)\varepsilon_t$$

where L is the lag operator, μ is the mean of the realized volatility, coefficients d, φ and θ are fixed and unknown, and ε_t is Gaussian white noise with zero mean and variance σ_ε^2. Based on this model, one can construct multifractal volatility model such as

$$(1-\varphi L)(1-L)^d[MFV_t-\mu]=(1-\theta L)\varepsilon_t$$

3.4.1 Markov-switching multifractal

As already mentioned, Markov-switching multifractal, or shortly, MSM is a model developed by Calvet and Fisher [15] that incorporates volatility components of heterogeneous durations. In summary, it is discrete-time Markov process with multi-frequency stochastic volatility even though it can be also specified in continuous time [15].

The specification is pure regime-switching and has multiple frequencies, arbitrarily many states, a dense matrix and requires only four parameters.

MSM volatility is then derived by multiplying a finite number first-order Markov components which are identical except for their switching probabilities which follow approximately geometric progression. The construction has closed-form likelihood.

What the authors claim is that compared to GARCH(1,1), MSM has higher likelihood than GARCH for all four daily currencies they tested it on both in- and out-of-sample. Since both models have the same number of parameters, using standard selection criteria such as BIC and AIC would also lead to picking MSM. Out-of-sample MSM has the accuracy of GARCH on short horizons such as one day and performs better at longer horizons such as 20 to 50 days, while authors claim that MSM outperforms MS-GARCH and FIGARCH out-of-sample.

The difference in regime-switching for MS-GARCH and MSM is that MS-GARCH uses regime-switching only for low-frequency events while MSM uses linear autoregressive transitions at medium frequencies and a thick-tailed conditional distribution of returns. On the other hand, MSM captures long-memory features, intermediate frequency volatility dynamics and thick tails in returns.

What is interesting about MSM is that, usually, one uses small number of regimes in the model since it is popular opinion that regimes do not switch frequently.

3.4.1.1 Definition

Let P_t be the price of an asset or exchange rate. Volatility is driven by a first-order Markov state vector which consists of $\bar{k}$ positive real number components:

$$M_t=(M_{1,t};M_{2,t};...;M_{\bar{k},t})$$

The components of M_t have the same marginal distribution but evolve at different frequencies. Assume that the volatility state vector has been constructed up to the point $t-1$

$$M_{t-1}=(M_{1,t-1};M_{2,t-1};...;M_{\bar{k},t-1})$$

Now M_t is built as follows. Each $M_{k,t}$ is drawn from some fixed distribution M with probability γ_k and is otherwise equal to its previous value $M_{k,t-1}$.

$$\begin{aligned} M_{k,t} \text{ drawn from distribution M} &\quad \text{with probability } \gamma_k \\ M_{k,t} = M_{k,t-1} &\quad \text{with probability } 1 - \gamma_k \end{aligned} \tag{3.7}$$

Switching events and new draws are independent across k and t. It is required that $M \geq 0$ and $\mathbb{E}(M) = 1$. While multipliers $M_{k,t}$ have the same distribution, their dynamics, i.e. their transition probabilities γ_k are different. All of the components are mutually independent which contributes to the parsimony of the model.

Stochastic volatility is modeled by

$$\sigma(M_t) = \bar{\sigma} \left(\prod_{i=1}^{\bar{k}} M_{k,t} \right)^{1/2}$$

where $\bar{\sigma}$ is a positive constant. For returns it holds then

$$r_t = \sigma(M_t)\varepsilon_t$$

where $\{\varepsilon_t\}$ are i.i.d standard Gaussians $\mathcal{N}(0,1)$. The transition probabilities are defined as

$$\gamma_k = 1 - (1 - \gamma_1)^{b^{k-1}}$$

where $\gamma_1 \in (0,1)$ and $b \in (1,\infty)$.

The transition probabilities of components with low frequencies grow approximately geometrically at rate b, while at high frequencies, the rate is smaller. Therefore, parameters $\gamma_{\bar{k}}$ and b alone define the transition matrix.

MSM is defined with one number - $\bar{k}$ which is the number of frequencies. Considering distributions, one simplest solution is binomial distribution where M takes only two values m_0 or m_1. In the simplest case, those two values occur with the equal probability. Since $\mathbb{E}(M) = 1$ it must be that $m_1 = 2 - m_0$. The parameters are then

$$\psi = (m_0, \bar{\sigma}, b, \gamma_{\bar{k}})$$

for binomial MSM, where, as previously mentioned, m_0 defines the distribution of multipliers, $\bar{\sigma}$ is the unconditional standard deviation of returns and $\gamma_{\bar{k}}$ and b define together the set of switching probabilities.

While this is the simplest possible MSM, it already produces good results. On the other hand there exist more complex models such as multinomial MSM and lognormal MSM. I refer the interested reader to the authors' book [15].

3.5 Discussion, future improvements and remarks

One of the very popular directions to go when forecasting volatility is to incorporate **multivariate models**, i.e. to look and predict multiple time series at the same time. While for univariate time series we consider variation of only one variable at the time and deal with one-dimensional data, for multivariate time series we take into account multiple variables and their relation. Multivariate GARCH attempts to jointly capture volatility dynamics in several financial markets and language R contains great number of packages for estimating multivariate GARCH models such as DCC.

Many believe that one way of improving forecasting power is to use exogenous variables. There are many possible directions in which this research can go since volatility might be linked to macroeconomic news, interest rates, trading volume, recessions or seasonal factors. What is unclear is how to use this specific information to improve volatility forecasts [21].

There is almost no discussion about the tight relationship between asset prices and market sentiment. This relationship is a well-studied topic, especially in the industry with financial companies having more and more NLP researcher roles. Therefore, the advancing in the area is very hard because of the lack of computational resources available to the independent researchers for analyzing huge amount of text. One promising alternative to this is to use Google trends. Google trends are new, promising and accessible data to anyone. It is a website by Google that analyzes the popularity of specific search queries in Google Search. It can be restricted to the region, period of time and even where it was searched for. This includes Image search, News search, Google shopping and YouTube search.

It comes as no surprise that volume of search is correlated with volatility [26]. One very simple example of correlation between S&P500 price and number of times it was searched through Google can be seen in figure 3.3

The Pearson correlation coefficient measures the linear relationship between two datasets where it is assumed that each dataset is normally distributed. The Pearson value is between -1 and 1, where 0 indicates no correlation. Like other correlation coefficients, this one varies between -1 and +1 with 0 implying no correlation. The Pearson correlation coefficient between S&P500 asset price and volume of search in Google engine is 0.66 which indicates very positive correlation.

Most of the papers do not take into account how models perform depending on the time horizon or in general the period of time they were evaluated. The topic of which models performs better when there is low volatility and which when there is high volatility is not well-covered. Even more interestingly, papers do not usually cover how models complement each other cross-sectionally and through time [21]. This is a big problem since performance of the models usually depends on the underlying asset and market conditions.

Models perform differently depending on the state of volatility. For example, comparing EGARCH and GJRGARCH out-of-sample, EGARCH overestimated volatility much more than GJRGARCH did even when overall based on the MSE, the model performed better. On the considered data EGARCH overestimated volatility 9-11% more than GJRGARCH. On the other hand, the EGARCH was usually closer

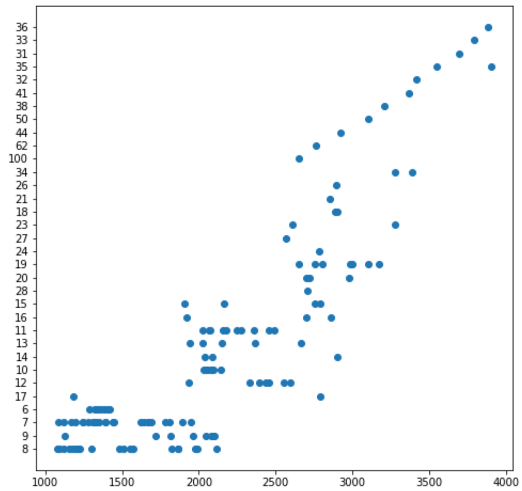

Figure 3.3: Correlation between S&P500 price and the number of times it was searched on Google

to the true volatility when they both underestimated. This lead to the specific weighting of the predictions be a slightly better model then the individual models alone. This, however, did not change the fact of (G)ARCH-like models adapting very slowly to the new regimes. As will be seen in the next chapter, they have all hugely underestimated volatility during the burst of the pandemic.

As already mentioned, simple combination of (G)ARCH models when combined with concrete weighting performs better than individual models. However, difference in the ensemble model compared to the individual ones is outshined by the slow adaptation to the new regime. If, however, combination could be combined with regime-switching, this could lead to good results.

Chapter 4

Forecast evaluation, empirical results and analysis

Backtesting is the method meant for testing predictive model on historical data. In other words, in the backtesting step results predicted by the model are compared with the actual historical data. Usually, historical data is separated into two parts, in-sample and out-of-sample data. In-sample data is used for model fitting, while out-of-sample data is reserved for backtesting and evaluating the model performance.

4.1 In-sample results

4.1.1 Notation

Ljung-Box test is a statistical test determining whether any of a group of auto-correlations of a time series is different from zero. In other words, this test tests the overall randomness of a model based on the number of lags instead of testing randomness at each distinct lag.

The Hurst exponent is one popular measure of long-term memory of time series. Usually if the value of Hurst exponent is around 0.5, series are considered random. If the value of Hurst exponent is close to 0 series are mean-reverting and values near 1 indicate a trending series.

Ideally, we would like to choose parsimonious models over complex ones or in other words use as few parameters as possible.

