MASTERARBEIT / MASTER’S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

„Applying GARCH-MIDAS Models To A Volatility Trading Strategy“

verfasst von / submitted by

Michael Dampf, BSc (WU)

gemeinsam mit / in collaboration with

Andrii Mudrak, Bak.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of-

Master of Science (MSc)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 974
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Banking and Finance UG2002
Betreut von / Supervisor:	Dr. Jérémy Leymarie, licence master

APPLYING GARCH-MIDAS MODELS TO A VOLATILITY TRADING STRATEGY

Abstract

The goal of this paper is to identify whether GARCH-MIDAS models are better suited to estimate the S&P 500 volatility risk premium (VRP) for a VIX trading strategy than conventional- and VIX futures term-structure- implied VRP models. To determine the best-suited model, the model-estimated VRP will act as an indicator in a long/short VIX trading strategy and the models will be compared based on the out of sample Sharpe ratio. Apart from the binary indicator, we also employ threshold and scaled indicators to take the magnitude of the volatility risk premium into account. Moreover, we will extend GARCH-MIDAS models to Double Asymmetric GARCH-MIDAS models (DAGM) to adjust for the asymmetry of volatility. Our results show that DAGM models that use the term-structure of SPX options as a long-term component with a threshold indicator yield a higher Sharpe ratio than conventional- and VIX futures term-structure implied VRP models. Moreover, we demonstrate that the performance of VIX trading strategies is heavily impacted by the market regime in which they are evaluated.

Contents

1 | Introduction

With the global financial system going through its second major crisis in the 21st century, the resulting low interest rate environment has sent the world of finance on a chase for returns from alternative investments. An asset class that has seen a rise in popularity in recent years is volatility, especially the volatility of the S&P 500. The rise of ETNs which allow investors to trade volatility without gaining exposure to options has not only triggered a large influx of retail investors into volatility trading, but also caught the interest of academia, resulting in more articles being published on this topic. While the biggest emphasis was put on volatility derivatives analysis and volatility forecasting model development, some of the articles focused on our research topic – volatility trading.

The literature on volatility trading models predominantly focuses on trading volatility through estimating the volatility risk premium (VRP). Ex-post, it is the difference between implied volatility (IVOL) and realized volatility (RVOL) and is indicative of a premium that can be earned through entering into a volatility position. Depending on whether the premium is positive or negative it can be earned by either entering into a long or short volatility position. Due to the negative correlation between volatility and market returns, selling volatility can be also seen as selling insurance against a market crash (Ilmanen, 2011). IVOL refers to the volatility expected by the market. It is derived by taking market prices on options and using the Black-Scholes formula as an interpolation tool (Hull, 2019). The most closely watched indicator for the S&P 500 implied volatility is an index called VIX, which shows the market-expected level of volatility 30 days in the future (CBOE, 2020). RVOL on the other hand refers to the actual standard deviation of returns of an asset (Hull, 2019).

In the case of the S&P 500, evidence suggests that IVOL is above RVOL most of the time, except the times of elevated volatility, when the opposite is usually observed (Ilmanen, 2011). The average VRP is estimated to be 2% according to Ilmanen (2011) and between 4% and 5%

based on Edwards and Preston (2017). While ex-post the VRP, as it was already mentioned, is calculated as the difference between IVOL and RVOL which is shown by Eraker (2008), Bollerslev et al (2009), Dufour et al (2012) and Ribeiro et al (2012), there is no consensus on the best suited methodology to estimate it ex-ante.

1.1 │VRP Estimation

Even though the amount of articles about volatility trading is limited, there is a disagreement about the best approach for estimating the VRP for a trading model. It is estimated either based on the term structure of VIX futures by Cooper (2013), Simon and Campasano (2014), Bordonado et al (2017) and Dondoni et al (2018) or based on the difference between the VIX and a forecast of RVOL (Cooper, 2013).

The majority of the literature estimates the VRP via the VIX futures term structure. As the VIX is mean reverting, its futures term structure adjusts to this property by being in contango when the current spot price is below its mean and in backwardation in the other case. Simon and Campasano (2014) use findings from Mixon (2007) and Nossman and Wilhelmsson (2009), who show that VIX futures do not have predictive power for the spot VIX, but the futures themselves and hence contain a premium. As a result, Simon and Campasano (2014) use this property and claim that VRP is nothing else than the difference between the VIX spot price and the price of the nearest VIX future. While Bordonado et al (2017) estimate the VRP in the same way, Dondoni et al (2018) also test futures with later expiration dates. Their findings however support the methodology from Simon and Campasano (2014) and Bordonado et al (2017) as the first expiring VIX future is indeed best suited to estimate the VRP based on the term structure.

Cooper (2013) uses a different approach and estimates the VRP based on the difference between IVOL and a forecast of RVOL. To forecast RVOL, he employs the simple moving average (SMA) and exponential generalized autoregressive conditional heteroskedasticity (EGARCH) models. In addition, Cooper (2013) also estimates the VRP based on VIX futures term structure, but achieves better results with an RVOL forecasting model generated VRP.

1.2 │VRP-based Trading

Once the VRP is estimated, it can be used for signal derivation and volatility trading. While the positive VRP always signals a short volatility position and vice versa, the papers analysed use two types of indicators: binary (Cooper, 2013) and threshold (Simon and Campasano 2014; Dondoni et al 2018). The binary indicator always follows the signal and suggests a long or short position according to it, whilst the threshold indicator is different as it compares the volatility risk premium forecast to a pre-set threshold and suggests a long or short position only if the condition is satisfied. In the literature, volatility is traded via two instruments: VIX futures (Simon and Campasano 2014; Dondoni et al 2018) and VIX ETNs (Cooper 2013; Bordonado et al 2017). What is more, Simon and Campasano (2014) and Bordonado et al (2017) also test hedged versions of their strategies. The performance of the generated models was outstandingly good: Simon and Campasano’s (2014) best result was achieved with a hedged portfolio and yielded a Sortino ratio of 1.15, Bordonado et al (2017) were able to generate a Sharpe ratio of 2.11 and a Sortino ratio of 3.17 with their unhedged portfolio and Cooper (2013) claims to reach a Sharpe ratio of 1.85 by estimating the VRP with an SMA model in an unhedged portfolio. Only Dondoni et al (2018) report a less remarkable Sharpe ratio of less than 0.5.

1.3 │Contribution

The comparison of the performance of the strategies we found in the literature does not allow us to arrive at a clear conclusion on which volatility trading approach is best due to a range of reasons. First of all, all the strategies were tested exclusively in the in-sample (IS) period, which makes fair performance evaluation and comparison impossible. Secondly, Cooper (2013) tested only SMA and EGARCH models for volatility forecasting ignoring the classic GARCH model as well as more recent variations of it. Thirdly, there is no clear evidence on whether hedged or unhedged models generate superior results, with Simon and Campasano (2014) and Bordonado et al (2017) reporting contradicting results. In addition, all previously cited volatility trading papers use either binary or threshold indicators and thus do not fully take the magnitude of the VRP and hence forecasting errors into account. Lastly, the backtests in the previously listed literature are conducted in the period from 2008 to 2013. As the trading volume of VIX derivatives has increased substantially since then, it is also not clear, how these models perform in a period of increased liquidity of VIX derivatives. We not only aim to answer the questions described in this paragraph, but also introduce a new scaled indicator that determines the allocation based on the magnitude of the VRP estimation and test our own volatility trading strategy that relies on newly developed RVOL forecasting models.

Those papers that use the terms structure approach rely on the well-documented relation between the VIX spot and front VIX future and thus do not leave much space for VRP estimation improvement. The paper by Cooper (2013), however, uses RVOL forecasting models, which are constantly reworked and improved. The most recent improvement for RVOL forecasting models are GARCH-MIDAS (GM) models which were introduced by Engle et al (2013). The idea behind these models is that there exist exogenous variables that have an impact on volatility and thus can be used in volatility forecasting. These variables can be macroeconomic factors as well as volatility observations of different frequencies. Asgharian et al (2013) demonstrate that GM models can deliver superior forecasts of S&P 500 volatility compared to GARCH models. GM models were further enhanced to Double Asymmetric GARCH-MIDAS (DAGM) models by Amendola et al (2019), who managed to capture volatility asymmetry with their extension. It was done by making both GARCH and MIDAS variable coefficients dependent on the sign of the respective observation in the dataset. To our knowledge, wea are the first to use GM and DAGM models to estimate the VRP for a VIX trading strategy.

1.4 │Outline

To wrap it up, we conduct a backtest that switches between long and short volatility positions based on the estimation of VRP. We derive the VRP based on the difference between the VIX and I) front end VIX futures, II) an SMA-based forecast of RVOL, III) a GARCH-based forecast of RVOL, IV) a GM/DAGM based forecast of RVOL. Out of these estimations, SMA, GARCH and VIX-Futures based ones denote the benchmarks which the GM/DAGM based estimations aim to beat. Each of these estimations is used for binary and threshold indicators, which is in line with existing literature and a scaled indicator, to take the estimation accuracy into account. What is more, all models are tested in a hedged and an unhedged portfolio to identify the best-suited approach. The model-derived estimations of VRP are then applied to a backtest discussed in the further chapters. The backtest period is 01.01.2007 – 27.11.2020, with the out-of-sample (OOS) period starting on 01.01.2016. To compare the models on a risk-adjusted basis, the results are assessed based on the OOS Sharpe ratio.

1.5 │Findings

Our results show that there are specifications that allow DAGM models to outperform conventional and VIX futures term structure implied VRP models. Our best performing models are not only able to achieve a Sharpe ratio of more than 1, but exhibit a negative correlation to the S&P 500 and a positive skewness of daily returns, making it an interesting strategy for portfolio diversification. Moreover, we demonstrate that the performance of VIX trading strategies does not vanish over the years, but is rather heavily impacted by the market regime in which they are evaluated. In addition, we find clear evidence that unhedged models outperform hedged models, even on a risk-adjusted basis. The results further demonstrate that establishing a threshold for the VRP as a trading indicator yields the best results, but at the same time the choice of the indicator is heavily influenced by the underlying applied forecasting model.

1.6 │Structure

The remainder of the paper is organized as follows: Section 2 is dedicated to the trading instruments and corresponding data that we used, Section 3 introduces the forecasting models, different VRP estimations, indicator derivations, and the backtest set-up and Section 4 reports the results of our backtest, which are concluded in Section 5.

2 │Data

2.1 │Trading Instrument Selection

"Trading the VIX" is not as straightforward as it might sound, as it is a purely mathematical indicator that shows the level of volatility expected by the market in 30 days and is therefore not tradable. There are however derivatives that allow to trade the volatility of the S&P 500, such as VIX ETNs, VIX futures and SPX options. The most obvious candidates for trading volatility would be SPX options as the VIX is derived based on their level of implied volatility. By constantly delta hedging and adjusting the exposure to replicate the VIX, one could use SPX options, but there is no paper known to us that applies this methodology for a volatility trading model. As we have already stated, VIX Futures and VIX ETNs are the most popular tools for backtesting a volatility strategy in the literature, which is why we will examine these instruments more closely.

VIX futures are the "purest" form of trading the VIX, as they provide a payoff at expiration which is determined by the difference between the futures price at the purchase date and the VIX spot price at the expiration date. The price of the futures is the market price for 30 day implied volatility beginning at the expiration date (McFarren, 2013). It is important to note that the lack of a purchasable spot price implies that there is no general cost-of-carry relationship between the spot and the future's price (Bittman 2007; Alexander et al 2015). The term structure of the VIX takes its mean-reverting nature into account by being in contango when the current spot price is below its mean and in backwardation in the other case. As markets crash less often than they do not, the term structure is in contango much more often than it is in backwardation. The steepness of the term structure is determined by how far the VIX is from its long-term mean. As a long position in the VIX acts as an insurance against a drop in the market, the expected return of continuously holding a long position in VIX futures is negative, which is demonstrated by the VIXSTER index, which replicates a rolling long VIX futures position and lost more than 94% of its value from 2008 until 2013 (Whaley, 2013).

An evolution of VIX futures are VIX ETNs, which have become increasingly popular in recent years and are a simple way of gaining exposure to VIX futures. There is a big range of VIX ETNs available which vary mainly based on the amount of leverage and exposure to different maturities of the underlying futures. The most heavily traded VIX ETN is the VXX, which aims to replicate the payoff of a long VIX futures position. Replication is achieved by constantly weighting and rolling futures in such a way that they always have an average expiration date of 30 days. This ETN is the most attractive ETN for trading volatility in our case, as VIX ETNs that use leverage would require a view on volatility-of-volatility, short VIX ETNs provide a less than 100% inverse exposure and Mid-Term Futures ETNs exhibit lower sensitivity to movements in the spot price of the VIX. It is noteworthy that just like their underlying VIX, long VIX ETNs have a negative expected return in the long run due to the mostly negative roll yield (Stanton, 2011). The advantage of VIX ETNs over VIX futures is that they require less initial capital and no constant rolling from the purchaser, which makes them a very attractive instrument for trading. As a result of their attractiveness, it can be observed that the demand for VIX ETNs drives the prices of the underlying VIX futures (Alexander and Korovilas 2013; Bollen et al 2017). VIX ETNs are used for volatility trading strategies in papers by Cooper (2013) and Bordonado et al (2017). Consequently, we will stick with the VIX ETN that aims to replicate the payoff of a long VIX futures position in this paper, as this makes our results comparable, not only to other papers using VIX ETNs, but also to papers using front-end VIX futures.

