The variance risk premium: evidence from international stock markets using different variance forecasting techniques

Master Thesis

Department of Finance

Stockholm School of Economics

Supervisor: Prof. Tobias Sichert

Submitted by: Luca Mauritz (42394)

Abstract

This thesis investigates the predictive power of the variance risk premium for a sample of 16 international stock market indices. The variance risk premium is defined as the difference between implied volatility and forecasted realized variances. To compute the variance risk premium, I test and compare a range of different variance forecasting models and select the best-performing model for each equity market. The results reveal that the variance risk premium significantly predicts stock market returns for the US and several European markets, especially at the short- and medium-term horizon. The findings for other markets are more heterogeneous. I also report evidence that the use of monthly returns, a common choice in the literature, is sensitive to the specific day of the month that is used to compute returns.

Contents

1 Introduction **1**

2 Literature review **2**

3 Data **5**

4 Realized variance forecasting **5**

4.1 Heterogeneous Autoregressive Model 8

4.2 Continuous HAR 10

4.3 Weighted-Least-Squares HAR 10

4.4 Robust Regression HAR Model 11

4.5 Random Forest 13

4.6 Computation of the variance risk premium 16

5 Stock market premium **19**

6 Robustness tests **24**

6.1 Beginning of the month data 25

6.2 Middle of the month data 26

6.3 Daily Data 28

6.4 Number of Newey-West lags 28

6.5 Estimation with WLS 31

7 Conclusion **33**

References **36**

A Appendix **38**

A.1 Monthly realized variance forecast results 38

A.2 Historical evolution of VRP 39

A.3 Detailed regression results 47

A.4 Robustness to Newey-West lags 56

1 Introduction

The question of whether and how future stock returns can be predicted has been at the heart of the financial literature for decades. Over this time, a broad range of different risk factors, carrying different risk-premia, have been proposed. One well-known risk factor is the Variance Risk Premium (VRP), which is commonly seen as a measure of risk-aversion in the marketplace (Bekaert and Hoerova, 2014). It tends to be particularly pronounced during times of high economic uncertainty, such as the global financial crisis or the outbreak of the Covid-19 pandemic. Following Zhou (2018), the VRP at time t can be defined as the difference between the risk-neutral expectation of stock market return variance and the statistical, or *physical* expectation of this variance for the next period:

$$VRP_t = E_t^Q[Var_{t,t+1}] - E_t^P[Var_{t,t+1}]. \tag{1}$$

The risk-neutral expectation of stock market return variance $E_t^Q[Var_{t,t+1}]$ is commonly measured via implied volatility extracted from option prices. The statistical expectation of asset price returns is usually defined using realized variance but has to be estimated from the data. Hence, one can compute the VRP via the following equation:

$$VRP_t = IV_{t+1} - E[RV_{t+1}], \tag{2}$$

where IV_{t+1} is the implied variance for the next month, extracted from option prices. RV_{t+1} is the realized variance during the next month and has to be estimated from past data. Hence, as the first term of the VRP can be observed in the market, the computation of the VRP depends on forecasting variance precisely. In this thesis, I estimate a number of state-of-the-art variance forecasting models and compare their performance across 16 international equity markets. The best-performing forecasting technique for each market is then used to compute an expectation of realized variance, which, in turn, can be used to compute a forward-looking measure of the VRP. I will then investigate whether the VRP can predict future stock returns for the respective stock market and whether the VRP carries a significant risk premium.

My findings suggest that the VRP works well for all four US stock markets in my sample and for a number of European markets. In general, I find larger effects in the short- and

medium-term, but several short-term prediction coefficients are not significantly different from zero. For long-term predictions, on the other side, I report smaller coefficients, but also smaller standard errors, leading to higher significance levels. The predictive power of the VRP, however, does not hold in all markets. More specifically, there are several non-US and non-European markets, where the VRP cannot predict future returns for any time horizon. Moreover, in contrast to existing research by Londono (2015), the prediction power did not improve when using the VRP based on the US stock market to predict local stock returns in other markets. Finally, I conduct a number of robustness tests, which reveal that the results are relatively sensitive to the construction of the sample and change notably depending on what day of the month is used to compute the monthly returns. Using daily, instead of monthly data as most of the existing literature does, may be seen as a valid alternative in order to make the estimation results more robust. I can also show that using weighted least squares (WLS) instead of ordinarily least squares (OLS), as in Johnson (2019), does not materially change the conclusions of my analysis.

The thesis proceeds in the following way: Section 2 summarizes the existing literature on the VRP and Section 3 introduces the different data types used across the 16 stock markets of my sample. Section 4 defines the different variance forecasting techniques and displays the best-performing forecasting models for each stock index. Section 5 presents the main analysis of this thesis by regressing future returns on the VRP for all 16 markets. In Section 6, I conduct robustness tests regarding the sample, the standard errors, and the estimation method used for the regression analysis. Finally, Section 7 concludes and provides some directions for future research.