In order to decide whether to drop specific parameter or not, we can use hypothesis test, better known as null-hypothesis. In the case of parameters, the null hypothesis is whether the parameter value is 0. If the hypothesis cannot be rejected, parameter should not be used in the model. The significance level is usually set to 0.05, meaning that the probability of observing the results in the data is 5%.

With ***p*-value**, one usually denotes the probability of obtaining the observed results of a test, assuming that the null hypothesis is true. In other words, it defines the chance that results happened by chance.

T-statistic is the statistical measure which is calculated as the estimated parameter value subtracted by its expected mean value and then divided by its standard error. T-statistic is supposed to measure how distant in terms of standard errors is the estimated parameter away from zero. The larger the distance, the more likely the parameter is not zero and therefore should be included in the model.

4.1.2 Data description

Descriptive statistics of the data used for performance evaluation in the time period between January 2010 and March 2021 is shown in the table below

Description	JNJ	OMX	ETH-USD	SPX
Mean	0.0	0.0	0.0	0.0
Var	1.14	1.45	41.97	1.21
Skew	-0.32	-0.78	38.12	-0.77
Kurtosis	17.66	18.20	31424.30	26.27
Hurst(returns)	$2 \cdot 10^{-4}$	$5.25 \cdot 10^{-5}$	$-3 \cdot 10^{-3}$	$2 \cdot 10^{-3}$
Q-stat(returns)	0.0	0.0	0.0	0.0
ADF(returns)	$2.8 \cdot 10^{-26}$	$4.74 \cdot 10^{-23}$	$4.65 \cdot 10^{-12}$	6.36%
Daily vol	1.28%	1.21%	6.48%	1.1%
Monthly vol	5.87%	5.53%	29.7%	5.04%
Annual vol	20.35%	19.15%	102.87%	17.47%

S&P500 (SPX or SPY) is a stock index based on the 500 largest companies listed on the New York Stock Exchange and Nasdaq. This index is considered to be the overall health of the economy and β measure is also defined in regard to this index.

OMX (Stockholm 30) is a stock index like S&P500 but for the Stockholm Stock Exchange. It is a capitalization weighted index of the 30 most-traded stock on the Nasdaq Stockholm stock exchange. It is slightly more volatile than S&P500, and is, naturally, traded less.

Ethereum is a decentralized, open-source blockchain with smart contract functionality and its native cryptocurrency is **Ether** (ETH). It is the second-largest cryptocurrency. It is chosen for model evaluation as one of the harder-to-predict assets because of its volatility.

Johnson & **Johnson** (JNJ) is one of the best-known names in the health industry. Johnson & Johnson is the example of a stable stock without much volatility.

4.1.3 Information criteria

Information criteria is a method for selecting a model. It is used as a measure of trade-off between goodness of fit and complexity of the model. In simpler terms, when fitting models it is possible to increase the likelihood by adding extra parameters and therefore increasing complexity of the model. This will usually result in overfitting and poor results on the new or test data. While information criteria considers the model likelihood, it also adds penalties for model complexity, i.e. if models A and B have the same likelihood values, but model A has less parameters, then model A has smaller information criteria score.

Two among most popular information criteria are Akaike information criteria and Bayesian information criteria.

4.1.3.1 Akaike information criteria

Akaike information criteria, or shortly AIC, is information criteria named after its inventor Akaike.

The main equation is the following one:

$$AIC = -2\ln(L) + 2k$$

where L is maximum likelihood estimation (log-likelihood) and k is the number of parameters.

4.1.3.2 Bayesian information criteria

Bayesian information criteria, or shortly BIC, is information criteria similar to AIC information criteria. The main difference is the amount of penalty they introduce for model complexity.

BIC usually leads to choosing a more parsimonious model since it imposes bigger penalties on parameters.

The BIC formula is:

$$BIC = -l\ln(L) + \ln(n)k$$

4.1.4 Comparison

Hyperparameter search for optimal models for the S&P500 data was performed. Based on AIC and BIC criteria, set of models among which the selection was performed contains GARCH, HARCH, FIGARCH, TARCH, EGARCH, GJRGARCH models of lag-values up to 10. There are two benchmark models included - RiskMetrics and GARCH(1,1). Hyperparameter search also includes search for the best possible distribution and mean model for all the mentioned models. Considered distributions are normal, student's t-distribution, skewed student's t-distribution and ged. Considered mean models are constant, zero, autoregressive model, heterogeneous autoregressive (HAR) model and least squares model. The period on which the models were selected is the period between 01.01.1995. and 31.12.2009. Reported are log-likelihood, AIC and BIC criteria for (G)ARCH-like models in the table below.

Model	mean	dist	AIC	BIC	LL
GARCH(p=1,q=1)	constant	t	10777.4	10808.6	-5383.69
RiskMetrics	constant	normal	58553.2	58578.1	-29272.6
FIGARCH(p=1,q=1,tr=500)	constant	skewt	10752.8	10796.5	-5369.41
FIGARCH(p=0,q=1,tr=1000)	constant	skewt	10755.8	10793.2	-5371.90
HARCH([1,7,23])	constant	skewt	10793.6	10837.3	-5389.82
HARCH([2,7,23])	constant	ged	10798.0	10835.4	-5392.99
GARCH(p=2,q=1)	constant	skewt	10760.0	10803.7	-5373.02
EGARCH(p=3,q=1,o=4)	constant	skewt	**10572.6**	10647.4	**-5274.30**
EGARCH(p=2,q=1,o=3)	zero	skewt	10581.4	**10637.5**	-5281.68
TARCH(p=0,q=3,o=2)	constant	skewt	10634.2	10690.3	-5308.09
TARCH(p=0,q=1,o=1)	zero	skewt	10643.8	10675.0	-5316.91
GJRGARCH(p=0,q=3,o=2)	constant	skewt	10642.0	10698.1	-5311.99
GJRGARCH(p=0,q=1,o=1)	zero	skewt	10654.5	10685.7	-5322.25

The same selection is performed for the period between 01.01.2010. and 31.03.2021. and below is the table described model performance for that time period.

Model	mean	dist	AIC	BIC	LL
GARCH(p=1,q=1)	constant	t	6909.71	6939.45	-3449.85
RiskMetrics	constant	normal	35993.7	36017.4	-17992.8
FIGARCH(p=0,q=1,tr=1000)	constant	skewt	6875.24	6910.92	-3431.62
HARCH([3,6,23])	constant	skewt	6896.60	6938.24	-3441.30
GARCH(p=2,q=1)	constant	skewt	6890.22	6931.85	-3438.11
GARCH(p=1,q=1)	constant	skewt	6895.08	6930.76	-3441.54
EGARCH(p=2,q=1,o=9)	constant	skewt	6733.55	6828.71	-3350.77
EGARCH(p=1,q=1,o=1)	zero	skewt	6759.12	6794.80	-3373.56
TARCH(p=0,q=2,o=2)	constant	skewt	6718.49	6766.08	-3351.25
TARCH(p=0,q=1,o=1)	zero	skewt	6725.18	6754.91	-3357.59
GJRGARCH(p=0,q=2,o=2)	constant	skewt	6760.40	6802.59	-3372.20
GJRGARCH(p=0,q=1,o=1)	constant	skewt	6766.90	6802.59	-3377.45

Model parameters and further descriptive statistics of models can be found in the respective appendices - In sample results 1995-2010, S&P500 and In sample results 2010-2021, S&P500.

MSM model in-sample results and parameters for the period of January 2010 - March 2021 are summarized in the table below.

$\bar{k}$	m_0	b	γ_k	σ	LL
1	1.80	1.50	0.05	1.36	-3668.81
2	1.70	11.44	0.11	1.53	-3539.59
3	1.62	6.79	0.12	1.79	-3505.62
4	1.59	4.68	0.14	1.46	-3493.90
5	1.59	4.93	0.14	1.16	-3494.80
6	1.48	2.63	0.16	1.54	-3489.04
7	1.43	2.13	0.19	1.38	-3489.48
8	1.40	1.79	0.16	1.25	-3488.93
9	1.38	1.65	0.18	1.15	-3489.03
10	1.35	1.53	0.18	1.16	-3488.84

4.2 Out-of-sample results

There are many things that in general need to be taken into consideration when comparing volatility forecasting out-of-sample results. If the evaluation is based on squared variance errors, the standard error will usually be quite large [21]. Well-known fact is that different cost functions will favor different models. While non-linear GARCH forecasts produce smaller mean absolute errors than the exponentially weighted moving average, tighter GARCH models produce more VaR violations.

One of the noted and sensible observations is that forecasting longer horizons leads to preferring simple models. For a larger than one-year horizons, Figlewski found that using low frequency data and historical volatility models produces best results [18].

Comparing different volatility forecasts can be tricky because forecasted values have to be compared against an ex post proxy of volatility, rather than its true, latent value. Patton [41] [40] has identified a class of loss functions that asymptotically generate the same ranking of models regardless of the proxy being used. Ranking ensures that models rankings achieved with proxies such as squared returns or realized volatility correspond to the ranking that would be achieved if forecasts were compared against the true volatility. Patton class rules out most of the traditionally used losses in the volatility forecasting literature such as mean absolute error. The two valid loss functions from the group of ones usually used are the quasi-likelihood loss which depends only on the multiplicative forecast error and mean squared error.

The quasi-likelihood loss is named for its close relation to the Gaussian likelihood and is defined as

$$L(\bar{\sigma}^2, h_{t|t-k}) = \frac{\bar{\sigma_t}^2}{h_{t|t-k}} - \log \frac{\bar{\sigma_t}^2}{h_{t|t-k}} - 1$$

where $\bar{\sigma_t}^2$ is an unbiased ex post proxy of conditional variance and $h_{t|t-k}$ is a volatility forecast based on $t-k$ information.

This loss metric will not be used and mentioned further in the thesis and we will focus on the mean squared error.