2.2 | Data Overview

Once we have established which trading instrument to choose, we can provide an overview of the data. If available, all daily datasets start on 20.12.2005, weekly on 30.12.2005 and monthly on 01.01.2006. The data used has been collected from Yahoo Finance, FRED, Refinitiv and the Realized Library from the Oxford-Man Institute of Quantitative Finance (Heber et al, 2009).

Table 1: Overview of the used data

Dataset	Time Frame	Source	Frequency
S&P 500	20.12.2005 – 27.11.2020	Yahoo Finance	Daily
SPVXSPID	20.12.2005 – 27.11.2020	Refinitiv	Daily
VIX	20.12.2005 – 27.11.2020	Yahoo Finance	Daily
VIX3M	13.11.2007 – 27.11.2020	Refinitiv	Daily
CVXCS00	20.12.2005 – 27.05.2020	Refinitiv	Daily
VIX 1st Future	01.01.2013 – 27.11.2020	CBOE	Daily
RV5	20.12.2005 – 01.12.2020	Oxford-Man Institute	Daily
NFCI	30.12.2005 – 18.12.2020	FRED	Weekly
INDPRO	01.01.2006 – 01.11.2020	FRED	Monthly
CFNAI	01.01.2006 – 01.11.2020	FRED	Monthly
HOUST	01.01.2006 – 01.11.2020	FRED	Monthly

Table 1 reports a description of the data sets used in the remainder of the thesis. In total, we use 11 datasets for our backtests. We use the SPX ticker for the S&P 500 data, SPVXSPID is the VIX ETN of our choice, VIX is the previously discussed implied volatility index and CVXCS00 is the proxy for the VIX 1st future that we will address in more detail in the following paragraphs. The remaining datasets VIX3M, RV5, NFCI, INDPRO, CFNAI and HOUST are the macroeconomic and volatility variables that are used exclusively in the GM and DAGM models and thus we will elaborate on them when introducing the respective models.

The period from 20.12.2005 – 31.12.2006 is used for gathering data, as some forecasting models require historic data of up to one year, and hence no optimization takes place in this period. All model parameters are optimized in the IS period from 01.01.2007 – 31.12.2015 and obtained parameters are then also used in the OOS period from 01.01.2016 – 27.11.2020. The reason for selecting the respective periods is that the IS period contains the global financial crisis of 2008, the subsequent bull market and a sideways market in 2015 and thus covers several different “regimes”. Similarly, the OOS period contains the bull market of 2016 and 2017, the ultra-low volatility environment of 2017, the second-largest documented one day spike of the VIX in 2018 with a subsequent sideways market in 2018 and the COVID-19 triggered market crash of 2020.

As a result of selecting these periods, there are a few issues that must be solved. First and foremost, the VXX was launched in January 2009 and replaced by its successor VXXB in January 2018. While the methodology of the VXXB is the same as for VXX, using VXX and VXXB would mean that we are not able to show backtest results for the 2008 financial crisis. Therefore, we will use the SPVXSPID index as a substitute for the VXX over the whole timeframe of the backtest. The SPVXSPID is an index that replicates the return of constantly weighing and rolling the two front end VIX futures such that they have an average expiration of 30 days and thus acts as the benchmark for the VXX. It is important to note that the SPVXSPID, just like the VXX itself, was subject to several reverse-share (or index for that matter) splits. As a result, it had a value of 495.31 on the first day of our OOS period, which implies that a bid/ask spread that is defined in absolute terms results in the underestimation of the trading costs (Barclays, 2019). Even though this issue vanishes over time, which is demonstrated by our index having a value of 25.47 on the last day of the OOS period, it is something that has to be kept in mind when discussing the assumptions for trading costs, which we do at the end of the next section.

What is more, the time-series for the front-end VIX future starts on the first trading day of 2013, which is why we use the CVXCS00 index as a proxy for the time before. As this is not a perfect substitute, we will stick with the VIX 1st Future time series from 01.01.2013 onwards, which means that the OOS period only contains the “original” time series. In addition, data for the 3-

month-VIX, which is required for the derivation of several indicators, is only available from 13.11.2007, which means that it is not possible to compute RVOL forecasts before that date. As a result, for this specific variable, no trading takes place until enough data is available. While this has an impact on the IS results of the backtests that use the VIX3M, it will not influence the validity of our results, as the data is available for the full time frame of the OOS period, which is what the models will be judged by. Table 2 provides a statistical description of the data that we use.

Table 2: Statistical properties of the used data

Dataset	count	mean	std	min	max	skew	kurt
SP500	3762	0.04%	1.28%	-11.98%	11.58%	-0.27	13.24
SPVXSPID	3762	-0.12%	4.53%	-25.95%	96.12%	3.54	59.08
VIX	3762	0.32%	8.07%	-29.57%	115.60%	2.10	16.55
CVXCS00	3631	0.00%	4.59%	-24.84%	86.81%	2.78	39.18
VIX 1st Future	3762	0.16%	5.56%	-30.77%	112.41%	3.33	49.36
RV5	3754	30.45%	122.99%	-93.86%	1917.71%	5.23	51.14
VIX3M	3620	0.13%	5.03%	-20.86%	65.18%	1.99	14.81
indpro	178	0.03%	1.35%	-12.68%	6.24%	-4.52	47.84
CFNAI[1]	179	-19.69%	158.20%	-1773.00%	596.00%	-7.43	86.83
HOUST	178	0.20%	9.16%	-26.40%	23.96%	0.12	-0.11
NFCI[2]	779	-28.35%	61.29%	-77.63%	272.01%	2.77	8.43

Table 3 describes the data further and allows us to draw similar conclusions to what was already pointed out previously: There is a strong negative relationship between the S&P 500 and the VIX, as well as VIX-related derivatives, which supports the thesis presented by Ilmanen (2011) that a long volatility position resembles buying insurance against a market decline. It also should not come as a surprise that there is a strong positive relationship between the VIX 1st future and the SPVXSPID, which rolls front-end VIX futures as well as the VIX and VIX

[1] CFNAI is a 0-based index that displays absolute change. Therefore, no percentage change was taken, and the min/max values show the absolute change

[2] The same applies as to CFNAI.

derivatives which, despite the diverging returns to spot VIX, are driven by movements in the VIX. What is most surprising is the relatively weak correlation between the high-frequency volatility data RV5 and all other datasets, which could hint at additional information that can be extracted by including RV5 in forecasting models.

Table 3: Correlation matrix of volatility related time series

	1	2	3	4	5	6	7
1. SP500	-						
2. SPVXSPID	-0.71	-					
3. VIX	-0.72	0.89	-				
4. 3MVIX	-0.77	0.92	0.95	-			
5. RV5	-0.22	0.29	0.35	0.32	-		
6. CVXCS00	-0.60	0.86	0.78	0.80	0.28	-	
7. VIX 1st Future	-0.61	0.88	0.81	0.82	0.29	0.81	-

3 │Methodology

3.1 │Implied Volatility

3.1.1 │*Derivation*

The level of implied volatility (IVOL) is derived from the assets volatility surface by taking market prices of options as given and then solving the Black-Scholes Formula for sigma (Hull, 2019). In a world that follows the Black-Scholes assumptions, one could take the market price of any option on the same underlying, solve for sigma and retrieve the level of volatility that is implied by the market (Black and Scholes, 1973). While this might be accurate in theory, there is broad evidence that in reality not all of the Black-Scholes assumptions hold and as a result, different levels of IVOL will be retrieved, depending on the delta and the time till the expiration of the option. Retrieving IVOL levels across different levels of delta and time till expiration finally allows to generate what is known as the volatility surface, which can take on different shapes, depending on the underlying asset class. As a result, Black-Scholes is essentially used as an interpolation tool that allows to retrieve the volatility surface (Hull, 2019).

As this paper aims to forecast the RVOL of the S&P 500, the index that we use as a proxy for IVOL is the VIX. The VIX was developed with the goal to provide a single index that shows the market's expectation of the 30-day volatility of the S&P 500 (CBOE, 2020). Equation 1 demonstrates, how the market implied level of variance is derived for a single expiration date.

$$\sigma^2 = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2} e^{RT}\, Q(K_i) - \frac{1}{T}\left[\frac{F}{K_0} - 1\right]^2 \tag{1}$$

with:

$$\Delta K_i = \frac{K_{i+1} - K_{i-1}}{2}$$

$$F = K_{ATM} + e^{RT} * (C_{ATM} - P_{ATM})$$

In the equation, T denotes the time till expiration, F the forward index level, K_0 the closest strike below F, K_i the strike price of the i^{th} OTM option, R the risk free rate and $Q(K_i)$ the mid-price for each option with the strike K_i. In the derivation of the forward index level, C_{ATM} and P_{ATM} represent the at-the-money call and put prices and K_{ATM} the at-the-money strike price, where the at-the-money level is determined via put-call-parity by the smallest difference between the mid-price of put and call options (CBOE, 2020).

The VIX considers standard SPX options that expire on the 3rd Friday of the month and weekly SPX options that expire every Friday, except for the 3rd Friday of the month. As IVOL differs depending on the strike price and the time till expiration, and there does not exist an option on every day that expires exactly 30 days from the current trading day, it is not enough to use put and call options across varying strikes, but this procedure must be used for two different expiration dates. As a result, SPX put and call options are considered as long as they have more than 23 and less than 37 days till expiration, which results in options with two different expiration dates being considered, with the one closer to expiration being the “near-term” option and the one with the later expiration being the “next-term” option. Once the near-term option has reached 23 days till expiration, the next-term option becomes the near-term option and a new next-term option is considered. In addition, the considered options are all European, out of the money and have non-zero bid prices which results in the possibility that the number of options that are considered can change with high frequency (CBOE, 2020).

$$VIX = 100 * \sqrt{\left\{T_1\sigma_1^2\left[\frac{N_{T2}-N_{30}}{N_{T2}-N_{T1}}\right] + T_2\sigma_2^2\left[\frac{N_{30}-N_{T1}}{N_{T2}-N_{T1}}\right]\right\} * \frac{N_{365}}{N_{30}}} \tag{2}$$

These steps allow to calculate two values for σ^2, σ_1^2 (IVOL based on near-term options) and σ_2^2 (IVOL based on next-term options). To derive the 30-day level of IVOL, these results are weighted in a way that their time till expiration is 30 days, as demonstrated by equation 2. While most of the variables for the formula above have already been introduced, some of them are new – N_{T1} (number of minutes to a settlement of the near-term options), N_{T2} (number of minutes to a settlement of the next-term options), N_{30} (number of minutes in 30 days) and N_{365} (number of minutes in a 365 day year). The result of this formula is the VIX, which will be taken as the level of S&P 500 IVOL in this paper (CBOE, 2020).

3.1.2 | *Properties*

From a statistical point of view, the VIX has the same stylized facts as RVOL, which will be covered in more detail in the next sub-section. There are further non-statistical properties however that are unique for the VIX due to the way it is calculated. As the formulas state, the VIX takes the sum of all available options to derive a level of variance instead of weighting them. Consequently, the level of the VIX also depends on the number of options that qualify for entering the equation, which theoretically means that the VIX could rise despite a decreasing level of IVOL. This is a highly unlikely scenario, however, as, other things equal, a lower level of IVOL results in lower options prices, which in its turn leads to more options with a 0-bid price, thereby reducing the number of options that are considered for the VIX. Moreover, the formula for the VIX also implies that options that are close to the at-the-money level are responsible for the bigger portion of the final value than the deep-out-of-the-money options. It is so as the main determinant of the impact of an option on the final level of VIX is its price. This becomes apparent when analyzing equation 1: The only term that is not the same for every option is $\frac{\Delta K_i}{K_i^2} e^{RT} Q(K_i)$. If the strike price interval stays constant, the only factors determining the weight of an option in the index are its strike price and price, with the option price playing the bigger role. As options become more expensive the closer they are to the ATM level, they play a larger role in determining the level of the VIX than options that are further away from the ATM level.

3.2 | Realized Volatility

RVOL refers to the standard deviation of asset returns, which for a single period can be simplified to the square root of the squared return of the underlying, assuming a one period mean return of 0. Before we talk about calculating RVOL though, we would like to discuss its notable properties that would be relevant for the further sections, where the volatility forecasting models we use are discussed. The first property is called volatility persistence, which means that large and small moves of RVOL in either direction tend to cluster. Consequently, large changes are followed by more large changes and vice versa (Mandelbrot 1963; Fama 1965). The second stylized fact of RVOL is mean reversion, which states that RVOL always eventually returns to its long-term average. Another empirical observation is RVOL asymmetry – negative and positive equity returns affect realized volatility differently, with negative returns having a bigger impact on future volatility. This leads not only to a negative correlation between RVOL and equity markets but is also reflected in the skew of implied volatilities of OTM equity options. In addition to the above-presented properties, various exogenous variables can have an impact on volatility, as asset prices are not independent of the market surrounding them. Lastly, asset returns are known to follow a leptokurtic distribution, which is another characteristic that has to be taken into account when talking about volatility modelling (Engle and Patton, 2007).