2 Literature review

For decades, whether and how stock market returns can be predicted has been central to financial literature. Researchers have proposed a large number of factors and trading strategies. In 2009, Bollerslev et al. (2009) published an influential study that introduced a new variable to forecast stock returns: the variance risk premium (VRP). The authors defined the VRP as the difference between implied volatility, deducted from option prices, such as the VIX for the S&P-500, and a *physical variance* composed by realized stock market variance. Bollerslev et al. (2009) find that the VRP outperforms other well-known

predictors in forecasting stock market returns for certain time horizons. Subsequent research has confirmed the importance of the VRP. Drechsler and Yaron (2011), for example, show that the VRP varies significantly over time and is a powerful predictor of excess stock returns for the next one to three months. Similarly, Bekaert and Hoerova (2014) find that the VRP significantly predicts future returns on the S&P-500. The VRP has also been found to significantly predict dividend growth (Bollerslev et al., 2015). Furthermore, Bali and Zhou (2016) report that stocks which are highly correlated with the VRP carry a higher stock market risk premium, compared to stocks that are uncorrelated with the VRP.

Interestingly, the predictive power of the VRP can also be found for other asset classes, underlining its economic significance. Wang et al. (2013), for example, find that the firm-level VRP can predict spreads on credit default swaps on individual companies. They find that the VRP is the most important macroeconomic predictor for explaining variation in credit spreads. Similarly, the VRP has been shown to be a good predictor of bond excess returns, especially over short forecasting horizons (Zhou, 2018).

The early research on the VRP has focused mainly on the US stock market, using the fact that the VIX provides a relatively long time series on implied volatility that is needed to estimate a measure of the VRP (see Equation 2). Yet, in subsequent years, several additional trading venues started to construct implied volatility indices for their respective stock markets (Fassas and Papadamou, 2018). This allowed for the study of the VRP across international markets. In this context, Bollerslev et al. (2014), for example, report a strong prediction power for the VRP in seven developed markets. Fassas and Papadamou (2018) find similar results for an even larger group of international markets, including some emerging markets such as India and South Korea. There is, however, some evidence in support of the idea that the role of the VRP may differ across international markets. Qiao et al. (2019), for example, conclude that the VRP in emerging markets is better able to predict stock returns over longer time horizons, whereas the VRP in developed markets is connected to return predictability for shorter horizons. Londono (2015), on the other hand, finds that the predictive power of the VRP for local stock returns does not hold for a number of international markets. However, he finds that the US' VRP has predictive power for stock returns in all the countries in his sample, which suggests that the VRP in the US may be particularly important given the size and depth

of its financial markets.

Finally, we can see from Equation 2 that the VRP consists of two parts: the implied variance of traded options, that can be observed in the market, and the physical measure of realized variance over the same period. The latter, however, has to be estimated from the data. Here, different approaches have been proposed. The early studies on the VRP simply used the historical variation of stock prices as a proxy of the variation over the next month (see Bollerslev et al., 2009). With improvements in realized variance forecasting over the last decade, however, more recent research has started to apply different forecasting techniques in order to arrive at a more precise measure of next month's realized variance. Bekaert and Hoerova (2014), for example, test a range of different forecasting models and then select a model containing the squared VIX and the lagged daily, weekly and monthly realized variance. The forecasts from this model are then used to compute the variance risk premium for each point in time. Similarly, Drechsler and Yaron (2011) and Johnson (2019) also apply a variance forecast to derive a measure of expected realized variance. Both studies employ a simple regression forecast using the fitted values of a regression of future realized volatility on historical volatility and the current value of the squared VIX. Moreover, in a recent study, Li and Tang (2022) use an ensemble model that combines different machine learning models for forecasting realized variance. The authors found that the forecasts from this ensemble model significantly predicted monthly stock returns for a sample of individual stocks as well as for the broader S&P-500.

For this thesis, I combine two important areas of research from the literature on the VRP. The first part of this thesis, investigated the performance of eight different variance forecasting models across a large sample of equity indices. Moving forward, I will use the best-performing forecasting model for each country to compute a forward looking estimate of physical variance which can then be used to comute the VRP over time for each of the indices. Finally, I test the predictive power of the VRP across all markets. Hence, I extend the variance forecasting approach of Bekaert and Hoerova (2014) to a sample of 16 international stock markets.

3 Data

The analysis in this thesis builds on three different samples. The first sample consists of 16 datasets of realized variance for the respective international equity markets. This data was taken from the Oxford-Man Institute of Quantitative Finance and consists of daily realized variances. Section 4 explains the notion of realized variance in more detail. Second, I use a sample of 16 volatility indices, for the different international stock markets, such as, for example, the VIX for the S&P-500. Here, the data was taken from Bloomberg and also consists of daily data. Finally, for the same 16 stock market indices, I am also using the daily closing prices of the respective stock index to compute returns over different time horizons. The stock market data was taken from Refinitiv and Bloomberg. Table 1 presents all indices in my analysis and the volatility index for the respective equity market. Most of the volatility indices are constructed following the VIX index and show the volatility over the next month as implied by current option prices. Yet, the computation of the individual index may differ between markets but a full presentation of the methodology of all 16 volatility indices is beyond the scope of this thesis.[1]

Note that the sample period differs across markets as the volatility indices were introduced at different points in time and not all indices are still in use. Furthermore, the final sample period for the regression analysis also depended on the availability of realized variance data for the respective stock market as the variable of interest is the variance risk premium. Hence, complete data for implied and realized volatilities was needed. The period that the regression analysis in Section 5 was ultimately based on can be found in Table 3 which will be introduced in the next section.