Mean squared error, or shortly MSE, is another possible measurement of back-testing results. It is defined as

$$L(\bar{\sigma}^2, h_{t|t-k}) = (\bar{\sigma_t}^2 - h_{t|t-k})^2$$

In this thesis we will compare models with mean squared error between squared returns and squared volatility. All the models are trained on returns data expressed in percentage change. Therefore, mean squared error might seem bigger than it is usually reported and when compared with results in other papers the right metric might be MSE divided by 10000 or 100 depending on a paper. It should also be noted that when squared errors are used as proxies the noise may be quite substantial.

4.2.1 Rolling window forecast

In the rolling-window forecast, in-sample data is used for model fitting after which 1-period ahead forecast is made. Continuing this process as time rolls, that is, fitting the model on existing data and providing the forecast for the next day, we can see how our model performs out-of-sample.

In the **expanding window forecast**, one starts with a sample of data and continuously adds new data. Expanding window forecast is more responsive to the most recent news, changes in economic cycles, etc.

Similarly to expanding window forecast, **fixed rolling window forecast** adds new data points as time moves forward. The difference, however, is that with each new addition we remove the oldest data point currently in the set. When opting for a fixed rolling window forecast, we have additional variable to determine - window size. If a window size is too wide, it may include obsolete data and, therefore, higher prediction bias. While wide window size would lead to overfitting, too small window size might not incorporate relevant data to the model which would give higher variance.

All of the selected models in-sample, i.e. from the previous section are evaluated out-of-sample on the data. We also added models that performed best in-sample on this time period. This was done in order to contrast in-sample results for periods containing 2008 crisis and also best possible model which was more recent. We conclude that models are quite robust and model selection does not depend much on the period on which it was selected.

Model	mean	dist	MSE
GARCH(p=1,q=1)	constant	t	103.61
RiskMetrics	constant	normal	123.57
FIGARCH(p=1,q=1,tr=500)	constant	skewt	104.18
FIGARCH(p=0,q=1,tr=1000)	constant	skewt	104.13
HARCH([1,7,23])	constant	skewt	108.40
HARCH([2,7,23])	constant	ged	101.01
HARCH([3,6,23])	constant	skewt	104.92
GARCH(p=2,q=1)	constant	skewt	104.95
GARCH(p=1,q=1)	constant	skewt	103.47
EGARCH(p=3,q=1,o=4)	constant	skewt	109.10
EGARCH(p=2,q=1,o=3)	zero	skewt	112.28
EGARCH(p=1,q=1,o=1)	zero	skewt	103.70
EGARCH(p=2,q=1,o=9)	constant	skewt	108.43
TARCH(p=0,q=3,o=2)	constant	skewt	92.60
TARCH(p=0,q=1,o=1)	constant	skewt	97.36
TARCH(p=0,q=1,o=1)	zero	skewt	98.01
TARCH(p=0,q=2,o=2)	constant	skewt	92.78
GJRGARCH(p=0,q=3,o=2)	constant	skewt	99.39
GJRGARCH(p=0,q=1,o=1)	zero	skewt	107.24
GJRGARCH(p=0,q=2,o=2)	constant	skewt	99.39
GJRGARCH(p=0,q=1,o=2)	constant	skewt	104.31
MSM(2)	\	\	148.01
MSM(3)	\	\	155.14
MSM(4)	\	\	148.69

In the appendix Comparison of in sample models selected on different time periods, we show the plots that compare in-sample and out-of-sample models.

4.3 Results depending on the data

4.3.1 OMX

Model	mean	dist	AIC	BIC	LL	MSE
GARCH(p=1,q=1)	constant	t	7006.28	7035.30	-3498.14	49.18
RiskMetrics	constant	normal	23809.8	23833.1	-11900.9	87.13
FIGARCH(p=0,q=1,tr=1000)	constant	skewt	6997.23	7032.05	-3492.61	49.72
FIGARCH(p=0,q=1,tr=1000)	zero	skewt	7001.86	7030.88	-3495.93	49.74
HARCH([3,7,22])	constant	skewt	7005.69	7046.32	-3495.85	49.58
HARCH([3,7,22])	zero	skewt	7010.14	7044.96	-3499.07	49.67
GARCH(p=1,q=1)	constant	skewt	6997.25	7032.08	-3492.62	49.20
GARCH(p=1,q=1)	constant	zero	7002.55	7031.57	-3496.27	49.32
EGARCH(p=2,q=1,o=9)	zero	skewt	6850.08	6937.14	-3410.04	46.80
EGARCH(p=2,q=1,o=2)	zero	skewt	6855.15	6901.58	-3419.58	46.07
TARCH(p=0,q=1,o=1)	zero	skewt	6873.13	6902.15	-3431.56	45.98
GJRGARCH(p=0,q=1,o=1)	zero	skewt	6884.32	6913.34	-3437.16	49.39
MSGARCH 1	\	\	6896.33	6960.18	-3437.16	35.26
MSGARCH 2	\	\	6872.02	6941.67	-3424.01	40.21
MSM(4)	\	\	\	\	-4103.36	48.33

MSGARCH 1 This model has 2 regimes, one being GARCH and other being GJRGARCH. Distribution of both models is student's t-distribution. Model is in state 1 0.2386% of the time and in the state 2 0.7614% of time. The transition matrix is

$$P = \begin{bmatrix} 0.0001 & 0.9999 \\ 0.3133 & 0.6867 \end{bmatrix}$$

MSGARCH 2 This model has 2 regimes, one being EGARCH with ged distribution and other being GARCH with skewed ged distribution. Model is in state 1 0.7077% of the time and in the state 2 0.2923% of time. The transition matrix is

$$P = \begin{bmatrix} 0.5970 & 0.4030 \\ 0.9757 & 0.0243 \end{bmatrix}$$

Models are compared in the figure 4.1

4.3.2 ETH-USD

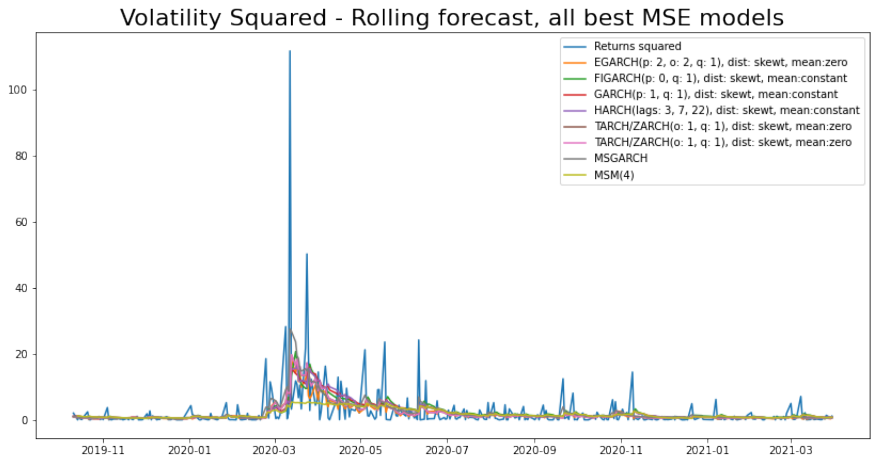

Figure 4.1: Rolling window forecast for OMX data

Model	mean	dist	AIC	BIC	LL	MSE
GARCH(p=1,q=1)	constant	t	10404.3	10431.5	-5197.15	3029.54
RiskMetrics	constant	normal	80714.1	80736.5	-40353.0	1512.08
FIGARCH(p=1,q=1,tr=500)	zero	ged	10387.2	10414.3	-5188.58	2959.16
FIGARCH(p=0,q=1,tr=500)	zero	ged	10387.3	10409.0	-5189.65	2957.83
HARCH([1,6,22])	zero	ged	10385.6	10412.8	-5187.81	2868.92
GARCH(p=1,q=1)	ged	zero	10385.0	10406.8	-5188.52	2802.73
EGARCH(p=1,q=1,o=2)	zero	ged	10382.8	10415.5	-5185.42	2762.36
EGARCH(p=1,q=1,o=1)	zero	ged	10383.0	10410.1	-5186.49	2763.77
TARCH(p=1,q=1,o=1)	zero	ged	10387.5	10414.7	-5188.76	2767.98
GJRGARCH(p=1,q=1,o=1)	zero	ged	10386.2	10413.3	-5188.08	2796.46
MSGARCH 1	\	\	10362.7	10422.5	-5170.4	1914.03
MSGARCH 2	\	\	10353.9	10419.1	-5164.95	1914.03
MSGARCH 3	\	\	10355.3	10469.4	-5156.65	1936.78
MSM(2)	\	\	\	\	-6286.18	1580.45

MSGARCH 1 This model has 2 regimes, one being GARCH and other being GJRGARCH. Distribution of both models is student's t-distribution. Model is in state 1 0.73% of the time and in the state 2 0.27% of time. The transition matrix is

$$P = \begin{bmatrix} 0.9171 & 0.0829 \\ 0.2240 & 0.7760 \end{bmatrix}$$

MSGARCH 2 This model has 2 regimes, one being EGARCH with ged distribution and other being GARCH with skewed ged distribution. Model is in the state 1 0.9407% of the time and in the state 2 0.0593% of time. The transition matrix is

$$P = \begin{bmatrix} 0.9565 & 0.0435 \\ 0.6892 & 0.3108 \end{bmatrix}$$

MSGARCH 3 This model has 3 regimes, one being GARCH with ged distribution, second being GJRGARCH with skewed ged distribution, and third being EGARCH with student's t-distribution. Model is in the state 1 0.5044% of the time, in the state 2 0.2754% of the time and in the state three 0.2202% of time. The transition matrix is

$$P = \begin{bmatrix} 0.7258 & 0.2742 & 0.0000 \\ 0.5021 & 0.3159 & 0.1820 \\ 0.0000 & 0.2276 & 0.7724 \end{bmatrix}$$