Coming back to calculating RVOL, the formula for it is shown in (3), where S_i denotes the closing price of the asset, u_i the asset's corresponding percentage change, and σ_n the level of RVOL on day n. To annualize RVOL, the square root of the result is then multiplied with the square root of the number of trading days, which is set to 252 for equities (Hull, 2019).

$$\sigma_n = \sqrt{u_{n-i}^2} * \sqrt{252} \quad (3)$$

with:

$$u_i = \frac{S_i - S_{i-1}}{S_{i-1}}$$

Applying this formula to the S&P 500 RVOL, the relationship between RVOL and changes in the S&P 500 becomes apparent. Assuming one day is one period, the one-period RVOL is simply the absolute value of the daily S&P 500 change, multiplied by the square root of 252. Having defined the estimation of RVOL, the next step is to apply forecasting models to derive future estimations of RVOL.

3.2.1 | *SMA*

The first forecasting model that we use is the SMA model, which Cooper (2013) deems to be the most suitable one. It is also the simplest forecasting model as it assumes that future volatility will be the mean of a certain number of historic observations of volatility (Hull, 2019). While this model takes mean reversion into account, it only partially captures persistence and completely ignores asymmetry of volatility (Marra, 2015). The model from Cooper, however, is a modification of a "classic" SMA model as shown by Hull (2019). While Cooper's (2013) SMA estimates RVOL in the same way as the usual SMA model does, he then also smoothes the VRP estimation by taking a simple moving average of the VRP forecasts. Therefore, we use two different models: A classic SMA model and the modified Cooper's (2013) version. We use them with his specification as well as with the specification achieved from our in-sample optimization. Both models employ formula 4 for their forecast, where the reader is already familiar with all the variables besides m, which denotes the number of historic observations:

$$\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}^2 \tag{4}$$

Subject to:

$$0 < m \leq 252$$

The model is then optimized by choosing the number of historical observations that minimize the MSE in the IS period.

3.2.2 │*GARCH*

In order to better adjust for mean reversion and persistency of RVOL, Engle (1982) introduced the ARCH model, which improves the SMA by giving some weight to a constant that measures the long term-term average of variance and applies an autoregressive weighting scheme on historic observations, instead of taking the mean. This model was further improved to GARCH by Bollerslev (1986), who additionally assigned a weight to previous estimations of volatility. Consequently, the GARCH model allows to capture both mean reversion and persistence of volatility (Marra, 2015). We stick to a GARCH (1,1) model, as defined by equation 5 and employ a Student-t distribution for the innovations to account for the fat tails of the S&P 500 return distribution (Engle and Patton, 2007).

$$\sigma_n^2 = \omega + \alpha u_{n-i}^2 + \beta\sigma_{n-i}^2 \tag{5}$$

Subject to:

$$\omega, \alpha, \beta \geq 0$$

The variables that are yet to be defined are α, which denotes the weight that is assigned to the most recent observation of RVOL, β which shows the weight assigned to the most recent forecast of the model and ω which is a constant representing the long-run variance. The restrictions are that α, β and ω must be larger than or equal to zero. The parameters of the model are fitted using Maximum Likelihood.

3.2.3 | *GARCH-MIDAS*

A further step in the development of volatility forecasting models was the GARCH-MIDAS (GM) model that was introduced by Engle et al (2013). While GARCH models were introduced in 1986 by Bollerslev, MIDAS was only introduced in 2006 by Ghysels et al. GM capitalizes on the previously mentioned fact that exogenous variables can have an impact on volatility. Using a GARCH model in a MIDAS framework allows to predict volatility not purely based on past volatility observations, but also based on macro-economic factors and volatility observations of different frequencies. Asgharian et al (2013) demonstrate that GM models can deliver superior forecasts of S&P 500 volatility compared to GARCH models. In addition, Conrad et al (2018) show that GARCH-MIDAS models are also able to deliver better results than GARCH models when it comes to predicting the volatility of cryptocurrencies. We use the model as it was defined by Conrad et al (2018).

Here, in equation 6, the daily return on the i th day of the period t is defined as

$$r_{i,t} = \sqrt{\tau_t * g_{i,t}} \tag{6}$$

where τ_t is the long-run component varying every time period t and $g_{i,t}$ is the short-run term varying each day i in the period t.

The short-run component is a GARCH-like equation demonstrated in (7). Same as with GARCH, we use Student-t conditional error density.

$$g_{i,t} = (1 - \alpha - \beta) + \alpha \frac{(r_{i-1,t})^2}{\tau_t} + \beta g_{i-1,t} \tag{7}$$

The long-run component is defined in (8)

$$\tau_t = \exp\,(m + \theta \sum_{k=1}^{K} \delta_k(\omega) X_{t-k}) \tag{8}$$

where K is the number of periods over which we smooth the volatility, X_t is the explanatory variable and $\delta_k(\omega)$ is a weighting scheme.

Furthermore, we test two weighting schemes for our models – Beta and Exponential. The Beta weighting scheme is defined as (9),

$$\delta_k(\omega) = \frac{(k/K)^{\omega_1 - 1}(1-k/K)^{\omega_2 - 1}}{\sum_{j=1}^{K}(j/K)^{\omega_1 - 1}(1-j/K)^{\omega_2 - 1}} \tag{9}$$

where k is the lag of interest, K [3] is the number of realizations to be considered and ω_1, ω_2 are parameters governing the weights of each k lag. For the exponential weighting schemes the set of variables stays the same, but the formula becomes (10).

$$\delta_k(\omega) = \frac{exp(\omega_1 k + \omega_2 k^2)}{\sum_{k=1}^{K} exp(\omega_1 k + \omega_2 k^2)} \tag{10}$$

3.2.4 | *Double Asymmetric GARCH-MIDAS*

DAGM is an extension of the GM model based on Amendola et al (2019). It changes both the short-run GARCH and the long-run MIDAS equations by introducing adjustments that help the model to capture volatility asymmetry. The model is called double asymmetric as it changes both the GARCH and MIDAS equations in a way that they become responsive to the sign of the realized volatility or MIDAS variable inputs.

[3] We have simulated backtests with different lag lengths from 1 to 20 and there was no lag length that achieved significantly better or worse results. As a result, we will report the results of having a lag length of 4 and 8, which corresponds to 1 month and 2 months of historic observations in case of weekly frequency and to 4 months and 8 months in case of monthly frequency.

The short-run component starts using GJR-GARCH instead of a regular GARCH as a basis, which is reflected in equation 11,

$$g_{i,t} = (1 - \alpha - \beta - \gamma/2) + (\alpha + \gamma * 1_{(r_{i-1,t}<0)}) \frac{(r_{i-1,t})^2}{\tau_t} + \beta g_{i-1,t} \tag{11}$$

where the newly introduced γ is the asymmetry coefficient and $1_{(r_{i-1,t}<0)}$ is an indicator function that is non-zero only if the argument is true. Similarly, our DAGM high-frequency component uses the Student-t distribution of errors to account for leptokurtic return distribution.

The long-run MIDAS equation becomes (12)

$$\tau_t = \exp\left(m + \theta^+ \sum_{k=1}^{K} \delta_k(\omega)^+ X_{t-k} 1_{(MV_{t-k} \geq 0)} + \theta^- \sum_{k=1}^{K} \delta_k(\omega)^- X_{t-k} 1_{(MV_{t-k}<0)}\right) \tag{12}$$

with two variations of θ being sign-specific asymmetric responses to changes in the underlying MIDAS variable and $\delta_k(\omega)$ being corresponding beta or exponential weighting scheme described in the GARCH-MIDAS section. For the DAGM models, we use the same set of MIDAS variables and the same set of K values as discussed in the GM subsection.

3.2.5 | *GM/DAGM Explanatory Variables*

The explanatory variables we use for X_t can be split into two parts: macroeconomic and volatility. The macroeconomic variables we use are provided by the FRED and are Industrial Production Index (INDPRO), National Financial Conditions Index (NFCI), New Privately Owned Housing Units Started dataset (HOUST) and Chicago Fed National Activity Index (CFNAI). The same macroeconomic variables were used by Conrad and Kleen (2020), INPRO and CFNAI were also used by Amendola et al (2019). The volatility variables set is the intraday daily realized variance calculated using 5-minute intervals (RV5), VIX (VIX), roll yield calculated as the absolute difference between the front-end VIX future and VIX spot (RY) and the volatility surface shape calculation that is the absolute difference between the VIX3M and the VIX (3MVIX). While realized variance and VIX as explanatory variables are not a novelty as they can be found in a paper by Conrad and Kleen (2020), RY and 3MVIX are something that is unique to our article. INDPRO, CFNAI and HOUST variables are observed at a monthly frequency, while the rest of the variables are observed at a weekly frequency. As originally RV5, VIX, RY and 3MVIX variables had daily frequencies, the frequencies were changed to weekly by assuming that the weekly value at the start of each week is the average of all daily observations in the previous week.

3.3 │VRP Calculation

3.3.1 │*RVOL implied VRP*

We have already outlined the most notable differences between RVOL and IVOL in the previous sections. RVOL exclusively represents the historic standard deviation of an asset and is unique for an asset. The VIX on the other hand contains the market's view on the future and is therefore a forward-looking index (Whaley, 2009). Ilmanen (2011) and Edwards and Preston (2017) among others showed that on average the VIX contains a premium over RVOL, the formerly discussed VRP. Ex-post, the VRP is calculated as the difference between IVOL and RVOL as shown in the literature (Eraker 2008; Bollerslev et al 2009; Ilmanen 2011; Dufour et al 2012; Ribeiro et al 2012). Bollerslev et al (2009) use the VIX to generate a model-free implied variance level IV_t and obtain realized variance RV_t in a similar model-free fashion, which leads to their estimation of a variance risk premium, shown in (13).

$$VRP_t = IV_t - RV_t \tag{13}$$

In our case, we use the aforementioned forecasting models to calculate RV_t and actual VIX data for IV_t. As shown in the whitepaper by CBOE (2020), the VIX displays an annualized level of volatility and is multiplied by 100, which is why we have to apply the same procedure to our estimations of RVOL, with the resulting formula being displayed in (14).

$$\widehat{VRP}_t = VIX_t - \sqrt{\sigma_n^2} * \sqrt{252} * 100 \tag{14}$$

We will use this formula to estimate the VRP based on the “classic” SMA, GARCH and all GARCH-MIDAS models. As already described, Cooper (2013) smoothes the difference between the SMA estimation and the VIX to retrieve an estimation of VRP, which is why he estimates the VRP with equation 15. To be able to replicate his results, we will estimate the VRP for his modified SMA model with $\widehat{VRP_{t_c}}$.

$$\widehat{VRP_{t_c}} = \frac{1}{m}\sum_{i=1}^{m} \widehat{VRP}_{n-i} \tag{15}$$

3.3.2 │*Term structure implied VRP*

The thesis that the term structure includes a risk premium is suggested by studies from Mixon (2007) and Nossman and Wilhelmsson (2009), who show that VIX futures do not have predictive power for the VIX spot price unless a risk premium is included. In the context of a trading model, this means that when the term-structure is in contango, the VIX is not expected to increase to the level of the future by its expiration date. It is true, however, that the term-structure does have predictive power for subsequent VIX futures prices, both when it is in contango and backwardation. The substantial forecasting power suggests that the term-structure can be used for a trading strategy, even though it has to be kept in mind that in a regression, the term-structure only accounts for about 10% of the variation in VIX futures price changes and hence this strategy carries a big exposure to equity markets, due to the strong negative correlation between VIX futures and the S&P 500 (Simon and Campasano, 2014). While this exposure could be hedged by shorting the S&P 500 when a short volatility position has been opened, and vice versa, Bordonado et al (2017) show that leaving these positions unhedged yields a significantly higher Sharpe ratio. As Bordonado et al (2017) were able to generate the highest Sharpe ratio among all the papers we analysed, we will estimate the VRP based on the roll yield both in the way that they suggest it in their paper ($\widehat{VRP_{t_b}}$), as shown in equation 17, as well as in a way that is consistent with our RVOL implied VRP estimation $\widehat{VRP_t}$ ($\widehat{VRP_{t_{RY}}}$) (16). In the equations below, $VX1_t$ denotes the price of the front end VIX future at time t.

$$\widehat{VRP}_{t_{RY}} = VX1_t - VIX_t \tag{16}$$

$$\widehat{VRP}_{t_b} = \left(\frac{VIX_t}{VX1_t} - 1\right) * 100 \tag{17}$$

Unless stated otherwise, we will from now on refer to all model estimated VRPs with $\widehat{VRP}_t$.

3.4 │Trading Indicator

Once the different estimations of the VRP have been conducted, the results are used to generate a trading indicator. The two main questions are when to go long or short and how big the allocation should be. Whether we go long or short volatility is based on the approach in the literature, where $\widehat{VRP}_t > 0$ indicates a short volatility position and $\widehat{VRP}_t < 0$ a long volatility position[4] (Cooper 2013; Simon and Campasano 2014; Dondoni et al 2018). In the next step, we can address the issue of how big the allocation to volatility should be. We denote the maximum available capital that can be allocated to a volatility position with γ, which is a variable that can be selected based on risk tolerance and margin requirements.

3.4.1 │*Binary*

The binary indicator is the most straightforward type of indicator and allocates the maximum amount of capital available to a long SPVXSPID position if $\widehat{VRP}_t < 0$ and vice versa when $\widehat{VRP}_t > 0$. This indicator is also used by Cooper (2013), and is not only the simplest, but also a very convenient one, as there are no variables that must be selected.