4 Realized variance forecasting

As explained in Section 1, I define the VRP as the difference between a market volatility index, such as, for example, the VIX for the S&P-500, and the realized volatility in that respective market. For this thesis, I followed the approach by Bekaert and Hoerova (2014), who forecast next month's realized volatility using several forecasting models. This allows

[1]Interested readers may look at Fassas and Papadamou (2018), who provide a detailed explanation of different volatility indices together with their history. Most of the indices they cover are also present in this thesis.

Table 1: Equity and Volatility Indices
The table presents all 16 stock market indices used for this thesis. The first column displays the Bloomberg Ticker of the respective index, the second column contains the official name, and the third column specifies the country where the index is located. The fourth column shows the volatility index used for the respective market. Note that the available data for the volatility indices varied widely across markets. The sample period included in the analysis is displayed in Table 3 as it also depends on the availability of the realized variance data for the respective market.

Ticker	Equity Index	Country	VOL Index
AEX	Euronext Amsterdam	Netherlands	AEX Volatility Index
AORD	S&P/ASX All Ordinaries	Australia	TSX 60 VIX Index
BVSP	Bovespa Index	Brazil	CBOE Brazil ETF Volatility Index
DJI	Dow Jones Index	USA	CBOE DJIA Volatility Index
FCHI	CAC40	France	CAC 40 Volatility Index
FTSE	FTSE 100 Index	UK	FTSE 100 Implied Volatility Index
GDAXI	DAX	Germany	VDAX
HSI	Hang Seng Index	Hongkong	HSI Volatility Index
IXIC	NASDAQ Composite	USA	NASDAQ-100 Volatility Index
KS11	KOSPI Composite Index	South Korea	VKOSPI
MXX	IPC Mexico	Mexico	VIMEX
N225	Nikkei	Japan	Nikkei Stock Average Volatility Index
NSEI	Nifty 50	India	India VIX
RUT	Russel 2000	USA	CBOE Russell 2000 Volatility Index
SPXCASH	S&P-500 Cash Index	USA	VIX
STOXX50E	Stoxx 50 Europe	Europe	VSTOXX

the construction of a forward-looking measure of the VRP which relies only on existing, historical data and can therefore be used for real-world investment strategies.

Within the scope of this thesis, I altered the approach by Bekaert and Hoerova (2014) in two important ways to add new insights to the existing literature. First, I investigate the role of the VRP across several international stock markets. To do this, I compared the performance of eigth different volatility forecasting techniques across all 16 equity markets in the first part of this thesis.[2] Second, I extended the variance forecasting task to a number of different models including different extensions of the well-known HAR model as well as three machine learning models. This extension to state-of-the-art forecasting

[2]Note that in the first part of this thesis, I used a larger sample of stock markets. Yet, not for all of those markets an implied volatility index exists, which is why I had to reduce the sample size for this part of the thesis that is focused on the VRP.

techniques allows me to investigate how more accurate variance forecasting techniques alter the predictive power of the VRP.

To evaluate the performance of the different forecasting techniques, I initially trained the models on the first 1000 days of my sample. After this, I re-estimate the model parameters every month. Following re-estimation, I produce 20 daily, out-of-sample forecasts. After each month, the sample that the model is trained on increases by the days of the last month and the parameters are re-estimated. Hence, I apply an expanding window scheme with recursive backtesting. Yet, even though re-estimation of the parameters takes place on a monthly basis, this approach allows daily realized variance forecasts, which are highly relevant for real-world risk management. Finally, the out-of-sample performance of all models and across all markets was evaluated based on three loss functions: the mean squared error (MSE), the quasi-likelihood error (QLIKE), and the mean absolute percentage error (MAPE). The final selection of the models for this part of the thesis was based on the MSE but results for the other loss functions as well as Diebold-Mariano test results can be found in the first part of the thesis. The reason for the monthly re-estimation of the models was the large number of markets, the relatively large sample size, and the fact that I also tested the performance of three machine learning models which are more computationally expensive. Daily re-estimation of all models was therefore not feasible within the scope of this thesis and monthly re-estimation seemed to strike a good balance between computational efficiency and real-world applicability.

The remainder of this section will explain the forecasting models selected for the respective countries. Table 2 displays the best-performing model for each of the 16 stock market indices used for this thesis. The results shown describe the forecasting performance at the monthly horizon. Hence, the forecasted variable was the average realized variance over the next month RV_t^m, defined on day t as

$$RV_t^m = \frac{1}{22} \sum_{i=t+1}^{t+21} RV_i^d, \tag{3}$$

where RV_i^d is the daily realized variance on day t and a month consists of 22 trading days. Daily realized variance, consisting of intra-day returns, can be defined as

$$RV_t^d = \sum_{i=1}^{M} = r_{t-j\Delta}^2, \tag{4}$$

where, following Corsi (2009), I define $M = h/\Delta$ as the number of intra-day returns taken into account. Hence, Δ is the frequency of intra-day returns, i.e., five minutes for my analysis and h is the number of trading hours in the respective market.[3] The intra-day returns can be expressed as

$$r_{t-j\Delta} = log(P_{(t-j\Delta)}) - log(P_{t-(j+1)\Delta}), \tag{5}$$

where t indexes the day and j the time during the day. Realized variance is a consistent estimator of integrated variance when the intra-day sampling frequency goes towards 0, $\Delta \to 0$, and hence, the number of intra-day returns becomes very large. This analysis is based on five-minute intra-day returns, a common choice in the literature on volatility forecasting (see for example Bollerslev et al., 2016, or Clements and Preve, 2021).