Models are compared in the figure 4.2

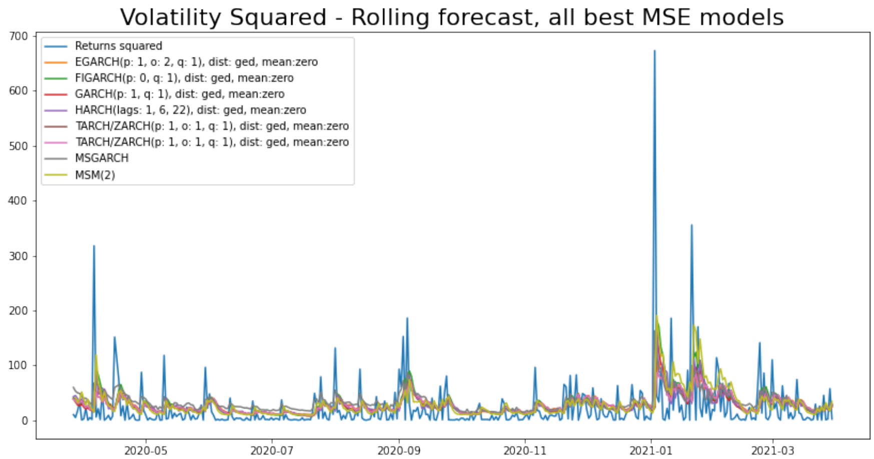

Figure 4.2: Rolling window forecast for ETH data

4.3.3 JNJ

Model	mean	dist	AIC	BIC	LL	MSE
FIGARCH(p=0,q=1,tr=5000)	constant	t	6167.48	6196.53	-3078.74	45.15
HARCH([2,6,24])	constant	t	6171.05	6205.90	-3079.52	45.83
GARCH(p=1,q=1)	constant	t	6168.68	6197.73	-3079.34	45.42
GARCH(p=2,q=2)	constant	t	6166.69	6207.35	-3076.34	44.97
EGARCH(p=1,q=1,o=7)	constant	t	6132.38	6202.09	-3054.19	48.05
EGARCH(p=1,q=1,o=1)	constant	t	6136.59	6171.44	-3062.29	48.53
TARCH(p=1,q=1,o=1)	constant	t	6136.00	6170.85	-3062.00	41.96
GJRGARCH(p=1,q=2,o=2)	constant	t	6141.02	6187.49	-3062.51	43.98
GJRGARCH(p=0,q=1,o=1)	constant	t	6143.99	6173.04	-3067.00	42.64
MSGARCH 1	\	\	6135.04	6198.94	-3056.52	38.63
MSGARCH 2	\	\	6131.79	6201.50	-3053.90	45.07
MSM(2)	\	\	\	\	-4117.03	68.30

MSGARCH 1

This model has 2 regimes, one being GARCH and other being GJRGARCH. Distribution of both models is student's t-distribution. Model is in state 1 0.3486% of the time and in the state 2 0.6514% of time. The transition matrix is

$$P = \begin{bmatrix} 0.9839 & 0.0161 \\ 0.0086 & 0.9914 \end{bmatrix}$$

MSGARCH 2

This model has 2 regimes, one being EGARCH with ged distribution and other being GARCH with skewed ged distribution. Model is in the state 1 0.7108% of the time and in the state 2 0.2892% of time. The transition matrix is

$$P = \begin{bmatrix} 0.989 & 0.011 \\ 0.027 & 0.973 \end{bmatrix}$$

Models are compared in the figure 4.3

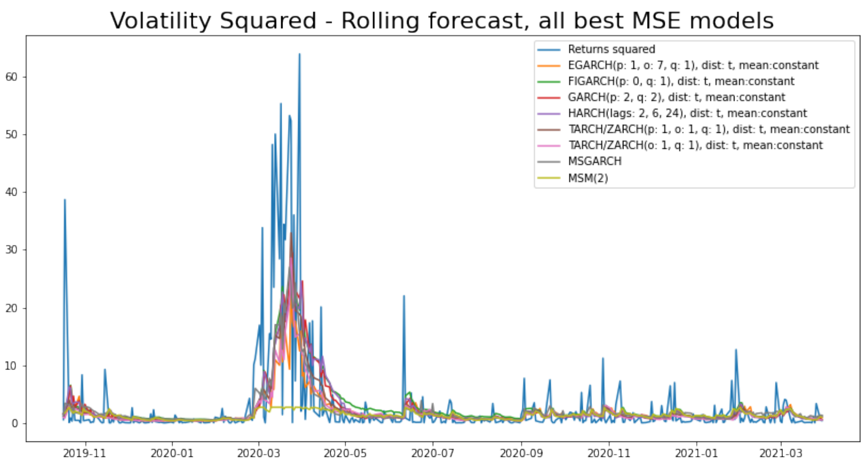

Figure 4.3: Rolling window forecast for JNJ data

4.4 Discussion, future improvement and remarks

We can see that preferable distribution for index data is skewed student's t-distribution while JNJ performs better with student's t-distribution and ETH with ged distribution. That means that distribution should be decided on the per asset basis.

Authors of the MSM model claimed that main advantage of MSM is its performance on longer horizons. MSM showed poor results in-sample and performed better than (G)ARCH-like models and MSGARCH out-of-sample only on ETH data. It is known that MSM usually underestimates volatility and therefore high volatility during the pandemic had a huge effect on the MSE score of MSM. It also apperead to be true for both MSM and MSGARCH that their performance was better for a smaller number of regimes. More concretely, models performed best overall with 2 regimes on average. Another fact that needs to be mentioned is that Multifractal models performed better than GARCH models on the data which did not include bursts of volatility.

As can be seen in the table below, MSM outperforms all (G)ARCH-like models out-of-sample for the period before the pandemic.

Model	MSE
EGARCH	3.37
FIGARCH	4.01
GARCH	3.97
HGARCH	3.99
TARCH	3.30
GJRGARCH	3.30
MSM(2)	2.97
MSM(2)+GJGARCH	2.46

We can also see that while MSM underestimates volatility for almost all low volatility periods, it does so also so for high volatility ones. This can be exploited so that when it is relatively low volatility state such as returns squared being smaller than $0.22 - 0.26$, we can forecast volatility in the following way:

1. When the volatility forecast by (G)ARCH-like model, in this case GJRGARCH is smaller than 0.26%, we take the prediction of MSM model.
2. When the volatility forecast by (G)ARCH-like model, in this case GJRGARCH is higher than 0.26%, we take the prediction of (G)ARCH-like model.

This shows the improvement of 21% over the best MSM model and improvement of 34% over the best (G)ARCH-like model out-of-sample. One can observe this in the 4.4.

We can also observe that in-sample MSM results do not significantly improve by increasing $\bar{k}$. This also manifested in out-of-sample results where MSM performed better for smaller k values where the best $\bar{k}$ depended on the asset.

Lastly, while MSM is parsimonious model, it requires significant computational effort since complexity grows exponentially with k.

We can also see that MSGARCH outperformed (G)ARCH-like and MSM models for the OMX data out-of-sample. While this was not a surprise, we can also find that RiskMetrics outperformed all the models out-of-sample for ETH data, with MSM on the second place.

Main improvement that would need to be made to the evaluation process is to evaluate these models using different metrics. While QL, briefly explained in

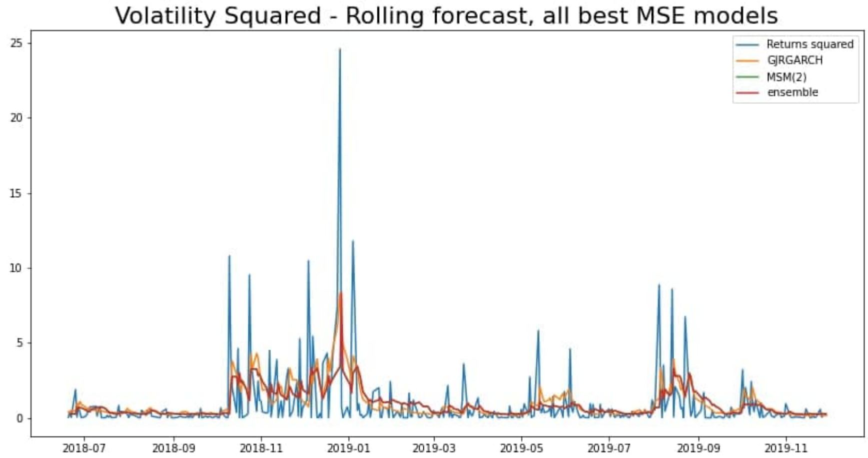

Figure 4.4: Ensemble - MSM - GJRGARCH comparison

this section, is one of the possible metrics, there are some remarkably better ones explained in [41] [40] and [42].

Another possible slight improvement and interesting thing to explore is to do a search of best possible models and distributions for MSGARCH. This could be done in simple way similarly to the GARCH hyperparameter search. What would be interesting to explore is the relationship between best (G)ARCH-like models and MSGARCH models using (G)ARCH-like models for the basis of regimes. Based on the data and results in this thesis, simple models as regimes work better than more complicated ones (see for example MSGARCH 1 versus MSGARCH 2 performance). This might be due to the fact that some of the features introduced to GARCH models are already modeled when switching regimes.

What seems like a possible idea is to model longer memory dependencies in GARCH using something like HARCH but assigning further lags bigger weights. This can be combined with TARCH idea of using different powers.