[4] We also pursued another approach where the long/short decision is based on the mean of the VRP ($\widehat{VRP}_m$). $\widehat{VRP}_m$ was estimated by: $\widehat{VRP}_m = \frac{1}{m}\sum_{i=1}^{m}\widehat{VRP}_{t_i}$. This approach yielded worse results across all indicators and models, which is why we decided not to report the results.

3.4.2 | *Threshold*

The threshold model establishes a threshold that $\widehat{VRP_t}$ must surpass to be considered large enough for a trading indicator. In the model, t^u denotes the upper threshold that has to be surpassed by $\widehat{VRP_t}$, t^l the lower threshold and k the amount of standard deviations from the mean that are considered to be relevant, as demonstrated in formulas 18 and 19. The drawback compared to the other indicators is that the variable k has to be selected arbitrarily. We set k to 0.25, 0.5, 0.75 and 1, which results in four threshold indicators for each model. Setting k lower than 0.25 would make the threshold indicator too similar to the binary indicator, and k > 1 would mean that we trade only very rarely.

$$t^u = +k\sigma \tag{18}$$

$$t^l = -k\sigma \tag{19}$$

Variations of this indicator have been used in the papers by Simon and Campasano (2014), Bordonado et al (2017) and Dondoni et al (2018). As Bordonado et al (2017) achieved the highest Sharpe ratio, we will also test his threshold when using $\widehat{VRP_{t_b}}$to replicate his model. He enters a position if $\widehat{VRP_{t_b}}$ > 8% or $\widehat{VRP_{t_b}}$ < -8% and closes an existing position if -6% < $\widehat{VRP_{t_b}}$< 6%. It is important to note that due to how $\widehat{VRP_{t_b}}$ is calculated $\widehat{VRP_{t_b}}$ > 8% indicates a long volatility position and $\widehat{VRP_{t_b}}$ < -8% a short volatility position, which is in contrast to our models, which have been defined previously.

3.4.3 | *Scaled*

The scaled indicator decides in the same way whether a long or short position is entered as the binary indicator, but the allocation to the position depends on how big $\widehat{VRP_t}$ is. It is important to note that we will use the absolute value of $\widehat{VRP_t}$, as otherwise very large negative values of $\widehat{VRP_t}$ will hardly be considered, due to the structure of the indicator. Therefore, $\left|\widehat{VRP_t}\right|$ is scaled between 0 and 1 using feature scaling, as described by equation 20. In the formula, x is the most recent observation of $\left|\widehat{VRP_t}\right|$, $\min(x)$ is the smallest observed value of $\left|\widehat{VRP_t}\right|$ and $\max(x)$ is the largest observed value of $\left|\widehat{VRP_t}\right|$ that are known up to the respective point in time.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{20}$$

The scaled indicator is then defined by the formula 21, where whether a long or short position is entered depends on whether ω_s is larger or smaller than 0.

$$\omega_s = \widehat{VRP_t} * x' \tag{21}$$

A big advantage of this model is that it makes the allocation directly dependent on $\widehat{VRP_t}$ and hence takes the estimation accuracy into account. This is a big difference to the binary indicator, where the only important factor is whether the directional forecast of RVOL was correct. As this is the only indicator where the allocation is not equal to the maximum possible allocation γ, the return of this model will be significantly smaller than the return of the binary or threshold model, since the only time the allocation equals γ is when a new all-time high of $\left|\widehat{VRP_t}\right|$ is achieved. This is not a problem for us however, as the results are assessed based on the Sharpe ratio. It also has to be noted that there is no paper known to us that uses feature scaling in the context of a volatility trading model.

3.5 | Backtest

Once the VRP estimation and indicator derivation have been conducted, there are a few decisions left before the simulation can be done. First of all, a value for γ, the maximal capital allocation is required. We set γ to 25%, as a lower level of γ decreases the likelihood of a margin call if the strategy was implemented under real-world conditions. Setting it even lower than 25% would mean that the exposure of any model that uses the scaled indicator would approach 0, as γ denotes the upper limit, which is hardly ever reached by the scaled indicator. In the hedged portfolios, we additionally need to determine a hedge ratio of how much capital should be allocated to the S&P 500 and the SPVXSPID. As on average, a -1% move in the S&P 500 corresponds to a +2.25% move in the VXX [5] (and for that matter the SPVXSPID), we will allocate $\gamma/2.25$ per cent of the portfolio to the SPVXSPID and $\gamma\%$ to the S&P 500 with a short position in the VXX corresponding to a short position in the S&P 500 and vice versa. What is more, we set the bid/ask spread to 0.06% per share, which is the 30-day median for the VXXB quoted by Fidelity[6]. In addition, we assume that it is possible to trade infinitely small fractions of shares of the SPVXSPID and the S&P 500, as otherwise, the reverse-split time series would heavily bias our results in the IS and the early stages of the OOS period. Lastly, it is assumed that there are no shorting fees and that our cash positions do not earn any interest.

Next, we can demonstrate the decision algorithm in figure 1 below. Every day at market close, each models estimation of the VRP is observed and used as an indicator. The binary and scaled indicators yield one backtest each, whereas the threshold indicator yields four backtests, with the levels of k being set to 0.25, 0.5, 0.75 or 1. Consequently, each estimation of VRP yields 6 portfolios, which will be denoted with b (binary), s (scaled) and t0_25, t0_50, t0_75 and t0_100 (threshold). All portfolios are rebalanced daily, except for the threshold model, which, depending on the indicator, might not enter a position on several consecutive days. It is worth noting that for the replication of the best benchmark model from Bordonado et al (2017) we used the flow-diagram outlined in his paper.

[5] Source : https://sixfigureinvesting.com/2014/08/hedging-market-with-vix-vxx-uvxy-vix-futures/, retrieved at 14.03.2021

[6] Source: https://screener.fidelity.com/ftgw/etf/goto/snapshot/quote.jhtml?symbols=VXX, retrieved at 27.02.2021

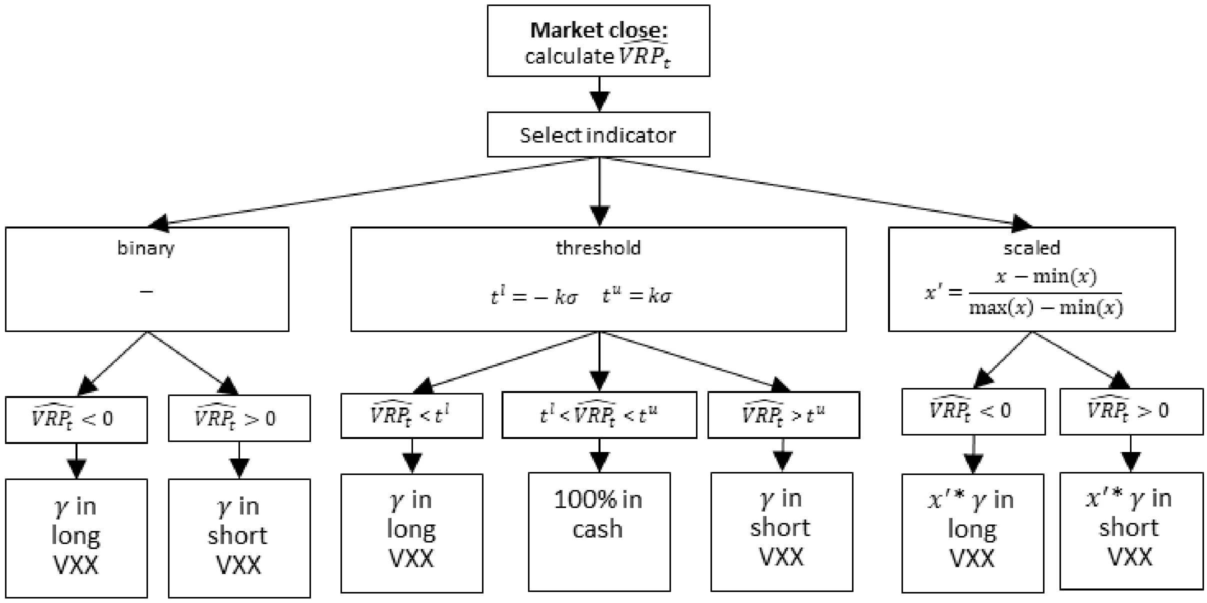

Figure 1: Decision-making process in the backtest. The allocation at the bottom of the figure displays the allocation in the unhedged portfolio, for the hedged portfolio the allocation to the SPVXSPID is divided by 2.25 and the allocation to the S&P 500 is γ, with a short SPVXSPID position corresponding to a short S&P 500 position.

The backtest period lasts from 01.01.2007 – 27.11.2020, which results in the backtest using IS forecasts from 01.01.2007 – 31.12.2015 and OOS forecasts from 01.01.2016 – 27.11.2020. As the goal of this paper is to identify which forecasting model is best suited for a volatility trading model, the forecasting models will be assessed based on the performance of the backtest in the OOS period. To assess the models on a risk-adjusted basis, the backtests will be assessed based on their OOS Sharpe ratio.

4 │Empirical Results

4.1 │Analysis Across All Models

In the first step, we will compare specifications that are shared by all backtests in order to identify whether any backtesting approaches are clearly worse and can therefore be excluded from the detailed analysis. The first comparison we are going to make is that of the hedged and unhedged portfolios, as it will allow us to substantially reduce the number of backtests we have to compare. In figure 2, we compare all hedged and unhedged strategies. The only exception is the backtest where we replicate the model from Bordonado et al (2017), as he has already shown that his model performs better when it is unhedged. As the boxplot demonstrates, unhedged models yield a higher Sharpe ratio, with only 5 out of 408 backtests yielding a higher Sharpe ratio when being hedged. Since none of these 5 models has a higher Sharpe ratio than the maximum Sharpe ratio of the unhedged models and only outliers are able to generate a positive Sharpe ratio for hedged portfolios, we will not consider hedged portfolios any further and claim that leaving the portfolios unhedged generates higher Sharpe ratios, both on average and among the top performing models. This result supports the claim from Bordonado et al (2017) who contradict the finding from Simon and Campasano (2014), by stating that unhedged portfolios outperform hedged portfolios, even on a risk-adjusted basis.

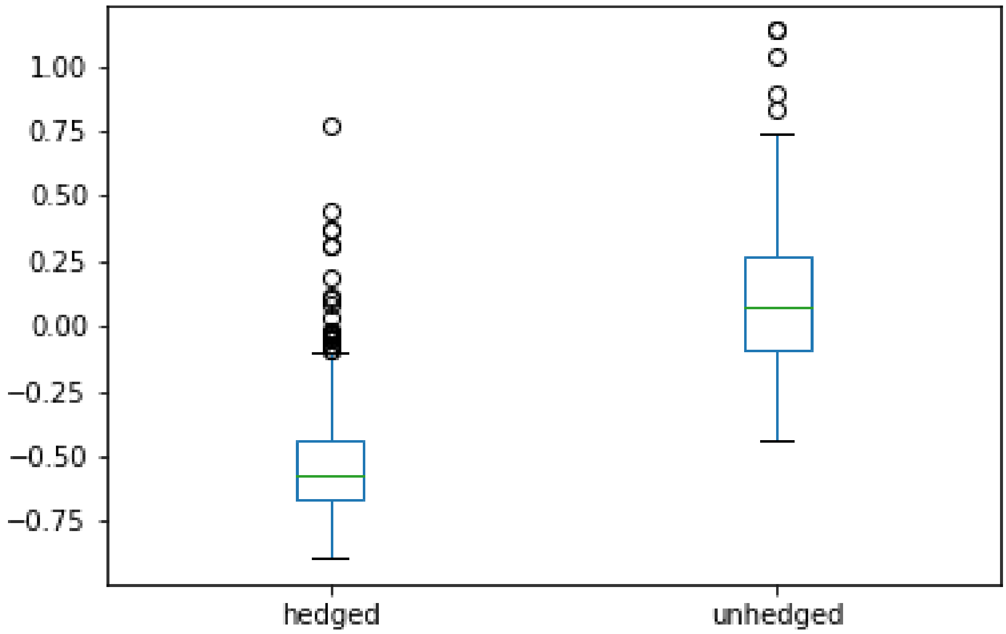

Figure 2: OOS Sharpe ratio comparison of hedged and unhedged portfolios

Another comparison that can be made across all models is that of whether there is an indicator that clearly out or underperforms the other indicators. Figure 3 shows the OOS Sharpe ratios of all models, grouped by their indicators. It becomes clear that the median values are highest for the binary indicator, followed by the scaled and then the threshold indicator. As we are however not interested in median values, but much rather in which indicator achieves the highest Sharpe ratios, the t100 indicator seems to be the most appealing one. It has the biggest range of Sharpe ratios of all indicators, and we will show in the next section that this is due to it working very well with some models, and not so well with others. As there are no clear best or worst indicators, we will proceed with all of them and since there are no more factors that all models have in common, we will now analyze the benchmark and GM/DAGM models in more detail.