Given the forecasting results, I will proceed to explain all models mentioned in Table 2. In addition, I will briefly introduce the benchmark HAR model, as several other models build on its main idea. Note, however, that the forecasting exercise included three additional models that are not mentioned here: the HAR-J, the Support Vector Machine, and the Long-Short-Term-Memory Model. A detailed explanation of these models as well as their performance in forecasting realized variance across all stock market indices can be found in the first part of this thesis. Furthermore, a comprehensive output table containing the numerical results for all indices can be found in Appendix A.1.

4.1 Heterogeneous Autoregressive Model

The benchmark Heterogeneous Autoregressive model (HAR), as introduced by Corsi (2009), estimates integrated volatility via realized volatility from intra-day returns (see Equation 4). The aim of the HAR model is to forecast future realized variance, based on daily, weekly, and monthly average realized variances that are known today. Daily realized variance was defined in Equation (4) whereas the weekly average realized variance can simply be expressed as

$$RV_t^w = \frac{1}{5}(RV_t^d + RV_{t-1}^d + \cdots + RV_{t-4}^d) \tag{6}$$

[3]For 6.5 trading hours and five-minute sampling frequency, for example, $M = 6.5 \times 12 = 78$

Table 2: Best-performing forecasting models across equity indices
The table presents all 16 stock market indices used for this thesis. The first column displays the Bloomberg Ticker of the respective index, the second column contains the official name of the index, and the third column specifies the country where the index is located. The fourth column shows the best-performing variance forecasting model for the respective market. For each market, the model with the lowest mean squared error (MSE) was selected as the best-performing model. The forecasted variable was monthly realized variance.

Ticker	Equity Index	Country	Best performing model
AEX	Euronext Amsterdam	Netherlands	Random Forest
AORD	S&P / ASX All Ordinaroes	Australia	RR-HAR
BVSP	Bovespa Index	Brazil	WLS-HAR
DJI	Dow Jones Index	USA	RR-HAR
FCHI	CAC40	France	Random Forest
FTSE	FTSE 100 Index	UK	C-HAR
GDAXI	DAX	Germany	RR-HAR
HSI	Hang Seng Index	Hongkong	RR-HAR
IXIC	NASDAQ Composite	USA	Random Forest
KS11	KOSPI Composite Index	South Korea	RR-HAR
MXX	IPC Mexico	Mexico	C-HAR
N225	Nikkei	Japan	RR-HAR
NSEI	Nifty 50	India	Random Forest
RUT	Russel 2000	USA	RR-HAR
SPXCASH	S&P-500 Cash Index	USA	RR-HAR
STOXX50E	Stoxx 50 Europe	Europe	C-HAR

Similarly, the monthly average realized variance is given by

$$RV_t^m = \frac{1}{22}(RV_t^d + RV_{t-1}^d + \cdots + RV_{t-21}^d), \tag{7}$$

assuming 22 trading days in a month. The benchmark HAR model can then be estimated as

$$RV_{t+k} = \beta_0 + \beta_1 RV_t^d + \beta_2 RV_t^w + \beta_3 RV_t^m + u_{t+k} \tag{8}$$

where k determines the forecast horizon. In the case of this thesis, $k = 22$, as I focus on forecasting monthly realized variance. In the standard HAR model, Equation (8) will be estimated using ordinary least squares (OLS).

4.2 Continuous HAR

The continuous HAR model (C-HAR) builds on the work by Barndorff-Nielsen and Shephard (2004), who show that bipower variation can be used as an estimator that is robust against rare jumps in the data. This allows us to separate realized volatility from such jumps. Bipower variation on day t is defined as

$$BPV_t = \frac{\pi}{2} \sum_{i=1}^{M-1} |r_{t,i}||r_{t,i+1}|. \tag{9}$$

Here, M is the sampling frequency as in Equation 4 and $r_{t,i}$ is the return on the respective asset on day t in sampling interval i. Bipower variation was then used by Corsi and Reno (2009) to expand the HAR model. The authors proposed a continuous volatility HAR, which uses the previous values of bipower-variation, instead of the previous values of realized variance, to explain today's realized variance. The C-HAR can hence be written as

$$RV_{t+k} = \delta_0 + \delta_1 BPV_t^d + \delta_2 BPV_t^w + \delta_3 BPV_t^m + \epsilon_{t+k}, \tag{10}$$

where BPV_t^w and BPV_t^m are the weekly and monthly averages of the bipower-variation, respectively. In a recent study by Rahimikia and Poon (2020) the C-HAR outperforms all other HAR versions tested.