What can be seen on the plots as well as model evaluations is that all the models performed bad during pandemics. While this was expected and was impossible to predict well, it could probably be improved up to the some extent. One idea would be to use something simple like google trends data briefly described in the previous chapter as the additional parameter. In order to encapsulate the exogenous variables, one should search for words such as war, pandemic, government policy etc. This could also be included in MSGARCH model either indirectly by incorporating such parameters in individual regimes or directly by modifying the transition matrix with such variables.

Chapter 5

Momentum strategy

Momentum investing, also known as trend-following, is a strategy of buying stocks or securities that have had high returns over the past time period, usually three to twelve months, while selling those that have had low returns over the same time period. The momentum strategy is contrasted in a way with the efficient-market hypothesis which states that asset prices reflect all the available information. This hypothesis states that since assets always trade at their fair value, it is impossible to beat the market unless one purchases riskier investments.

Traditional asset models are not good at explaining success of momentum strategy [32] even though it has proven robust for over 200 years and across the globe. The problem with momentum strategy is that it brings risk. Namely, momentum investing usually experiences steeper drawdowns during crisis than other strategies where by drawdown is meant percentage between the peak and the subsequent jaz.

Jegadeesh and Titman [29] quantitatively show the idea of momentum in their famous paper where they find that stocks that performed well in the past can outperform stocks that have performed poorly in the past, at least in the near future. It is a popular view that humans tend to overreact to information which has a consequence of stock prices also overreacting to the information. The authors also show that it is profitable to buy past winners and sell past losers simultaneously.

There are many attempts in explaining momentum strategy and why it works. Many believe that momentum is related to over-confidence of investors, confirmation biases or in other words the tendency to interpret new evidence as the confirmation of one's existing beliefs or theories, as well as under-reaction, over-reaction and herding which is a tendency of investors to follow what other investors are doing rather than relying on their own analysis.

5.1 Cross-sectional momentum strategy

The cross-sectional momentum strategy portfolio is constructed based on asset's performance compared or relative to other assets. One of the popular strategies coming from the idea of momentum strategy is that for each decile, buying the top decile from the previous time period and selling the bottom decile leads to profit.

This type of momentum strategy is also known as winner-minus-loser, or shortly, WML strategy and it proves to produce at least 1% per month in the holding period of one year. Another momentum strategy called winner-only, or shortly WO, also proved good. It only buys the top decile, without shorting the bottom decile.

Even though, most of the time WML strategy performs well, Daniel and Moskowitz [16] as well as Barroso and Santa-Clara [10] found out that market crashes can lead to big losses when using this strategy. Size of the loss is such that it is possible that all returns accumulated over the years can easily be wiped out during such a crash. As was pointed out in those papers, momentum crashes are partly forecastable since they occur in panic states when market volatility is high. Barosso and Santa-Clara therefore propose momentum trading strategy but with volatility-scaling technique.

Another type of momentum strategy is time series momentum strategy. This type of momentum strategy considers only the absolute returns for the buy and sell trade. Unlike cross-sectional momentum strategies, it does not compare stocks to each other, rather it compares stocks to their own previous value.

In [32], Kim proposed a strategy taking into account both success of WML and WO and its downsides during large volatility periods. They proposed putting buy trade for top decile if the expected return of its holding period is positive and sell trade for bottom decile if the expected return of its holding period is negative. Otherwise they proposed no trades. Therefore, the whole strategy is based on predicting returns for top and bottom deciles.

5.2 Description of different momentum strategies explored in thesis

5.2.1 Gold momentum strategy

It has been observed that there is an inverse correlation between price of gold and health of economy. Namely, it was suggested that during high volatility periods, investors buy gold in order to hedge risk since gold is considered a stable asset which is not expected to easily drop in value by big margin.

The strategy that was evaluated is as follows. For 200 days, last being in the May this year, we predict volatility of the following day. If the squared volatility is above 0.9 and we do not already have gold stock in the possession, we buy a stock. In the opposite case, namely, when the volatility is below 0.9 and we do not own the stock - we buy one. This strategy gave return of 17% over 2000 days and performed worse than buy-and-hold strategy for gold which had return of 27%. This could possible be improved by changing very arbitrary 0.9 threshold for volatility as well as considering having multiple stocks and selling when volatility is above some smaller value than buy-threshold. What was also not accounted for is price and one could definitely improve the strategy by waiting for both values to be good enough.

5.2.2 Volatility-scaled momentum strategy

This strategy has been evaluated on the set of stocks containing Facebook, Apple, IBM, Microsoft and Google. It is evaluated on the time period from January 2010 to March 2021. Firstly, we compute monthly cumulative returns for each stock from the group. After that we look at the returns of past eleven months, avoiding the most recent month. The idea is to buy the best performing stock at the beginning of each month and to sell the worst performing stock at the same time. This is the previously mentioned WML strategy. However, before deciding on winner and loser for the respective month, we scale the returns by the volatility prediction for the future month in order to buy less volatile stocks for which momentum strategy is known to perform well. Although the strategy did perform well and had 136% excess return, it still performed worse than the buy-and-hold strategy for index S&P500. The strategy, however, outperformed S&P500 on 60 months and underperformed on 59 months. It also performed better than buy-and-hold strategy for IBM and momentum strategy for S&P500 where we buy top decile stocks and sell bottom decile stocks at the beginning of each month in the similar way as described previously but without volatility-scaling.

5.2.3 Turtle single stock strategy

This strategy considers single stock, in this case TSLA and time period of 500 days. Volatility forecast used in this model is expanding rolling-window forecast with the start of 500 days before the trading strategy backtesting starts. We look at 20 day moving average of the price and volume of trading. At the beginning of the strategy we buy a stock and later we have at most one stock of that same kind in any point of time. If the volume of trading is higher than the moving average and the volatility is higher than the specified threshold, we buy if the price is smaller than the moving average. Similarly, we sell when it is the opposite. This strategy performs better when transaction costs are not included than simple buy-and-hold strategy for the same stock.

5.3 Discussion, future improvement and remarks

Volatlity index or **VIX** is derived from S&P500 options prices and it basically represents US stock market expected volatility and market sentiment. VIX demonstrates volatility clustering as might be expected since it is based on the volatility. VIX is a security one can trade and it is an annualized number that represents what the market's expectation of 30-day forward-looking volatility is. Opposite to other market products, VIX cannot be bought or sold directly, but is instead traded and exchanged via derivative contract, derived ETFs and ETNs which most commonly track VIX futures indices.

Hedging strategy is a method of reducing exposure to risk in the event that an asset in the portfolio has bigger chances of decreasing. Ideally, they would limit losses but not significantly reduce rate of return. One of the main ideas of hedging is to buy securities inversely correlated with the risky asset.

It has been observed that there exists a negative correlation between S&P500 and VIX. Usually, when VIX goes up, S&P500 goes down which can be seen on the figure 5.1.

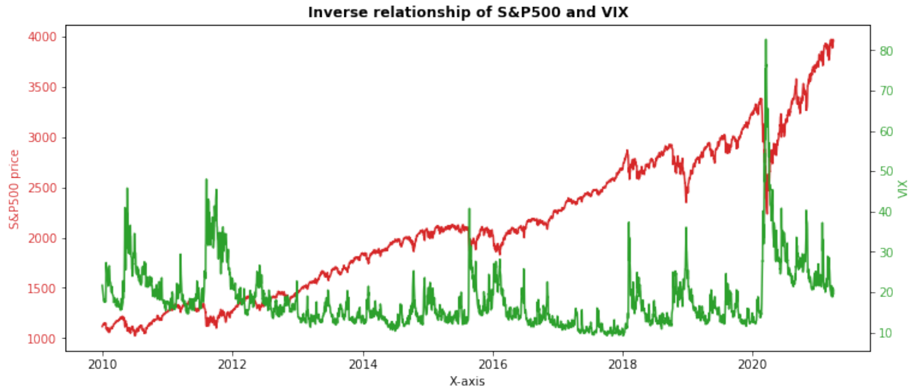

Figure 5.1: Negative correlation between S&P500 and VIX.

Volatility can be directly traded via VIX. If we expect an increase in volatility, we can go to long futures position while if we expect decrease in volatility, we can go into short futures position. This can be exploited to create the potential improvement in momentum strategy.

Options are financial derivatives that give buyers the right to buy or sell an underlying asset at the agreed price and date. Options are special in that buyer is not obliged to buy or sell, but only presented with a possibility. Call options give buyer the right to buy whereas put options give the buyer right to sell the specific bond, stock, commodity or other asset or instrument. Put options are type of protection. Best time to buy them would be when price is low but volatility could grow in future. Again, common way of protecting portfolio is by buying put options. The higher the volatility, the higher is option priced. This can also be exploited in order to enhance the momentum strategy.

However, this was not explored further in the thesis due to the difficulty in obtaining the appropriate data.

Chapter 6

Conclusion

During the last six months I have participated in many workshops, webinars and talks on the topic of automatic trading, algorithmic trading and finance in general. I attempted in many ways to understand the gap between what is published and what is actually going on in the financial world. What sounded strange 6 months ago when I talked with professor Sornette now makes sense. The best ideas are not published and while in all of the companies quant researchers have weekly meetings in order to read newly published papers from academia, academia does not get to know most of the work done in best companies.

When I asked professor Sornette to do a thesis with him, I must say that I have hoped that there will be some natural language processing part, the area I am also very passionate about. I was very quickly discouraged from doing that by professor Sornette with the explanation that in academia we do not have sufficient computational resources and that without that one cannot that easily find new results in the intersection of finance and natural language processing. Today I am very grateful to professor Sornette for that advice since that proved to be true. However, while in the periods of low volatility models explored in this thesis do perform well, out-of-sample results presented in this thesis are a showcase of their bad performance when the crisis occurs. Markets are in great deal indeed driven by news, sentiment on twitter which started to expand to the other platforms such as reddit with the remarkable example of Game Stop volatility at the beginning of this year[2].