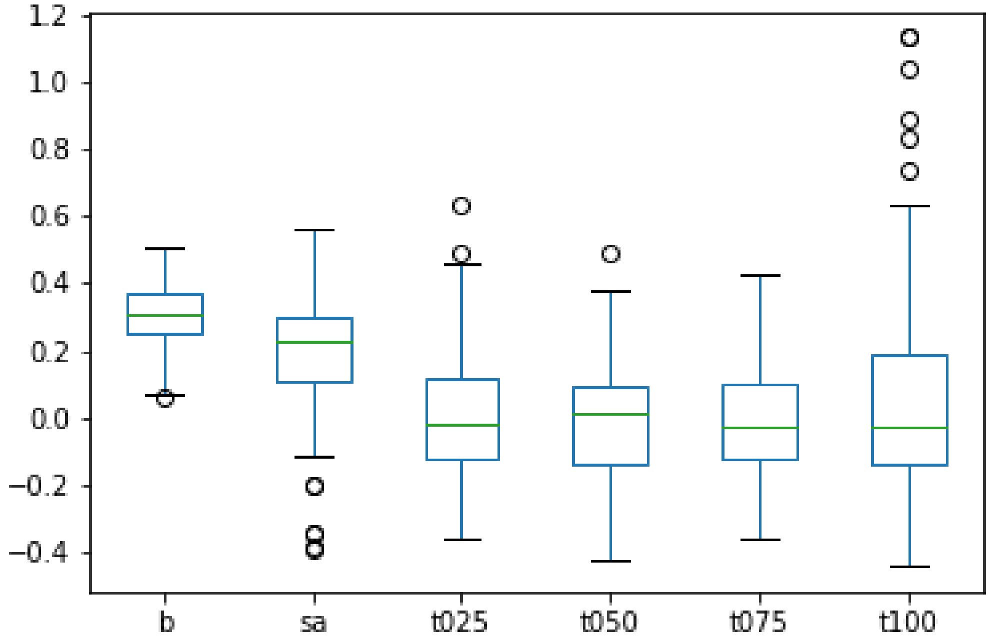

Figure 3: OOS Sharpe ratio comparison of all applied indicators

4.2 | Benchmark Model Analysis

The OOS Sharpe ratios of the benchmark models are listed in table 4 below and the IS results are reported in table 5 in the appendix, as the IS results will not be used for performance evaluation. Out of the SMA models, the specification from Cooper (2013) turns out to be the best one (SMAC), outperforming SMA and SMA2S with nearly all indicators. On average, there is hardly any difference between the performance of the GARCH (G) model and the SMAC model. The best performing benchmark RVOL-forecast derived VRP model is the G_t025 model. Table 4 demonstrates quite clearly however that across all benchmark models, the best performing model is the roll yield (RY) derived VRP model, where all except for one specification outperform the G_t025. What is noteworthy is that RY models require either a scaled or a threshold indicator to deliver a satisfying Sharpe ratio as the RY model using a pure binary indicator turns out to be the worst performing benchmark model. This also explains why Cooper (2013) concludes that RVOL-forecasting models are better suited than the RY to estimate the VRP, as he applied a binary indicator in his paper. Furthermore, it demonstrates that it depends on the VRP estimation methodology which indicator is best suited, and we can therefore underline the previous result that there is no overall best indicator. The overall best benchmark model is the RY model with the specification used by Bordonado et al (2017), which allows us to conclude that alternative optimization and indicator methods are not able to beat the ones used in their paper and is the reason why will resume with this model being the benchmark that the GM/DAGM models, which will be analyzed in the next section, have to beat.

Table 4: OOS Sharpe ratios of benchmark models

Indic.	G	RY	SMA	SMA2S	SMAC	mean
b	0.43	0.09	0.25	0.36	0.43	0.31
bord[7]	-	0.74	-	-	-	-
sa	0.36	0.56	0.18	0.30	0.15	0.31
t025	0.48	0.63	0.18	0.32	0.46	0.41
t050	0.25	0.49	0.14	0.38	0.30	0.31
t075	0.03	0.43	0.08	0.28	0.20	0.20
t100	0.28	0.40	0.07	0.29	0.36	0.28
mean	0.31	0.48	0.15	0.32	0.32	-

4.3 │GM and DAGM Model Analysis

In the next step, it must be determined which GARCH-MIDAS specification works best, as well as which low frequency variable is best suited. The Sharpe ratios of both the DAGM and

[7] Being an exact replication of the strategy from Bordonado et al (2017), this backtest only uses the RY indicator.

GM models are plotted based on their low frequency variables in figure 4 below, which allows us to draw the following conclusions: Low frequency variables observed on a weekly basis yield higher Sharpe ratios than low frequency variables observed on a monthly basis. This is demonstrated by 3MVIX, HFQ, NFCI, RY and VIX, which are weekly observations, all having higher medians, max and min values than CFNAI, HOUSING and INDPRO, which are monthly observations. This difference still holds when the models are split into DAGM and GM models, even though it is more pronounced for the DAGM models.

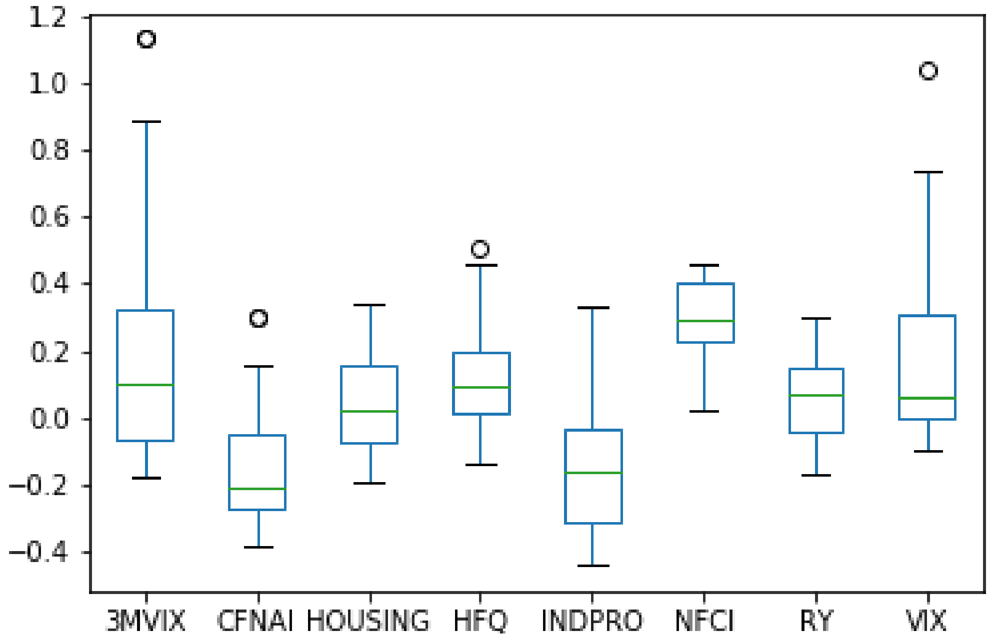

Figure 4: OOS Sharpe ratios of all applied MIDAS components

The results are presented in more detail in the tables below, with table 6 displaying the DAGM and table 7 the GM OOS Sharpe ratios and the IS Sharpe ratios being displayed in the appendix with table 8 reporting the DAGM and table 9 the GM results. Overall, there is no clear conclusion as to whether DAGM or GM models perform better on average, but when it comes to the best performing models, a comparison of table 6 and table 7 demonstrate quite clearly that DAGM models should be chosen over GM models. A comparison of DAGM and GM models is shown in figure 5 in the appendix. When it comes to the indicators, it becomes apparent that on average, the binary indicator is suited best, followed by the scaled and the threshold indicators. Just like for the benchmark model, we must point out that it is impossible to select the best indicator, as it is dependant on other factors, in this case, the MIDAS variables.

Table 6: OOS Sharpe ratios of DAGM models

weight	lag	Indic.	3MVIX	C[8]	H	HFQ	I	N	RY	VIX	mean
BETA	K4	b	0.33	0.15	0.29	0.51	0.13	0.36	0.27	0.33	0.30
		sa	0.36	-0.01	0.15	0.22	-0.39	0.46	0.27	0.35	0.18
		t025	-0.07	-0.21	-0.07	0.08	-0.36	0.34	0.07	0	-0.03
		t050	-0.18	-0.3	0.02	0.09	-0.21	0.21	0.15	0.06	-0.02
		t075	-0.07	-0.21	-0.07	0.08	-0.36	0.34	0.07	0	-0.03
		t100	1.13	-0.29	-0.18	-0.08	-0.05	0.06	-0.11	0.19	0.08
	K8	b	0.32	0.09	0.33	0.37	0.09	0.39	0.15	0.06	0.23
		sa	0.25	-0.05	0.26	0.17	-0.34	0.44	0.2	0.37	0.16
		t025	-0.12	-0.23	0.08	-0.04	-0.15	0.42	-0.04	0.04	-0.01
		t050	-0.01	-0.38	0.09	-0.07	-0.4	0.23	0.03	0.02	-0.06
		t075	-0.12	-0.23	0.08	-0.04	-0.15	0.42	-0.04	0.04	-0.01
		t100	0.89	-0.35	-0.05	-0.13	-0.44	0.02	0.06	1.04	0.13
EXP	K4	b	0.31	0.16	0.28	0.46	0.14	0.36	0.3	0.39	0.30
		sa	0.36	-0.06	0.15	0.23	-0.38	0.46	0.27	0.37	0.18
		t025	-0.06	-0.24	-0.09	0.12	-0.36	0.34	0.07	0	-0.03
		t050	-0.17	-0.32	0.01	0.08	-0.21	0.21	0.06	0.03	-0.04
		t075	-0.06	-0.24	-0.09	0.12	-0.36	0.34	0.07	0	-0.03
		t100	1.13	-0.31	-0.19	-0.08	-0.06	0.02	-0.09	0.18	0.08
	K8	b	0.33	-	0.34	0.36	0.18	0.41	0.15	0.32	0.30
		sa	0.27	-	0.26	0.18	-0.34	0.45	0.23	0.25	0.19
		t025	-0.13	-	0.06	0.01	-0.21	0.37	-0.08	-0.1	-0.01
		t050	-0.01	-	0.08	0.01	-0.42	0.2	0.09	-0.08	-0.02
		t075	-0.13	-	0.06	0.01	-0.21	0.37	-0.08	-0.1	-0.01
		t100	0.83	-	-0.04	-0.07	-0.42	0.04	0.08	0.74	0.17
	mean		0.22	-0.17	0.07	0.11	-0.22	0.30	0.09	0.19	-

As we aim to find the ideal specification, we are however much more interested in identifying the parameters of the best performing models, instead of average values, and here the results are clear. Using the highest threshold model t100 in combination with the 3MVIX variable (t100_3MVIX) yields the best results, with 8 out of the 10 best performing GM/DAGM models using the t100_3MVIX combination. The remaining 2 are t100_VIX models, which allows us

[8] The applied R optimization tool failed to yield a result for all DAGM_C_K8_EXP models

to claim that adding a measure of implied volatility does add significant value to the model. What is more, when it comes to the best performing models, DAGM models work much better than GM models as all DAGM models outperform GM models in the t100_3MVIX combination and all the best five models are DAGM models. Seeing that the 3MVIX models only perform well with a threshold should not come as a surprise either, as the term-structure of S&P 500 options implied volatility can also be interpreted as a measure of roll yield. Consequently, this is consistent with what we already saw in the benchmark models, where the RY model required a threshold indicator to achieve satisfying results.

Table 7: OOS Sharpe ratios of GM models

weight	lag	Indic.	3MVIX	C	H	HFQ	I	N	RY	VIX	mean
BETA	K4	b	0.29	0.11	0.24	0.42	0.29	0.43	0.25	0.43	0.31
		sa	0.27	-0.1	0.11	0.22	-0.2	0.42	0.27	0.3	0.16
		t025	-0.05	-0.22	-0.09	0.04	-0.17	0.23	0.09	0.15	0.00
		t050	-0.17	-0.33	-0.05	0.03	-0.14	0.23	0.10	-0.05	-0.05
		t075	-0.05	-0.22	-0.09	0.04	-0.17	0.23	0.09	0.15	0.00
		t100	0.59	-0.26	-0.13	-0.14	-0.3	0.25	-0.12	0	-0.01
	K8	b	0.3	0.3	0.33	0.39	0.33	0.4	0.15	0.31	0.31
		sa	0.21	0.03	0.18	0.19	-0.1	0.42	0.19	0.25	0.17
		t025	-0.02	-0.13	-0.04	0.1	0.04	0.23	-0.09	-0.03	0.01
		t050	-0.14	-0.14	0.05	0.02	-0.13	0.23	0.06	-0.05	-0.01
		t075	-0.02	-0.13	-0.04	0.1	0.04	0.23	-0.09	-0.03	0.01
		t100	0.61	-0.28	-0.17	-0.03	-0.17	0.25	0.04	0.01	0.03
EXP	K4	b	0.31	0.07	0.24	0.42	0.28	0.39	0.25	0.37	0.29
		sa	0.27	-0.1	0.11	0.23	-0.2	0.42	0.12	0.31	0.15
		t025	-0.05	-0.23	-0.12	0.11	-0.15	0.23	-0.17	0.15	-0.03
		t050	-0.16	-0.33	-0.05	0.01	-0.12	0.23	0.05	-0.03	-0.05
		t075	-0.05	-0.23	-0.12	0.11	-0.15	0.23	-0.17	0.15	-0.03
		t100	0.59	-0.3	-0.13	-0.14	-0.3	0.24	-0.03	0.01	-0.01
	K8	b	0.28	0.3	0.31	0.36	0.3	0.4	0.25	0.38	0.32
		sa	0.24	0.02	0.23	0.19	-0.11	0.42	0.12	0.27	0.17
		t025	-0.06	-0.17	0	0.14	0.02	0.23	-0.17	-0.09	-0.01
		t050	-0.12	-0.14	0.03	0.01	-0.15	0.23	0.05	-0.1	-0.02
		t075	-0.06	-0.17	0	0.14	0.02	0.23	-0.17	-0.09	-0.01
		t100	0.63	-0.33	-0.12	-0.05	-0.18	0.24	-0.03	0.06	0.03
	mean		0.15	-0.12	0.03	0.12	-0.06	0.29	0.04	0.12	-

As a result, we can conclude that 3MVIX is the best suited MIDAS variable, with the t100 indicator being the best performing indicator across all 3MVIX models, all t100_3MVIX performing better when using DAGM models and a shorter lag beating a longer lag across all DAGM_t100_3MVIX models.