4.3 Weighted-Least-Squares HAR

As mentioned above, the standard HAR model, as well as the C-HAR, is estimated via OLS. Yet, given the known characteristics of return data and realized volatility, such as outliers, conditional heteroskedasticity and a return distribution that does not follow a normal distribution (Clements and Preve, 2021), one should consider other estimation techniques. This seems particularly important as estimation results from OLS are known to be vulnerable to violations of the assumptions described above. Hence, Clements and Preve (2021) propose a range of new specifications of the HAR model, that follow the basic concept of HAR models, but use different input variables and different estimation techniques. One such specification of the standard HAR model is the weighted-least-squares HAR model (WLS-HAR). Clements and Preve (2021) find that the WLS-HAR

outperforms the benchmark HAR and other well-known HAR extensions. While OLS minimizes the sum of squared residuals, weighted least squares (WLS) weighs these residuals and minimizes the weighted sum. Therefore, if an adequate weighting scheme is chosen, WLS can be a more efficient estimator compared to OLS which would result in more precise volatility forecasting. One can estimate the WLS-HAR as in Clements and Preve (2021) via the following minimization problem:

$$\min_{\phi_0,\phi_1,\phi_2,\phi_3} \sum_{t=23}^{n} w_t (RV_{t+k} - \phi_0 - \phi_1 RV_t^d - \phi_2 RV_t^w - \phi_3 RV_t^m)^2, \tag{11}$$

where w_t is the weight on every observation t and n is the number of days in the sample. Clements and Preve (2021) test a range of different weighting schemes. The authors identify a model using weights that are inversely proportional to the realized quarticity day t to be the model with the best performance in forecasting realized variance. Realized quarticity is the fourth moment of intraday returns and can be seen as the volatility of the volatility. Therefore, using these weights, the WLS-HAR will put less weight on days on which there is high uncertainty around the volatility measure. However, Clements and Preve (2021) also test a model using weights $w_t = 1/RV_{t-1}$ and found its performance to be comparable to the model using realized quarticity due to the high correlation between realized variance and realized quarticity. Therefore, they conclude that such a model represents a good alternative for situations in which realized quarticity is not available for all observations. Hence, as realized quarticity was not available for a large share of indices used for this thesis, I estimated the WLS-HAR using the above-mentioned weights that are inversely proportional to the first lag of realized variance.

4.4 Robust Regression HAR Model

The final HAR model that will be used for this thesis is the HAR estimated via Robust Regression (RR-HAR). This estimation approach was also proposed by Clements and Preve (2021). Similarly to the WLS-HAR, Clements and Preve (2021) conclude that estimating the HAR model using robust regression, instead of OLS, significantly improves its performance. Robust regression (RR) was pioneered by Huber (1964) and represents an estimation technique that is more robust to outliers in the data. These outliers can pose a severe problem for OLS, known to perform poorly under the existence of unusual

observations in the data. If we wish to estimate a linear model

$$\begin{aligned} Y_i &= \alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \epsilon_i \\ &= \mathbf{x_i'}\boldsymbol{\beta} + \epsilon_i \end{aligned} \tag{12}$$

as in the case of the HAR estimation, RR applies an objective function $\sum_{i=1}^{n} \rho(\epsilon_i)$ to the residual ϵ_i. Estimators that minimize an objective function are called *M estimators*. One of the most common *M estimators* is Tukey's biweight function (Fox, 1997) which I will use for this analysis, in line with the work by Clements and Preve (2021). Tukey's biweight function can be defined as

$$\rho(\epsilon) = \begin{cases} \frac{k^2}{6}\{1 - [\,1 - (\frac{\epsilon}{k})^{\,2}\,]^{\,3}\} & \text{for } |\epsilon| \leq k \\ \frac{k^2}{6} & \text{for } |\epsilon| > k, \end{cases} \tag{13}$$

where k is the tuning constant that is used to distinguish between the center and the tails of the distribution of residuals.[4] The tuning parameter k is usually expressed as a function of the scale of the dataset, for example, $k = cS$, where variable S is defined as a quantification of scale (Fox, 1997). A commonly used and robust choice for S is the *median absolute deviation* (MAD), which is defined as

$$MAD = \text{median}|y_i - \hat{\mu}|, \tag{14}$$

where $\hat{\mu}$ is an estimate of the sample mean. As we can see in Equation (13), the objective function flattens out for residuals above k. Hence, a smaller tuning parameter can be used to adjust the robustness of the RR model.[5] Finally, *iterative weighted least squares* (IWLS), can be used to compute the *M estimators*. For this, we define a weight function $w(\epsilon) = \frac{\rho'(\epsilon)}{\epsilon}$, where $\rho'(\epsilon)$ is the first derivative of the objective function in (13) with respect

[4]Note that for the sake of readability, I suppressed the subscripts of the residuals as defined in Equation 12

[5] I used R's "rlm" function from the "MASS" package and its default value for c, when applying Tukey's biweight function and the MAD as a measure of scale.

to ϵ. The weight function can therefore be written as

$$w(\epsilon) = \begin{cases} [\,1 - (\frac{\epsilon}{k})^{\,2}\,]^{\,2} & \text{for } |\epsilon| \leq k \\ 0 & \text{for } |\epsilon| > k. \end{cases} \tag{15}$$

The weight function illustrates the fact that no weight will be put on observations with residuals that exceed the tuning parameter k. For regression analysis, the *M estimator* is applied to the linear model (see Fox, 1997, for all intermediate steps), and, using the weight function, the estimation equation becomes

$$\sum_{i=1}^{n} w_i(y_i - \mathbf{x_i'}\boldsymbol{b})\mathbf{x_i} = \mathbf{0}, \tag{16}$$

which minimizes the weighted sum of squares. Yet, the weights depend on the residuals, the residuals depend on the coefficients, and those depend on the weights again (Fox, 1997). Hence, this minimization problem must be solved with an iterative method. IWLS sets initial regression coefficients and uses those to back out the residuals, which allows us to retrieve the scale S and thereafter the weights w_i. Then, IWLS proceeds at each iteration by solving equation 16 with the current weights to derive new regression coefficients $\mathbf{b}$. The algorithm follows this iteration process until the changes in the regression coefficients become insignificant and $\mathbf{b}^{(I)} - \mathbf{b}^{(I-1)} \approx 0$, with I being the iteration counter (Fox, 1997).