In chapter 2 of this thesis, we discuss stylized facts about volatility with the case example of S&P500 data. In chapter 3, models which are used in the thesis are explained with the idea of incorporating mentioned stylized facts. The idea of incorporating exogenous variables such as Google trends data is proposed at the end of the chapter 3. In chapter 4, we evaluate those models on the four different assets, in- and out-of-sample. We do so by including pandemic period to the out-of-sample data. Best selected models varied by assets, as well as best distribution of residuals. We also discuss possible improvements, of which the simplest but well-performing one is using the multiple volatility predictors depending on the state of volatility.

Finally, in chapter 5 we briefly discuss momentum strategies and their possible extensions and improvements with the additional volatility information.

Chapter 7

Software

In this section we briefly mention software packages used in this thesis.

The **MSGARCH** package [5] is implemented in R while using C++ in the background for the reasons of efficiency. MSGARCH makes it possible to create simulations and perform maximum likelihood estimation. It supports single-step as well as multi-step ahead forecasts. Lastly, advantages that risk managers can enjoy consist of broad functionality such as estimation of conditional volatility, value-at-risk and expected-shortfall [6].

The Python **ARCH** toolbox [1] contains routines for univariate volatility models, bootstrapping, multiple comparison procedures as well as the required testing infrastructure.

Statsmodels [48] is a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration. The module encapsulates a for each estimator. The results are tested against existing statistical packages to ensure that they are correct.

Bibliography

[1] Arch package. https://arch.readthedocs.io. Accessed: 2021-05-22.

[2] Yigit Alparslan and Edward Kim. Extreme volatility prediction in stock market: When gamestop meets long short-term memory networks. *CoRR*, abs/2103.01121, 2021.

[3] Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. Modeling and forecasting realized volatility. Working Paper 8160, National Bureau of Economic Research, March 2001.

[4] David Ardia, Keven Bluteau, Kris Boudt, and Leopoldo Catania. Forecasting risk with markov-switching garch models:a large-scale performance study. *International Journal of Forecasting*, 34(4):733–747, 2018.

[5] David Ardia, Keven Bluteau, Kris Boudt, Leopoldo Catania, and Denis-Alexandre Trottier. Markov-switching GARCH models in R: The MSGARCH package. *Journal of Statistical Software*, 91(4):1–38, 2019.

[6] David Ardia, Keven Bluteau, Kris Boudt, Leopoldo Catania, and Denis-Alexandre Trottier. Markov-switching garch models in r: The msgarch package. *Journal of Statistical Software, Articles*, 91(4):1–38, 2019.

[7] Francesco Audrino and Yujia Hu. Volatility forecasting: Downside risk, jumps and leverage effect. *Econometrics*, 4(1), 2016.

[8] Francesco Audrino and Yujia Hu. Volatility forecasting: Downside risk, jumps and leverage effect. *Econometrics*, 4(1), 2016.

[9] Richard T. Baillie, Tim Bollerslev, and Hans Ole Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30, 1996.

[10] Pedro Barroso and Pedro Santa-Clara. Momentum has its moments. *Journal of Financial Economics*, 116(1):111–120, 2015.

[11] Geert Bekaert and Guojun Wu. Asymmetric Volatility and Risk in Equity Markets. *The Review of Financial Studies*, 13(1):1–42, 06 2015.

[12] Tim Bollerslev. Glossary to arch (garch). Creates research papers, Department of Economics and Business Economics, Aarhus University, 2008.

[13] Jean-Philippe Bouchaud, Andrew Matacz, and Marc Potters. Leverage effect in financial markets: The retarded volatility model. *Physical Review Letters*, 87(22), Nov 2001.

[14] Christian Brownlees, Robert Engle, and Bryan Kelly. A practical guide to volatility forecasting through calm and storm. *Journal of Risk*, 14, 08 2011.

[15] Laurent E. Calvet. Multifractal volatility : theory, forecasting, and pricing, 2008.

[16] Kent Daniel and Tobias J. Moskowitz. Momentum crashes. *Journal of Financial Economics*, 122(2):221–247, 2016.

[17] Carmen Fernandez and Mark F. J. Steel. On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371, 1998.

[18] Stephen Figlewski. Forecasting volatility using historical data. 05 1994.

[19] Stephen Figlewski and Xiaozu Wang. Is the 'leverage effect' a leverage effect? *SSRN Electronic Journal*, 11 2000.

[20] Lawrence R. Glosten, Ravi Jagannathan, and David E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. Staff Report 157, Federal Reserve Bank of Minneapolis, 1993.

[21] Clive Granger and Ser-Huang Poon. Practical issues in forecasting volatility. *Financial Analysts Journal*, 61, 02 2005.

[22] Jeremy Greenwood and Boyan Jovanovic. The information-technology revolution and the stock market. *American Economic Review*, 89(2):116–122, May 1999.

[23] Maxim Gusev, Dimitri Kroujiline, Boris Govorkov, Sergey V. Sharov, Dmitry Ushanov, and Maxim Zhilyaev. Predictable markets? a news-driven model of the stock market, 2014.

[24] Markus Haas. A new approach to markov-switching garch models. *Journal of Financial Econometrics*, 2(4):493–530, 2004.

[25] Markus Haas, Stefan Mittnik, and Marc Paolella. A new approach to markov-switching garch models. *Journal of Financial Econometrics*, 2, 11 2003.

[26] Alain Hamid and Moritz Heiden. Forecasting volatility with empirical similarity and google trends. *Journal of Economic Behavior Organization*, 117:62–81, 2015.

[27] James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1 edition, January 1994.

[28] Peter Hansen and Asger Lunde. Consistent ranking of volatility models. *Journal of Econometrics*, 131(1-2):97–121, 2006.

[29] Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91, 1993.

[30] Zhi-Qiang Jiang, Wen-Jie Xie, Wei-Xing Zhou, and Didier Sornette. Multifractal analysis of financial markets: a review. *Reports on Progress in Physics*, 82(12):125901, Nov 2019.

[31] Ujwal Kandi, Sasikanth Gujjula, Venkatesh Buddha, and V S Bhagavan. Visualizing the financial impact of presidential tweets on stock markets, 2021.

[32] Saejoon Kim. Enhancing the momentum strategy through deep regression. *Quantitative Finance*, 19(7):1121–1133, 2019.

[33] Y. Lempérière, C. Deremble, P. Seager, M. Potters, and J. P. Bouchaud. Two centuries of trend following, 2014.

[34] Feng Ma, Yu Li, Li Liu, and Yaojie Zhang. Are low-frequency data really uninformative? a forecasting combination perspective. *The North American Journal of Economics and Finance*, 44:92–108, 2018.

[35] B. B. Mandelbrot. *The Fractal Geometry of Nature*. 1982.

[36] Benoit B. Mandelbrot. Multifractal power law distributions: Negative and critical dimensions and other “anomalies,” explained by a simple example. *Journal of Statistical Physics*, 2003.

[37] Philippe Masset. Volatility stylized facts. *SSRN Electronic Journal*, 09 2011.

[38] Christopher J. Neely and Paul A. Weller. Predicting exchange rate volatility: genetic programming vs. GARCH and RiskMetrics. Working Papers 2001-009, Federal Reserve Bank of St. Louis, 2001.

[39] Daniel B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991.

[40] Andrew Patton and Kevin Sheppard. Evaluating volatility and correlation forecasts. *Handbook of Financial Time Series*, 02 2008.

[41] Andrew J. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256, 2011. Realized Volatility.

[42] Andrew J. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256, 2011. Realized Volatility.

[43] Thanasarn Porthaveepong. Multifractal model - Parsimonious model for volatility in financial market. 2017.

[44] Davide Raggi and Silvano Bordignon. Long memory and nonlinearities in realized volatility: A markov switching approach. *Computational Statistics Data Analysis*, 56(11):3730–3742, 2012. 1st issue of the Annals of Computational and Financial Econometrics Sixth Special Issue on Computational Econometrics.

[45] Ashish Rajbhandari. Estimating Markov-switching regression models in Stata. 2015 Stata Conference 24, Stata Users Group, August 2015.

[46] Richard Roll. Orange juice and weather. *American Economic Review*, 74(5):861–80, 1984.

[47] Adrienne Ross, Tobias Moskowitz, Ronen Israel, and Laura Serban. Implementing momentum: What have we learned? *SSRN Electronic Journal*, 01 2017.

[48] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[49] Didier Sornette, Peter Cauwels, and Georgi Smilyanov. Can we use volatility to diagnose financial bubbles? lessons from 40 historical bubbles. *Quantitative Finance and Economics*, 2:1–105, 03 2018.

[50] Didier Sornette, Yannick Malevergne, and J.F. Muzy. What causes crashes? *Risk*, 16:67–71, 02 2003.

[51] Alexander Wehrli, Spencer Wheatley, and Didier Sornette. Scale-, time- and asset-dependence of hawkes process estimates on high frequency price changes. *Quantitative Finance*, 21(5):729–752, 2021.

[52] Yu Wei and Peng Wang. Forecasting volatility of ssec in chinese stock market using multifractal analysis. *Physica A: Statistical Mechanics and its Applications*, 387(7):1585–1592, 2008.

Appendix

In sample results 1995-2010, S&P500

In sample results of S&P500 volatility forecasts between 01.01.1995. and 31.12.2009. For each model best parameters are selected according to BIC and AIC criteria.