4.4 │Best Models Overview

Table 10 provides an overview of the best performing models of each category of benchmark models, as well as the best GM/DAGM model. As table 10 clearly tells, the DAGM_t100_3MVIX_K4_EXP, the best performing GM/DAGM model, outperforms all benchmark models based on all risk-adjusted performance metrics. In addition, the DAGM_t100_3MVIX_K4_EXP model, just like RY_bord, has a negative correlation to the S&P 500, which makes it an interesting strategy for portfolio diversification. Another interesting measure is that of skewness of daily returns, which is positive for the DAGM_t100_3MVIX_K4_EXP and RY_bord models which again separates them from the other two models and equity markets in general, which tend to be negatively skewed. The fact that the biggest one-day movements are upwards and not downwards is also supported by the Sortino ratio of 2.32, which is drastically higher than the Sharpe ratio of 1.13 for the DAGM_t100_3MVIX_K4_EXP model. Moreover, both the skewness and the negative SPX correlation demonstrate that the performance is not generated by predominantly shorting volatility and thus collecting the equity risk premium.

What is more, it is notable that all models are threshold models, which again supports the statement made earlier that using a threshold indicator is better suited to generate top-performing VIX trading models than pure binary or scaled indicators. How high the threshold should be is highly dependent on how the VRP is estimated, as those model that take any form of term structure into account (DAGM_t100_3MVIX_K4_EXP, RY_bord) require a higher threshold, whereas those that don’t (G_t025, SMAC_t025), use a threshold that makes them close to the binary indicator and hence only exclude VRP estimates that are close to 0. Having had an extensive discussion on whether the term-structure of VIX futures or a forecast of RVOL is better suited to estimate the VRP, it is interesting to see that the best performing backtest makes use of an RVOL forecasting model that also considers the term structure, be it the term structure of implied volatility on S&P 500 options. We therefore make the claim that DAGM models are better suited to estimate the VRP for a volatility trading model than the term-structure estimated VRP, but highlight the finding that the best performing DAGM model still incorporates some form of term-structure.

Table 10: OOS Sharpe ratios of each categories best model

	DAGM_t100_3MVIX_K4_EXP	RY_bord	G_t025	SMAC_t025
CAGR	10.95%	10.97%	7.93%	7.64%
Sharpe ratio	1.13	0.74	0.48	0.46
Sortino ratio	2.32	1.67	0.62	0.60
Calmar ratio	0.98	0.60	0.27	0.24
Max. drawdown	-11.19%	-18.27%	-29.23%	-31.27%
Skewness	4.74	12.36	-5.49	-5.42
Kurtosis	65.44	296.38	98.43	97.60
SPX correlation	-0.21	-0.29	0.23	0.21

4.5 │Looking Beyond the Numbers

What becomes apparent when analyzing the results is that the performance measures of the strategies that were already published in previous papers have worsened substantially. While the CAGR will naturally be lower as we only use 25% of our portfolio, this does not have an impact on risk-adjusted performance measures, such as the Sharpe ratio. Bordonado et al (2017) reported a Sharpe ratio of 2.11, which has dropped to 0.74 in our OOS period and Cooper's (2013) Sharpe ratio of 1.85 has dropped to 0.43 (SMAC_b model). What is more, not even our best performing model with a Sharpe ratio of 1.13 came anywhere near the 2.11 reported by Bordonado et al (2017), despite its substantial outperformance in our OOS period. A potential reason for that is the selected evaluation period: Our results are evaluated based on their performance in the OOS period, which starts in 2016. The Sharpe ratios in the previously mentioned papers were evaluated for a period that ends in mid-2013 (Cooper 2013; Bordonado et al 2017), which either hints at the performance of the model being heavily determined by the market “regime”, or the potential gains from these strategies vanishing over time.

In order to analyze these hypotheses, we will analyze how the Sharpe ratios of all models have changed from the IS to the OOS period. Even though our IS periods lasts till the end of 2015, which is longer than mid-2013 in the previously mentioned papers, we are convinced that comparing the results of that period will yield valuable insights.

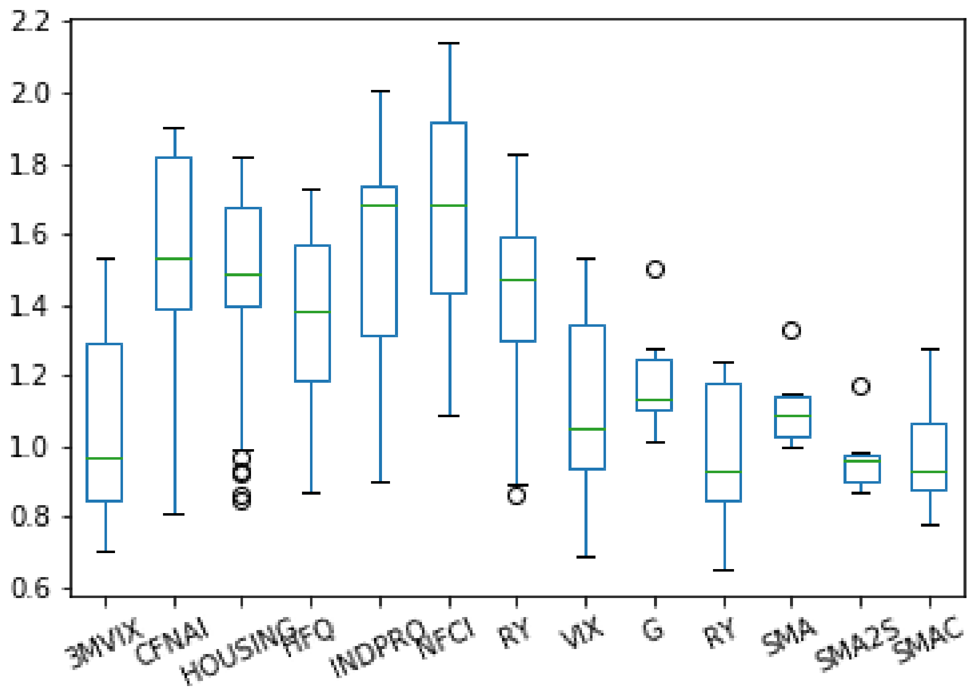

Figure 6: IS Sharpe ratios of all Benchmark, GM and DAGM models

Figure 6 shows the Sharpe ratios of the benchmark models and all GM/DAGM models in the IS period and it is remarkable to see that there is not a single model with a Sharpe ratio below 0.6 and the median of all but one GM/DAGM models is above 1.0. Even more importantly, all GM/DAGM models achieve max values of more than 1.5, with several models yielding a Sharpe ratio of above 2.0 and the top GM model generating a Sharpe ratio of 2.12, which slightly surpasses the Sharpe ratio of Bordonado et al (2017), despite our IS period being more than a year longer than the one used in their paper. While this finding underlines that strategies with an impressively high Sharpe ratio can indeed be generated when applied to this particular time frame, this result further hints at these strategies being either dependent on a certain market regime, or the returns vanishing over time. While one could argue that the models were simply overspecified and hence yielded such high Sharpe ratios, a remark should be made. The models are not optimized to obtain the maximum Sharpe ratio but the highest prediction accuracy in the IS period, and perhaps even more importantly, not all models were optimized in the first place. Of the benchmark models, none of the RY models are dependent on any optimization process and all SMAC models took the specification as published by Cooper (2013).

In order to analyze this issue in more detail, the performance of all backtests is split into four regimes and displayed in table 11 and the figure below. While table 11 contains data of all backtests that were conducted, figure 7, for illustrative puproses only contains the equity curve of all binary models. The performance of the remaining indicators are displayed in figure 8, figure 9, figure 10, figure 11 and figure 12 in the appendix, in order to demonstrate that the same behaviour can be observed across all indicators.

Table 11: Sharpe ratios across all models split by regime

	Regime 1	Regime 2	Regime 3	Regime 4
min	0.97	-0.51	-0.98	-1.07
max	2.45	0.96	0.78	2.35
median	1.78	0.40	-0.60	1.28
mean	1.78	0.36	-0.55	1.21
positive	409	351	14	379
negative	0	58	395	30
% positive	100.00%	85.82%	3.42%	92.67%
% negative	0.00%	14.18%	96.58%	7.33%

The first regime lasts till 01.01.2013 and is one, where not a single backtest yields a negative return and the average Sharpe ratio is 1.78. Even more impressively, the minimum Sharpe ratio in that regime is 0.97 which is more than twice the average Sharpe ratio of the S&P 500 of

0.44[9]. What is more, out of 409 backtests, 66 backtests yield a higher Sharpe ratio than Bordonado et al (2017) and 179 a higher Sharpe ratio than Cooper (2013). To us this further demonstrates that in that period, a wide range of models was able to achieve outstanding results. It has to be noted, however, that most VIX ETNs were not available for trading across this whole period and the trading volume of existing VIX ETNs was much smaller than nowadays, requiring further evidence to answer the question of whether the return of such strategies vanishes over time.

The second regime lasts from 01.01.2013 – 01.01.2018 and returns in this period are highly dependent on the model and indicator that is used. As the table shows, 85.82% of all models still deliver a positive performance, but the average Sharpe ratio has dropped significantly from 1.78 to 0.36. As the models are only assessed based on the OOS period and have already been discussed earlier, we will not go into more detail here, but the figures show that there are still many model indicator combinations that are able to generate substantial gains. This is no longer true for all models however and hence it tells us that the choice of model and indicator has a big impact on the performance of the backtest in that period and performance is not only determined by the regime.

The third regime lasts from 01.01.2018 – 01.03.2020 and is a period where 96.58% of all models yield a negative return and the best performing model has a lower Sharpe ratio than the worst performing model in regime 1. There are two reasons for that, with the first being that 05.02.2018 was a defining day for volatility trading, with the VIX posting its largest one-day movement in the entire backtesting timeframe. This event and its impact on the market requires deeper analysis and will therefore be addressed later in more detail. The second reason is that

[9] The average S&P 500 Sharpe ratio is measured over the backtesting timeframe 01.01.2008 – 27.11.2020

what followed was a market that was going sideways for the remainder of 2018, which is an environment that generally seems to be challenging for all VRP strategies. Nevertheless, some models still yielded Sharpe ratios that were higher than the S&P 500 average, demonstrating again that the right model can perform well even in the most challenging environment.

The fourth regime lasts from 01.03.2020 – 27.11.2020 and contains the Corona crisis with its following recovery rally. The behaviour of most strategies resembles that of regime 1, with 92.67% of strategies yielding a positive Sharpe ratio, and the average Sharpe ratio being 1.21. This period demonstrates clearly that the returns of VRP strategies did not vanish over time but are heavily driven by the market regime.

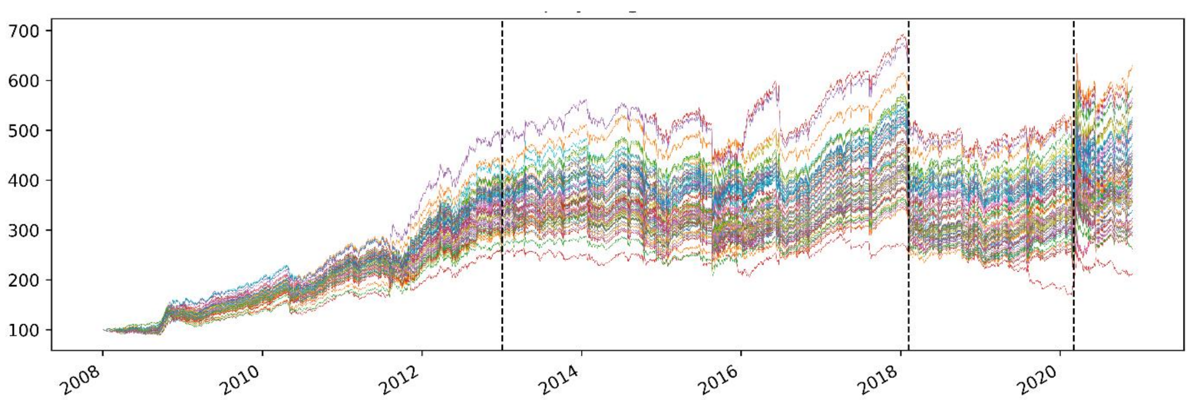

Figure 7: Performance comparison of all binary models, split by regimes

Does this allow us to conclude that the only factor that matters is the market regime? While regimes one and four certainly make it seem like the choice of model is almost irrelevant, it is regimes three and four, which demonstrate that the choice of the correct model and indicator are becoming very important when there is a challenging market regime and can make the difference between losing a substantial amount of previously made gains, and collecting profits, even in a challenging market environment. This is also demonstrated by our best performing OOS strategy DAGM_t100_3MVIX_K4_EXP boasting positive Sharpe ratios across all four regimes. We, therefore, claim that market regimes that suit VRP strategies, in general, will lead to spectacular returns with almost all VRP estimation methodologies, but the choice of the correct model is of substantial importance in regimes that suit the strategy not so well.