4.5 Random Forest

In addition to the different HAR model variations, I also tested the forecasting performance of several machine learning models. Here, the best-performing one was the random forest.[6] Random forests can be described as a machine learning algorithm that generates and combines a large number of decision trees to reduce the noise embedded in individual trees. Decision trees can be used for classification and regression problems, however, this explanation will focus solely on so-called regression trees given the numerical nature of

[6]Note that the two other machine learning models I tested, Support Vector Machines and the Long-Short-Term-Memory model, also performed well for several equity indices but under different loss functions than the MSE. Yet, across all loss functions and forecasting horizons, the first part of this thesis revealed that the random forest seemed to be the best-performing of the three machine learning model.

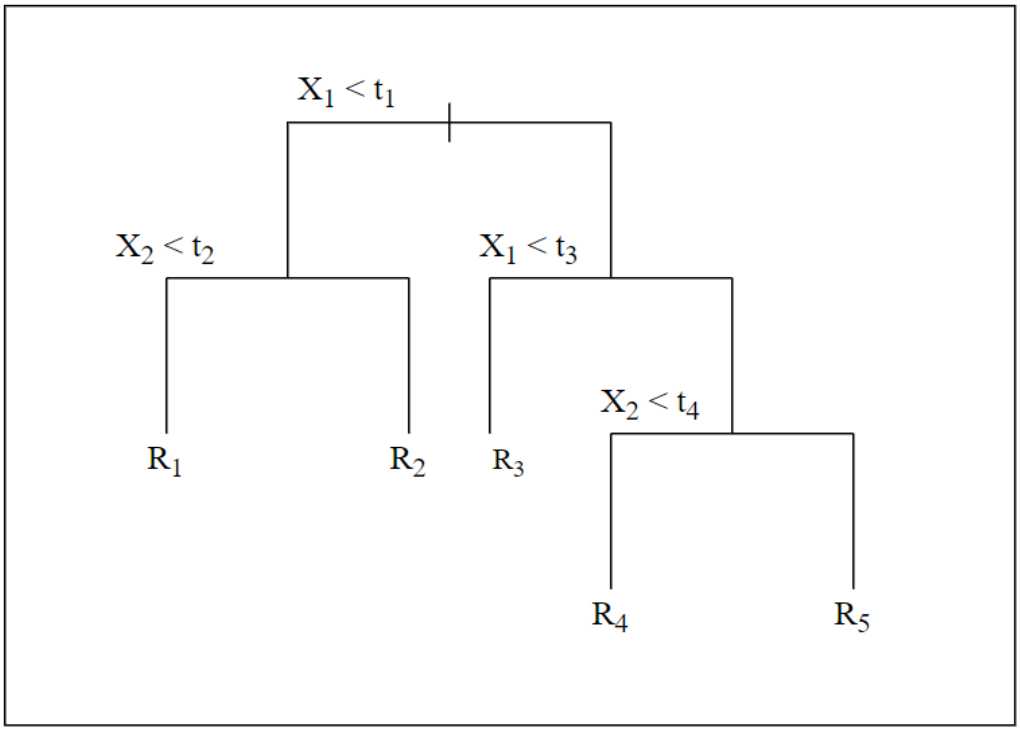

Figure 1: Illustration of a simple regression tree
The figure displays the structure of a simple regression tree using two predictors, X_1, X_2 and four internal nodes $t_1, \ldots, t_4$, resulting in five different external nodes $R_1, \ldots, R_5$. The figure is displayed as in Hastie et al. (2009).

forecasting volatility. Regression trees apply a set of splitting rules to each observation of the data set. Figure (1) shows the shape of a simple regression tree using two predictors, X_1, X_2. A regression tree is characterized by its *internal nodes* and its *terminal nodes* or *leaves* (Witten and James, 2013). Internal nodes are the points inside the tree where the individual observation is assigned to the left or right branch of the tree, depending on a certain parameter. On the top of the tree in Figure (1), for example, we see that an observation is assigned to the left part of the tree if $X_1 < t_1$, where t_1 is the first parameter used to separate the data. External nodes, also called *regions*, are the different outcomes of the regression trees: five in the case of Figure (1).

In order to construct the tree, all variables are scanned to find the combination of splitting variables and splitting points that split the data set most efficiently. This procedure is repeated at each internal node using the remaining observations in the respective branch of the tree (Hastie et al., 2009). If regression trees are grown sufficiently deep, meaning that they have a high number of internal nodes, they tend to fit the data well and show low bias. Yet, regression trees are known to produce noisy estimates.

Therefore, several machine learning algorithms have been proposed to solve this problem, *random forests* being one of them (Hastie et al., 2009).