GARCH(1,1) - Benchmark

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0746	1.313e-02	5.682	1.328e-08	[4.888e-02, 0.100]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	6.0748e-03	2.237e-03	2.716	6.616e-03	[1.690e-03,1.046e-02]
alpha[1]	0.0684	9.625e-03	7.101	1.236e-12	[4.949e-02,8.722e-02]
beta[1]	0.9298	9.413e-03	98.783	0.000	[0.911, 0.948]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	8.0259	1.117	7.183	6.827e-13	[5.836, 10.216]

GARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0645	1.363e-02	4.732	2.229e-06	[3.779e-02,9.123e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	8.4598e-03	3.168e-03	2.670	7.584e-03	[2.250e-03,1.467e-02]
alpha[1]	0.0200	1.577e-02	1.267	0.205	[-1.093e-02,5.088e-02]
alpha[2]	0.0668	2.121e-02	3.150	1.630e-03	[2.525e-02, 0.108]
beta[1]	0.9102	1.423e-02	63.968	0.000	[0.882, 0.938]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	8.3271	1.164	7.153	8.491e-13	[6.045, 10.609]
lambda	-0.0681	2.000e-02	-3.402	6.688e-04	[-0.107,-2.885e-02]

Model selected by BIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0645	1.363e-02	4.732	2.229e-06	[3.779e-02,9.123e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	8.4598e-03	3.168e-03	2.670	7.584e-03	[2.250e-03,1.467e-02]
alpha[1]	0.0200	1.577e-02	1.267	0.205	[-1.093e-02,5.088e-02]
alpha[2]	0.0668	2.121e-02	3.150	1.630e-03	[2.525e-02, 0.108]
beta[1]	0.9102	1.423e-02	63.968	0.000	[0.882, 0.938]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	8.3271	1.164	7.153	8.491e-13	[6.045, 10.609]
lambda	-0.0681	2.000e-02	-3.402	6.688e-04	[-0.107,-2.885e-02]

EGARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0343	1.325e-02	2.590	9.598e-03	[8.349e-03,6.030e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	2.2971e-03	2.320e-03	0.990	0.322	[-2.250e-03,6.844e-03]
alpha[1]	-0.1386	4.283e-02	-3.235	1.217e-03	[-0.223,-5.461e-02]
alpha[2]	0.1647	6.178e-02	2.665	7.692e-03	[4.358e-02, 0.286]
alpha[3]	0.0941	4.235e-02	2.223	2.622e-02	[1.114e-02, 0.177]
gamma[1]	-0.2112	2.733e-02	-7.727	1.103e-14	[-0.265, -0.158]
gamma[2]	-0.0426	3.686e-02	-1.155	0.248	[-0.115,2.968e-02]
gamma[3]	0.1188	3.565e-02	3.332	8.619e-04	[4.891e-02, 0.189]
gamma[4]	0.0387	2.480e-02	1.561	0.119	[-9.892e-03,8.730e-02]
beta[1]	0.9870	2.699e-03	365.644	0.000	[0.982, 0.992]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	11.0480	1.892	5.840	5.219e-09	[7.340, 14.756]
lambda	-0.0872	2.145e-02	-4.066	4.781e-05	[-0.129,-4.517e-02]

Model selected by BIC criteria has standardized skew student's t-distribution and zero mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	6.5427e-03	1.786e-03	3.663	2.496e-04	[3.042e-03,1.004e-02]
alpha[1]	-0.1398	4.739e-02	-2.950	3.173e-03	[-0.233,-4.694e-02]
alpha[2]	0.2590	4.822e-02	5.372	7.804e-08	[0.164, 0.353]
gamma[1]	-0.2060	2.791e-02	-7.382	1.555e-13	[-0.261, -0.151]
gamma[2]	-0.0512	3.720e-02	-1.376	0.169	[-0.124,2.173e-02]
gamma[3]	0.1570	2.788e-02	5.632	1.776e-08	[0.102, 0.212]
beta[1]	0.9848	2.702e-03	364.507	0.000	[0.979, 0.990]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	10.9746	1.897	5.786	7.188e-09	[7.257, 14.692]
lambda	-0.0963	2.049e-02	-4.699	2.609e-06	[-0.136,-5.614e-02]

TARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0330	1.357e-02	2.434	1.493e-02	[6.434e-03,5.963e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0314	6.757e-03	4.650	3.324e-06	[1.817e-02,4.466e-02]
gamma[1]	0.1075	2.319e-02	4.635	3.568e-06	[6.203e-02, 0.153]
gamma[2]	0.1437	3.369e-02	4.263	2.013e-05	[7.761e-02, 0.210]
beta[1]	0.3749	0.150	2.499	1.244e-02	[8.091e-02, 0.669]
beta[2]	0.0614	0.132	0.465	0.642	[-0.197, 0.320]
beta[3]	0.4377	0.152	2.886	3.907e-03	[0.140, 0.735]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	10.6551	1.839	5.795	6.839e-09	[7.051, 14.259]
lambda	-0.0903	2.064e-02	-4.373	1.224e-05	[-0.131,-4.982e-02]

Model selected by BIC criteria has standardized skew student's t-distribution and zero mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0187	3.128e-03	5.987	2.136e-09	[1.260e-02,2.486e-02]
gamma[1]	0.1322	1.327e-02	9.966	2.144e-23	[0.106, 0.158]
beta[1]	0.9339	7.116e-03	131.237	0.000	[0.920, 0.948]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	10.6430	1.861	5.720	1.064e-08	[6.996, 14.290]
lambda	-0.1037	1.982e-02	-5.231	1.689e-07	[-0.143,-6.483e-02]

FIGARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0634	1.382e-02	4.585	4.545e-06	[3.628e-02,9.047e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0232	8.131e-03	2.850	4.366e-03	[7.240e-03,3.911e-02]
phi	0.1151	4.363e-02	2.637	8.355e-03	[2.956e-02, 0.201]
d	0.6070	9.305e-02	6.524	6.863e-11	[0.425, 0.789]
beta	0.7020	7.053e-02	9.953	2.449e-23	[0.564, 0.840]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	8.2096	1.081	7.596	3.063e-14	[6.091, 10.328]
lambda	-0.0679	2.026e-02	-3.350	8.090e-04	[-0.108,-2.816e-02]

Model selected by BIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0632	1.370e-02	4.610	4.029e-06	[3.631e-02,9.003e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0135	6.270e-03	2.155	3.117e-02	[1.223e-03,2.580e-02]
d	0.8503	6.195e-02	13.725	7.151e-43	[0.729, 0.972]
beta	0.8359	5.239e-02	15.956	2.603e-57	[0.733, 0.939]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	8.0515	1.063	7.573	3.648e-14	[5.968, 10.135]
lambda	-0.0697	2.020e-02	-3.451	5.582e-04	[-0.109,-3.013e-02]

GJRGARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0358	1.360e-02	2.635	8.421e-03	[9.177e-03,6.249e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0213	5.682e-03	3.742	1.824e-04	[1.013e-02,3.240e-02]
gamma[1]	0.0919	2.733e-02	3.361	7.768e-04	[3.829e-02, 0.145]
gamma[2]	0.1869	3.954e-02	4.726	2.285e-06	[0.109, 0.264]
beta[1]	0.2958	0.114	2.597	9.406e-03	[7.255e-02, 0.519]
beta[2]	0.0889	0.143	0.620	0.535	[-0.192, 0.370]
beta[3]	0.4608	0.152	3.034	2.416e-03	[0.163, 0.759]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	10.9177	2.000	5.458	4.829e-08	[6.997, 14.839]
lambda	-0.0951	2.049e-02	-4.644	3.414e-06	[-0.135,-5.499e-02]

Model selected by BIC criteria has standardized skew student's t-distribution and zero mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0118	2.618e-03	4.494	7.002e-06	[6.633e-03,1.689e-02]
gamma[1]	0.1387	1.776e-02	7.812	5.609e-15	[0.104, 0.174]
beta[1]	0.9260	8.888e-03	104.179	0.000	[0.909, 0.943]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	10.7175	1.967	5.450	5.050e-08	[6.863, 14.572]
lambda	-0.1087	1.974e-02	-5.504	3.715e-08	[-0.147,-6.996e-02]

HARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0654	1.366e-02	4.791	1.658e-06	[3.867e-02,9.220e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.1094	2.601e-02	4.206	2.603e-05	[5.842e-02, 0.160]
alpha[1]	0.0000	2.848e-02	0.000	1.000	[-5.582e-02,5.582e-02]
alpha[7]	0.2893	7.456e-02	3.880	1.046e-04	[0.143, 0.435]
alpha[23]	0.6780	8.068e-02	8.403	4.340e-17	[0.520, 0.836]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	8.3643	1.164	7.187	6.627e-13	[6.083, 10.645]
lambda	-0.0686	1.963e-02	-3.493	4.772e-04	[-0.107,-3.010e-02]

Model selected by BIC criteria has generalized error distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0765	1.328e-02	5.763	8.281e-09	[5.050e-02, 0.103]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.1153	2.785e-02	4.140	3.467e-05	[6.072e-02, 0.170]
alpha[2]	0.0000	4.251e-02	0.000	1.000	[-8.331e-02,8.331e-02]
alpha[7]	0.2841	8.697e-02	3.266	1.089e-03	[0.114, 0.455]
alpha[23]	0.6796	8.217e-02	8.270	1.341e-16	[0.519, 0.841]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	1.4199	5.527e-02	25.690	1.512e-145	[1.312, 1.528]

RiskMetrics

	coef
mu	0.0000

	coef
omega	0.0000
alpha[1]	0.9400
beta[1]	0.0600

In sample results 2010-2021, S&P500

GARCH(1,1) - Benchmark

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0935	1.166e-02	8.023	1.032e-15	[7.070e-02, 0.116]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0266	5.976e-03	4.447	8.691e-06	[1.487e-02,3.829e-02]
alpha[1]	0.1855	2.317e-02	8.007	1.175e-15	[0.140, 0.231]
beta[1]	0.8067	2.011e-02	40.104	0.000	[0.767, 0.846]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	4.9941	0.465	10.745	6.249e-27	[4.083, 5.905]

GARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0742	1.289e-02	5.755	8.641e-09	[4.891e-02,9.942e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0337	8.198e-03	4.105	4.046e-05	[1.758e-02,4.972e-02]
alpha[1]	0.1160	2.863e-02	4.051	5.095e-05	[5.987e-02, 0.172]
alpha[2]	0.1015	4.061e-02	2.500	1.242e-02	[2.193e-02, 0.181]
beta[1]	0.7656	3.099e-02	24.701	1.039e-134	[0.705, 0.826]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	5.4697	0.556	9.840	7.585e-23	[4.380, 6.559]
lambda	-0.1063	2.412e-02	-4.409	1.037e-05	[-0.154,-5.907e-02]

Model selected by BIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0747	1.287e-02	5.806	6.385e-09	[4.950e-02,9.995e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0254	5.665e-03	4.486	7.264e-06	[1.431e-02,3.651e-02]
alpha[1]	0.1783	2.149e-02	8.295	1.083e-16	[0.136, 0.220]
beta[1]	0.8095	1.964e-02	41.225	0.000	[0.771, 0.848]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	5.4597	0.553	9.868	5.734e-23	[4.375, 6.544]
lambda	-0.1037	2.401e-02	-4.321	1.555e-05	[-0.151,-5.668e-02]

EGARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0279	1.291e-02	2.161	3.068e-02	[2.598e-03,5.319e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	1.1853e-03	4.441e-03	0.267	0.790	[-7.519e-03,9.890e-03]
alpha[1]	0.0771	4.270e-02	1.805	7.111e-02	[-6.628e-03, 0.161]
alpha[2]	0.0693	4.167e-02	1.663	9.636e-02	[-1.238e-02, 0.151]
gamma[1]	-0.2556	3.037e-02	-8.414	3.946e-17	[-0.315, -0.196]
gamma[2]	-0.0206	3.788e-02	-0.543	0.587	[-9.480e-02,5.367e-02]
gamma[3]	0.0744	3.838e-02	1.938	5.264e-02	[-8.482e-04, 0.150]
gamma[4]	-0.0368	3.797e-02	-0.969	0.333	[-0.111,3.762e-02]
gamma[5]	0.0782	4.157e-02	1.882	5.988e-02	[-3.253e-03, 0.160]
gamma[6]	-0.0358	4.495e-02	-0.797	0.425	[-0.124,5.227e-02]
gamma[7]	0.0566	4.122e-02	1.372	0.170	[-2.422e-02, 0.137]
gamma[8]	-0.0370	4.331e-02	-0.854	0.393	[-0.122,4.791e-02]
gamma[9]	0.0702	3.587e-02	1.957	5.031e-02	[-9.594e-05, 0.141]
beta[1]	0.9811	6.308e-03	155.532	0.000	[0.969, 0.993]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	6.1880	0.765	8.092	5.858e-16	[4.689, 7.687]
lambda	-0.1713	2.691e-02	-6.365	1.954e-10	[-0.224, -0.119]

Model selected by BIC criteria has standardized skew student's t-distribution and zero mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	7.3972e-03	5.595e-03	1.322	0.186	[-3.569e-03,1.836e-02]
alpha[1]	0.2014	2.338e-02	8.611	7.227e-18	[0.156, 0.247]
gamma[1]	-0.2184	1.813e-02	-12.047	2.021e-33	[-0.254, -0.183]
beta[1]	0.9551	7.331e-03	130.277	0.000	[0.941, 0.969]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	5.7886	0.662	8.740	2.322e-18	[4.491, 7.087]
lambda	-0.1778	2.341e-02	-7.596	3.053e-14	[-0.224, -0.132]

TARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0259	1.246e-02	2.074	3.805e-02	[1.425e-03,5.028e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0770	1.156e-02	6.663	2.680e-11	[5.436e-02,9.967e-02]
gamma[1]	0.2332	2.574e-02	9.061	1.287e-19	[0.183, 0.284]
gamma[2]	0.2781	2.783e-02	9.995	1.596e-23	[0.224, 0.333]
beta[1]	0.0231	4.332e-02	0.534	0.593	[-6.178e-02, 0.108]
beta[2]	0.7111	3.971e-02	17.908	1.030e-71	[0.633, 0.789]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	6.1081	0.727	8.398	4.532e-17	[4.683, 7.534]
lambda	-0.1709	2.648e-02	-6.454	1.089e-10	[-0.223, -0.119]

Model selected by BIC criteria has standardized skew student's t-distribution and zero mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0435	6.052e-03	7.189	6.518e-13	[3.165e-02,5.537e-02]
gamma[1]	0.2792	2.338e-02	11.944	6.984e-33	[0.233, 0.325]
beta[1]	0.8587	1.247e-02	68.877	0.000	[0.834, 0.883]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	6.0352	0.721	8.372	5.647e-17	[4.622, 7.448]
lambda	-0.1860	2.411e-02	-7.714	1.224e-14	[-0.233, -0.139]

FIGARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0729	1.272e-02	5.725	1.033e-08	[4.791e-02,9.779e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0364	9.977e-03	3.650	2.621e-04	[1.686e-02,5.597e-02]
d	0.6001	7.737e-02	7.756	8.733e-15	[0.448, 0.752]
beta	0.4665	7.890e-02	5.913	3.360e-09	[0.312, 0.621]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	5.5246	0.544	10.160	3.001e-24	[4.459, 6.590]
lambda	-0.1114	2.430e-02	-4.585	4.533e-06	[-0.159,-6.380e-02]

Model selected by BIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0729	1.272e-02	5.725	1.033e-08	[4.791e-02,9.779e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0364	9.977e-03	3.650	2.621e-04	[1.686e-02,5.597e-02]
d	0.6001	7.737e-02	7.756	8.733e-15	[0.448, 0.752]
beta	0.4665	7.890e-02	5.913	3.360e-09	[0.312, 0.621]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	5.5246	0.544	10.160	3.001e-24	[4.459, 6.590]
lambda	-0.1114	2.430e-02	-4.585	4.533e-06	[-0.159,-6.380e-02]

GJRGARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0380	1.281e-02	2.966	3.013e-03	[1.289e-02,6.310e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0617	9.860e-03	6.256	3.949e-10	[4.236e-02,8.101e-02]
gamma[1]	0.2618	4.521e-02	5.790	7.033e-09	[0.173, 0.350]
gamma[2]	0.3639	5.028e-02	7.239	4.523e-13	[0.265, 0.462]
beta[1]	4.0761e-04	3.726e-02	1.094e-02	0.991	[-7.263e-02,7.344e-02]
beta[2]	0.6527	3.734e-02	17.478	2.104e-68	[0.579, 0.726]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	5.7430	0.636	9.024	1.815e-19	[4.496, 6.990]
lambda	-0.1515	2.583e-02	-5.864	4.508e-09	[-0.202, -0.101]

Model selected by BIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0385	1.275e-02	3.020	2.524e-03	[1.352e-02,6.351e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.0317	5.315e-03	5.971	2.361e-09	[2.132e-02,4.215e-02]
gamma[1]	0.3315	4.385e-02	7.559	4.071e-14	[0.246, 0.417]
beta[1]	0.8167	1.825e-02	44.742	0.000	[0.781, 0.853]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	5.7282	0.637	8.996	2.332e-19	[4.480, 6.976]
lambda	-0.1488	2.554e-02	-5.827	5.641e-09	[-0.199,-9.878e-02]

HARCH

Model selected by AIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0731	1.301e-02	5.621	1.893e-08	[4.762e-02,9.861e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.1332	2.297e-02	5.799	6.684e-09	[8.817e-02, 0.178]
alpha[3]	0.2461	8.058e-02	3.054	2.257e-03	[8.817e-02, 0.404]
alpha[6]	0.4020	0.102	3.946	7.932e-05	[0.202, 0.602]
alpha[23]	0.2886	6.064e-02	4.760	1.940e-06	[0.170, 0.407]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	5.3607	0.533	10.049	9.315e-24	[4.315, 6.406]
lambda	-0.1077	2.431e-02	-4.431	9.369e-06	[-0.155,-6.008e-02]

Model selected by BIC criteria has standardized skew student's t-distribution and constant mean.

	coef	std err	t	P> |t|	95.0% Conf. Int.
mu	0.0731	1.301e-02	5.621	1.893e-08	[4.762e-02,9.861e-02]

	coef	std err	t	P> |t|	95.0% Conf. Int.
omega	0.1332	2.297e-02	5.799	6.684e-09	[8.817e-02, 0.178]
alpha[3]	0.2461	8.058e-02	3.054	2.257e-03	[8.817e-02, 0.404]
alpha[6]	0.4020	0.102	3.946	7.932e-05	[0.202, 0.602]
alpha[23]	0.2886	6.064e-02	4.760	1.940e-06	[0.170, 0.407]

	coef	std err	t	P> |t|	95.0% Conf. Int.
nu	5.3607	0.533	10.049	9.315e-24	[4.315, 6.406]
lambda	-0.1077	2.431e-02	-4.431	9.369e-06	[-0.155,-6.008e-02]

RiskMetrics

	coef
mu	0.0000

	coef
omega	0.0000
alpha[1]	0.9400
beta[1]	0.0600

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

MULTIFRACTAL VOLATILITY PREDICTORS FOR ENHANCED PORTFOLIO STRATEGIES

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
PEPIC	SELENA

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
18.05.2021.	[signature]

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.