Now the question arises what makes a good regime for a VRP trading strategy? The two best regimes are regime one and four, and what they both have in common is a severe market crash, followed by a lasting recovery rally. The worst regime for VRP strategies was regime three, which had a massive one-day crash, where the panic vanished almost as quickly as it came, which was followed by a sideways moving market. This is again something that regime three shares with regime two, which contains the Greek debt crisis, as well as the sideways market of 2015, which both proved to be challenging times for VRP strategies. Taking these points into consideration, we can claim that trending markets, in either direction, are regimes in which VRP strategies thrive, whereas sideways markets tend to be a challenging environment.

As previously mentioned, another important factor that must be mentioned is that there was one day in the OOS period, which had a massive impact on the performance of the model. On February 5th 2018, the VIX spiked 115%, and while the traded instrument only spiked 36%, getting this day right could make the difference and either ruin or substantially improve the performance over the OOS period. Out of all the models in our backtest, the only model that was long volatility on the day of the spike was the model from Bordonado et al (2017) (RY_bord), which also explains why it outperforms the other benchmark models in the OOS period. Among the GM/DAGM models, several threshold models did not trade on that day and consequently avoided a massive drop in their portfolio value. Unsurprisingly, one of these models is also our best performing model DAGM_t100_3MVIX_K4_EXP which did not trade on the day where the VIX spiked, correctly anticipated the drop in the VIX the following day and entered into a short volatility position due to the substantial VRP passing the threshold. An interesting aspect in that context is that despite this day being the seemingly perfect case for using the hedged portfolios, the hedged portfolios still underperformed the unhedged portfolios.

The reason for that is that the relationship between the VXX and SPX, which is assumed to be stable in general, but that was not the case on that day and the VXX spiked beyond the 2.25 multiplier assumed earlier. To illustrate how severe the impact of that one day was on the performance of the backtests, figure 13 compares the performance of the model that performed best in the IS period, with the best performing model in the OOS period.

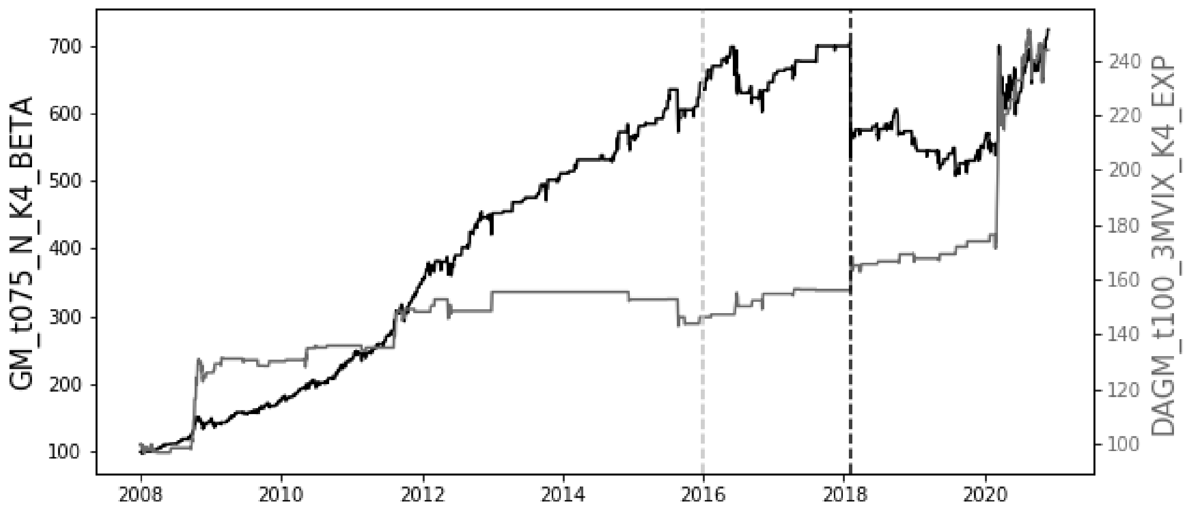

Figure 13: Performance comparison of the best OOS (DAGM_t100_3MVIX_K4_EXP) and IS (GM_t075_N_K4_BETA) model

The vertical dashed black line marks this respective day and the effect of being right or not trading at all on this day are illustrated. The GM_t075_N_K4_BETA model drops massively, and even though it recovers on the next day, it is far from gaining back its initial gains. This results not only in it being underwater for more than a year but also in the model hardly gaining a positive CAGR after the beginning of the OOS period, which is indicated by the dashed vertical silver line. The DAGM_t100_3MVIX_K4_EXP model on the other hand restrains from trading on that day, capitalizes on the drop in the VIX on a subsequent day and returns a significantly positive return in the OOS period. This however does not imply that we consider the DAGM_t100_3MVIX_K4_EXP model to be the best one out of pure luck, as the chart clearly demonstrates that it performs very well across the whole period, but it is nevertheless an important remark to make. The extent to which this day should be considered is of course debatable. On the one hand, it can be argued that such a spike in the VIX is one of the risks in any VIX trading strategy and short sellers are compensated for that by the VRP. On the other hand, it can be argued that estimating the VRP before that day in such a way that it either does not surpass the threshold or is negative, is a lucky punch as there is no other comparable day in the OOS period, making it difficult to tell whether this prediction can be replicated. As our best performing model did not trade on that day, and yet outperformed the only model that had a long volatility position going into the day, we will restrain from analyzing the OOS results where this day is excluded, but we still consider it an important point that had to be mentioned.

5 | Conclusion

The results of our backtests allow us to state that out of all estimation methodologies using an RVOL forecasting model to estimate the VRP, DAGM models that take the term-structure of SPX options into account are best suited to estimate the VRP for a volatility trading strategy. These DAGM models were also able to outperform models that estimate the VRP based on the term structure of VIX futures. Due to their positive skewness and negative correlation, we also were able to show that their performance does not stem from a predominantly short volatility position and that these strategies can act as a great tool for portfolio diversification. Furthermore, we showed that the performance of volatility trading strategies is highly dependent on the market regime and can therefore vary significantly, depending on the period in which they are evaluated. What is more, we demonstrate that the best Sharpe ratio can be achieved by using a threshold that the VRP has to surpass, as this means that the strategy only enters into a trade, when a significant amount of VRP can be earned. In addition, our results indicate clearly that leaving the models unhedged yields better risk-adjusted results. Unhedged models constantly perform better than models that are hedged with the S&P 500, as the relationship between VIX futures and the SPX does not always hold in extreme market downturns.

Reference List

Alexander, C., Korovilas, D., (2013). Volatility Exchange-traded Notes: Curse or Cure?. *Journal of Alternative Investments,* 15(2), 52-70

Alexander, C., Kapraun, J., & Korovilas, D. (2015). Trading and investing in volatility products. *Financial Markets, Institutions & Instruments*, 24(4), 313-347.

Amendola, A., Candila, V., & Gallo, G. M. (2019). On the asymmetric impact of macro–variables on volatility. *Economic Modelling*, *76*, 135-152.

Asgharian, H., Hou, A. J., & Javed, F. (2013). The importance of the macroeconomic variables in forecasting stock return variance: A GARCH-MIDAS approach. *Journal of Forecasting*, *32*(7), 600-612.

Barclays (2019). iPath Series B S&P 500 VIX Short-Term Futures ETN Prospectus. Retrieved from: https://www.ipathetn.com/US/16/en/documentation.app?instrumentId=341408&documentId=6906743

Bittman, J. (2007). VIX, it protects and diversifies. *Futures*, 36(9), 38.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, *81*(3), 637-654.

Bollen N., O'Neill M.J., Whaley R. (2017). Tail Wags Dog: Intraday Price Discovery in VIX Markets. *Journal of Futures Markets*, 37(5), 431–451.

Bollerslev, T. (1986). Glossary to ARCH (GARCH). Volatility and Time Series Econometrics Essays in Honor of Robert Engle. MarkWatson, Tim Bollerslev and Je¤ rey.

Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. The Review of Financial Studies, 22(11), 4463-4492.

Bordonado, C., Molnár, P., & Samdal, S. R. (2017). VIX exchange traded products: Price discovery, hedging, and trading strategy. Journal of Futures Markets, 37(2), 164-183.

CBOE (2020). VIX White Paper. CBOE. Retrieved from: https://www.cboe.com/vix

Conrad, C., Custovic, A., & Ghysels, E. (2018). Long-and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. *Journal of Risk and Financial Management*, *11*(2), 23.

Conrad, C. and Kleen, O., (2020). Two are better than one: Volatility forecasting using multiplicative component GARCH-MIDAS models. *Journal of Applied Econometrics,* 35(1), 19-45.

Cooper, T. (2013). Easy Volatility Investing. *Available at SSRN 2255327*.

Dondoni, A., Montagna, D., & Maggi, M. (2018). VIX Index Strategies: Shorting volatility as a portfolio enhancing strategy. Available at SSRN 3104407.

Dufour, J. M., Garcia, R., & Taamouti, A. (2012). Measuring high-frequency causality between returns, realized volatility, and implied volatility. Journal of Financial Econometrics, 10(1), 124-163.

Edwards, T., & Preston, H. (2017). A practitioners guide to reading VIX R. S&P Global: S&P Dow Jones Indices (Education, Strategy 201), 1-10.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987-1007.

Engle, R. F., & Patton, A. J. (2007). What good is a volatility model?. In Forecasting volatility in the financial markets (pp. 47-63). Butterworth-Heinemann.

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

Eraker, B. (2008). The volatility premium. Manuscript.

Fama, E. F. (1965). The behavior of stock-market prices. The journal of Business, 38(1), 34-105.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. Journal of Econometrics, 131(1-2), 59-95.

Heber, Gerd, Asger Lunde, Neil Shephard and Kevin Sheppard (2009) "Oxford-Man Institute's realized library" Library Version: 0.3, Oxford-Man Institute, University of Oxford

Hull, J. C. (2019). Options, futures and other derivatives.

Ilmanen, A. (2011). Expected returns: An investor's guide to harvesting market rewards. John Wiley & Sons.

Mandelbrot, B. (1972). Certain speculative prices (1963). The Journal of Business, 45(4), 542-543.

Marra, S. (2015). Predicting Volatility. Lazard Asset Management. Retrieved from: https://www.lazardassetmanagement.com/docs/-m0-/22430/predictingvolatility_lazardresearch_en.pdf

Mixon, S. (2007). The implied volatility term structure of stock index options. Journal of Empirical Finance, 14(3), 333-354.

Nossman, M., & Wilhelmsson, A. (2009). Is the VIX futures market able to predict the VIX index? A test of the expectation hypothesis. The Journal of Alternative Investments, 12(2), 54-67.

McFarren, T. (2013). VIX Your Portfolio. *BlackRock Investment Insight*.

Ribeiro, R., Kolanovic, D., Silvestrini, T. Lee (2012). Risk premia in volatility markets: exploiting volatility spillover and clustering. *Technical report, J.P. Morgan*.

Simon, D. P., & Campasano, J. (2014). The VIX futures basis: Evidence and trading strategies. *The Journal of Derivatives*, *21*(3), 54-69.

Stanton, C. W. (2011). Volatility as an Asset Class: The Potential of VIX as a Hedging Tool and the Shortcomings of VIX Exchange Traded Notes. *Journal of Investment Consulting*, 12(1), 23-30.

Whaley, R. E. (2009). Understanding the VIX. *The Journal of Portfolio Management*, *35*(3), 98-105.

Whaley, R., (2013). Trading Volatility:At What Cost? *The Journal of Portfolio Management*, 40(1), 95-108.

Appendix

German Abstract

Das Ziel dieses Papers ist es, herauszufinden, ob GARCH-MIDAS Modelle besser geeignet sind um die S&P 500 Volatilitäts Risiko Prämie (VRP) für eine VIX Trading Strategie zu schätzen, als herkömmliche, sowie VIX Futures Terminkurven implizierte VRP Modelle. Um das beste Model zu identifizieren wird die berechnete VRP als Indikator für eine long/short VIX Trading Strategie genutzt und die Modelle werden basierend auf deren Out-of-Sample Sharpe Ratios evaluiert. Wir werden nicht nur einen Binären, sondern auch einen Threshold- und einen Scaled-Indikator anwenden, um die Größe der VRP zu berücksichtigen. Darüberhinaus werden wir GARCH-MIDAS zu Double Asymmetric GARCH-MIDAS (DAGM) Modellen erweitern, um die Asymmetrie von Volatilität zu berücksichtigen. Die empirischen Resultate zeigen, dass DAGM Modelle, die die Terminkurven Struktur von SPX Optionen als Langzeit Komponente nutzen, eine bessere Sharpe Ratio als konventionelle und VIX Terminkurven Struktur implizierte VRP Modelle erzielen. Darüberhinaus zeigen wir, dass die Performance von VIX Trading Strategien stark vom jeweiligen Markt Regime beeinflusst wird.