The concept behind random forests is to generate a large amount of uncorrelated trees and compute the average of their predictions. The main characteristic of random forests is, that in order to obtain uncorrelated trees, a subset of $m \leq p$ of the predictors is drawn before generating a new tree, where p is the total number of independent variables in the data set. Usually, the number of predictors used per tree is selected as $m = \sqrt{p}$ (Hastie et al., 2009). Furthermore, the individual trees are trained on a sub-sample of the data set which means that not all available observations are used for each tree. Hence, each new tree does not see all of the existing data, which leads to a lower correlation between trees. For regression problems, the final prediction of the random forest will be the average of all generated trees. As the data sets generated from the sub-sampling are identically distributed, the expectation of their averages will be identical. Therefore, the only possibility to improve the performance of the forecast is by reducing the variance (Hastie et al., 2009). The variance of the average can be written as

$$\rho\sigma^2 + \frac{1-\rho}{N}\sigma^2, \tag{17}$$

where ρ is the correlation between trees, N is the number of trees generated, and σ^2 is the variance of the individual tree. Equation 17 shows that the term on the right side converges to zero as N increases and the term on the left side decreases with a lower correlation between trees (Hastie et al., 2009). The size of the sub-sample, the number of predictors in each sub-sample, and the number of trees must be selected by the researcher in a way that minimizes the variance of the prediction. The selection of hyperparameters is often carried out via grid search or comparing the performances of different models. This approach, however, was not feasible for this thesis, given the large number of stock indices and the monthly re-estimation of the models. Hence, hyperparameters were selected based on well-known heuristics or in line with the literature on volatility forecasting. The number of trees was set to 500 and each tree was trained on a sub-sample of 50% of the observations as in Li and Tang (2022). The number of trees should be sufficiently large as this generally improves results (Probst et al., 2019) if it does not significantly impair computation time. The number of predictors, m, that each individual tree was trained on, was selected based on the heuristic $m = \sqrt{p}$ which is a common choice in the literature

and was found to yield similar results compared to other heuristics (Bartz et al., 2023). The predictors that the random forest was trained on consisted of the past 22 lags of realized variance as well as lagged daily, weekly, and monthly bipower variation.

Overall, past research has illustrated the strong performance of the random forest in forecasting realized variance. Luong and Dokuchaev (2018), for example, show that using the random forest algorithm can significantly improve the forecasting ability of the HAR model. Similarly, Christensen et al. (2023) finds that random forest outperforms the HAR model for the one-day forecast horizon when a wide range of variables are used, not just the main variables from the HAR model. For a one-week horizon, the random forest produces a smaller error than the HAR model, even if using only the core HAR variables. Li and Tang (2022) also use the random forest as part of their ensemble model that combines different machine learning techniques. The authors report that the random forest outperforms their HAR model at the monthly and quarterly horizon.

4.6 Computation of the variance risk premium

Finally, we can use the realized variance forecasts described above to compute the VRP for every equity index. For this, we have to square the volatility index of each country and divide it by 12 as it is stated in annualized terms and as a standard deviation. Then we can compute the VRP on day t as

$$VRP_t = IV_t^2 - E[RV_t^m], \tag{18}$$

where IV_t^2 is the monthly, squared implied volatility index for each index and $E[RV_t^m]$ is the realized variance forecast based on the best-performing model for each stock index. The best-performing models were selected as described in Table 2. Figure 2 illustrates the VRP for the four US stock market indices in the sample. Figure 3 displays the VRP for four important European stock market indices. The graphs for all indices can be found in Appendix A.2.

The two figures reveal some important characteristics of the VRP. First of all, we see clear spikes in the VRP around the heights of the Global Financial Crisis in 2008 and the outbreak of the Covid-19 pandemic in early 2020. One also observes a significant increase in the VRP in most European markets during the euro-crisis from 2010 to 2012.