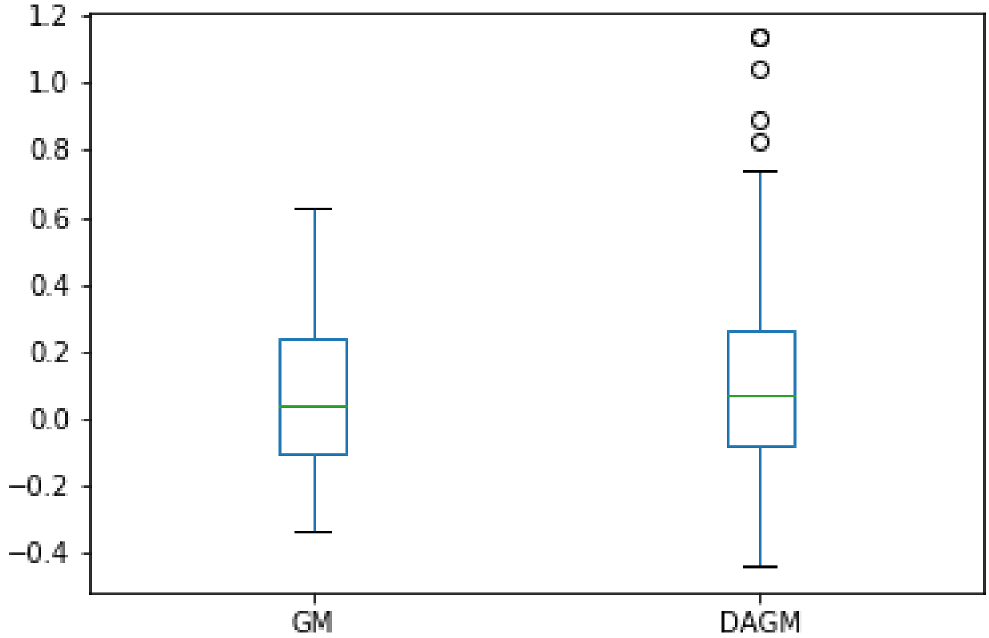

Figure 5: OOS Sharpe ratio comparison of all DAGM and GM models

Table 5: Benchmark models IS Sharpe ratios

	G	RY	SMA	SMA2S	SMAC	mean
b	1.01	0.65	1.06	0.97	0.87	0.91
bord		1.2				1.20
sa	1.50	0.79	1.33	1.17	1.28	1.21
t025	1.15	0.93	1.02	0.87	0.97	0.99
t050	1.28	1.24	1.11	0.95	0.78	1.07
t075	1.12	1.15	1	0.88	0.89	1.01
t100	1.1	0.9	1.15	0.98	1.1	1.05
mean	1.19	0.98	1.11	0.97	0.98	-

Table 8: DAGM models IS Sharpe ratios

weight	lag	Indic.	3MVIX	C	H	HFQ	I	N	RY	VIX	mean
BETA	K4	b	1.23	0.84	0.97	0.89	1.15	1.13	0.91	1	1.02
		sa	1.29	1.48	1.45	1.37	1.73	1.63	1.47	1.35	1.47
		t025	0.9	1.82	1.62	1.62	1.69	1.9	1.59	0.98	1.52
		t050	1.3	1.53	1.4	1.33	2.01	1.87	1.29	1.31	1.51
		t075	0.9	1.82	1.62	1.62	1.69	1.9	1.59	0.98	1.52
		t100	0.77	1.58	1.43	1.27	0.96	1.22	1.35	0.75	1.17
	K8	b	1.05	0.89	1.07	1.1	1.05	1.24	0.9	1.08	1.05
		sa	1.32	1.48	1.47	1.51	1.74	1.64	1.52	1.43	1.51
		t025	0.94	1.76	1.77	1.7	1.86	1.96	1.61	0.97	1.57
		t050	1.39	1.4	1.36	1.47	1.66	1.91	1.38	1.53	1.51
		t075	0.94	1.76	1.77	1.7	1.86	1.96	1.61	0.97	1.57
		t100	0.86	1.53	1.61	1.1	1.26	1.14	1.3	1.19	1.25
EXP	K4	b	1.22	0.85	0.99	0.92	1.1	1.17	0.93	1	1.02
		sa	1.29	1.49	1.45	1.37	1.73	1.63	1.47	1.35	1.47
		t025	0.88	1.86	1.63	1.58	1.71	1.88	1.52	1.02	1.51
		t050	1.27	1.43	1.38	1.31	1.99	1.87	1.29	1.3	1.48
		t075	0.88	1.86	1.63	1.58	1.71	1.88	1.52	1.02	1.51
		t100	0.79	1.69	1.49	1.23	0.95	1.22	1.33	0.74	1.18
	K8	b	1.13	-	1.09	0.93	1.11	1.18	0.89	0.83	1.02
		sa	1.32	-	1.47	1.47	1.74	1.64	1.51	1.35	1.50
		t025	1	-	1.76	1.73	1.9	1.94	1.61	0.94	1.55
		t050	1.4	-	1.33	1.28	1.66	1.9	1.38	1.38	1.48
		t075	1	-	1.76	1.73	1.9	1.94	1.61	0.94	1.55
		t100	0.82	-	1.62	1.19	1.29	1.13	1.29	0.85	1.17
		mean	1.08	1.50	1.46	1.38	1.56	1.62	1.37	1.09	-

Table 9: GM models IS Sharpe ratios

weight	lag	Indic.	3MVIX	C	H	HFQ	I	N	RY	VIX	mean
BETA	K4	b	0.81	0.81	0.86	0.87	0.92	1.12	1.03	0.9	0.92
		sa	1.24	1.5	1.48	1.38	1.68	1.65	1.48	1.34	1.47
		t025	0.78	1.89	1.77	1.65	1.76	2.14	1.54	1.14	1.58
		t050	1.43	1.39	1.46	1.41	1.62	1.71	1.32	1.51	1.48
		t075	0.78	1.89	1.77	1.65	1.76	2.14	1.54	1.14	1.58
		t100	0.82	1.69	1.65	1.2	1.35	1.59	1.3	0.76	1.30
	K8	b	1.05	0.88	0.92	0.98	1	1.09	0.86	0.89	0.96
		sa	1.32	1.47	1.48	1.42	1.71	1.66	1.53	1.36	1.49
		t025	0.9	1.84	1.81	1.55	1.73	2.09	1.62	1.02	1.57
		t050	1.44	1.39	1.48	1.3	1.54	1.75	1.4	1.51	1.48
		t075	0.9	1.84	1.81	1.55	1.73	2.09	1.62	1.02	1.57
		t100	0.7	1.56	1.55	1.16	1.61	1.5	1.33	0.73	1.27
EXP	K4	b	0.76	0.83	0.85	0.89	0.9	1.09	1.17	0.92	0.93
		sa	1.24	1.5	1.48	1.38	1.68	1.65	1.57	1.34	1.48
		t025	0.8	1.9	1.79	1.62	1.76	2.12	1.83	1.15	1.62
		t050	1.43	1.37	1.43	1.4	1.61	1.71	1.36	1.52	1.48
		t075	0.8	1.9	1.79	1.62	1.76	2.12	1.83	1.15	1.62
		t100	0.85	1.64	1.63	1.18	1.32	1.57	1.64	0.76	1.32
	K8	b	1	0.92	0.93	0.92	1	1.09	1.17	0.92	0.99
		sa	1.31	1.47	1.5	1.4	1.71	1.65	1.57	1.36	1.50
		t025	0.85	1.85	1.82	1.57	1.74	2.1	1.82	1.15	1.61
		t050	1.53	1.36	1.4	1.34	1.55	1.75	1.36	1.52	1.48
		t075	0.85	1.85	1.82	1.57	1.74	2.1	1.82	1.15	1.61
		t100	0.84	1.54	1.52	1.17	1.61	1.56	1.6	0.69	1.32
		mean	1.02	1.51	1.50	1.34	1.53	1.71	1.47	1.12	-

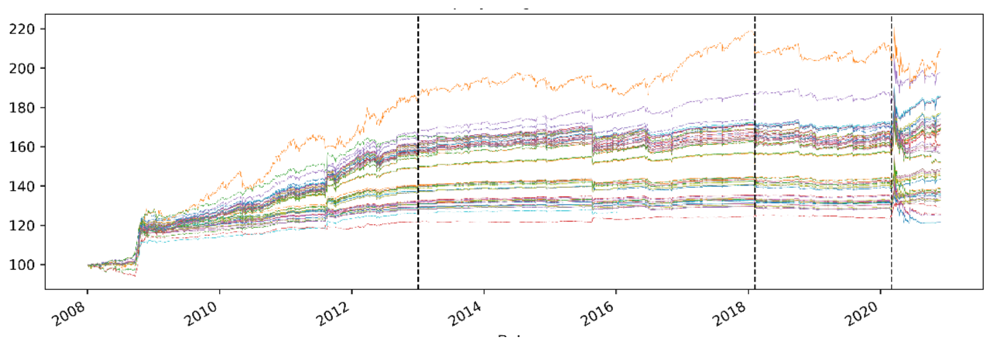

Figure 8: Performance comparison of all scaled models, split by regimes

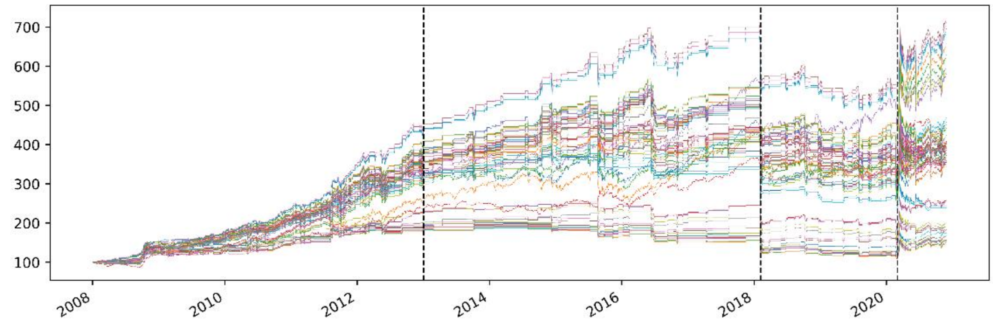

Figure 9: Performance comparison of all t025 models, split by regimes

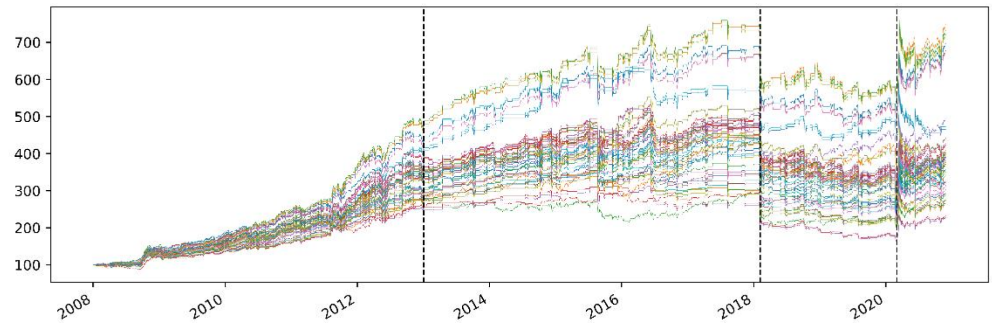

Figure 10: Performance comparison of all t050 models, split by regimes

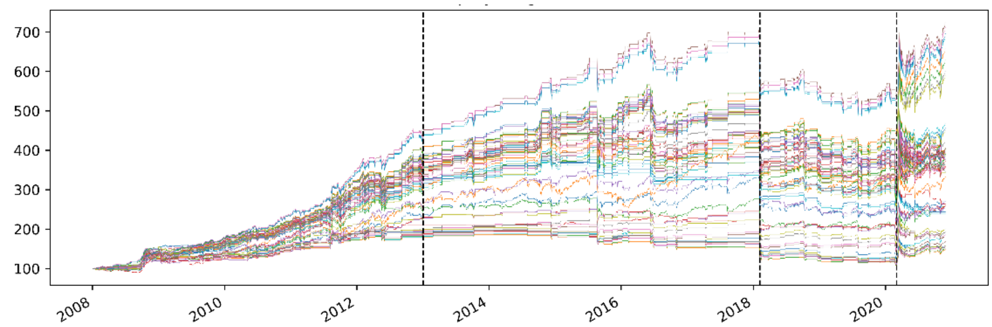

Figure 11: Performance comparison of all t075 models, split by regimes

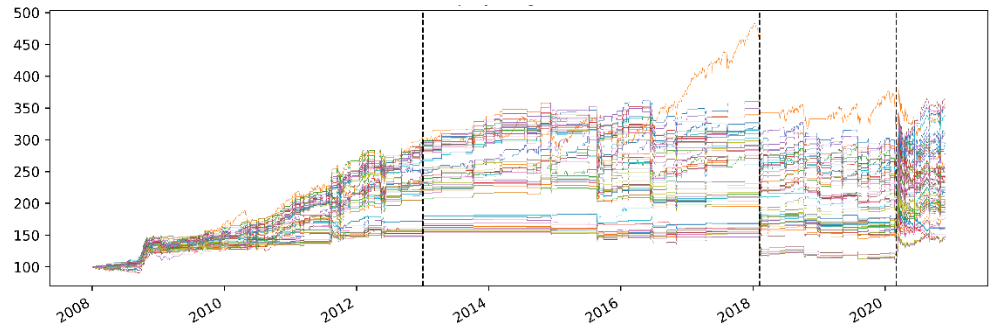

Figure 12: Performance comparison of all t100 models, split by regimes