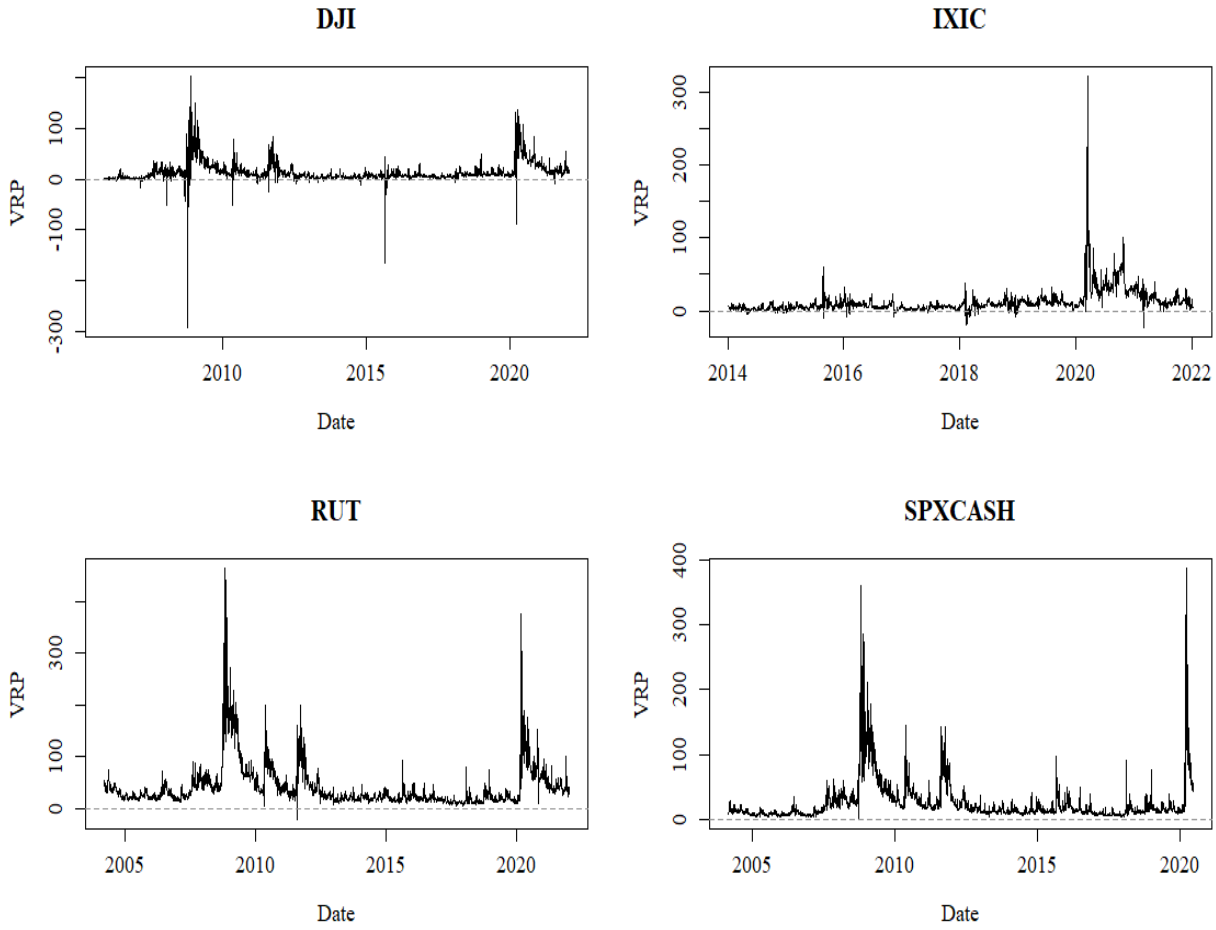

Figure 2: Variance Risk Premium - US Stock Market Indices
Historical evolution of the Variance Risk Premium (VRP) for the four US stock market indices in the sample: Dow Jones Index (DJI), NASDAQ-100 (IXIC), Russel 2000 (RUT) and the S&P500 Cash index (SPXCASH). The VRP is defined as in Equation 18. Graphs for all other stock market indices can be found in Appendix A.2.

Furthermore, we observe short but pronounced periods of negative VRP in line with the findings presented by Bollerslev et al. (2015). Further, Cheng (2019) conclude that these periods of negative premium are not due to mismeasurement of realized variance but tend to forecast high ex-post volatility. These periods of negative VRP, however, cannot be found in all stock markets in my sample. Yet, a common pattern of the VRP across different markets seems to be a baseline VRP slightly above zero interrupted by intervals of high VRP during periods of financial market stress. Table 3 displays summary statistics on the VRP across all 16 equity markets in the sample and illustrates significant heterogeneity across markets. The maximum VRP, in particular, varies significantly between equity markets. Here, three emerging markets can be observed as showing significantly higher

Table 3: Summary Statistics - Variance Risk Premium
Mean, maximum value, minimum value, standard deviation, kurtosis, and skewness for all 16 stock market indices included in the analysis. The last two columns display the start and end of the sample for each equity index. The periods for the individual indices vary as the availability of implied volatility indices was not constant across indices. The variance risk premium was computed following Equation 18. Monthly expected realized volatility was forecasted with the best-performing forecasting model for the respective market as displayed in Table 2. The indices' full names can be found in Table 2.

	Mean	Max	Min	St.Dev	Kurtosis	Skewness	Start	End
AEX	19.6	606.6	-52.1	36.3	63.1	6.6	02/2004	12/2020
AORD	22.1	296.2	-216.6	29.3	21.8	3.3	01/2008	01/2022
BVSP	85.8	1374.7	-94.4	84.9	80.4	6.9	03/2011	01/2022
DJI	13.3	201.5	-291.4	19.3	34.9	0.3	11/2005	01/2022
FCHI	15.3	424.2	-62.7	24.2	84.9	7.4	02/2004	12/2020
FTSE	8.5	238.5	-227.3	19.6	26.4	3.1	03/2004	01/2022
GDAXI	26.0	877.2	-116.0	41.8	107.5	8.1	10/2006	01/2022
HSI	64.3	1202.7	5.7	79.3	27.4	4.2	04/2004	01/2022
IXIC	12.3	320.5	-22.0	17.6	82.6	6.8	01/2014	01/2022
KS11	17.4	159.7	-64.7	13.4	17.2	3.2	04/2009	01/2022
MXX	24.4	312.4	-133.4	35.9	22.9	4.1	03/2004	06/2017
N225	33.0	466.6	-36.0	41.8	29.7	4.7	05/2004	12/2021
NSEI	133.7	2309.8	5.5	198.9	16.3	3.4	03/2008	01/2022
RUT	41.6	462.8	-20.0	41.4	17.4	3.5	03/2004	01/2022
SPXCASH	25.2	386.3	0.3	31.5	25.7	4.2	02/2004	06/2020
STOXX50E	14.5	260.6	-168.9	23.5	19.9	3.1	03/2004	01/2022

maximum values than the rest of the sample: the BVSP (Bovespa Index, Brazil), the HSI (Hang Seng Index, Hongkong), and the NSEI (Nifty 50, India). It is important to note that the large values of the VRP were driven primarily by large spikes in the implied volatility which could be due to less liquid option markets in these jurisdictions. However, these jumps in implied volatility all happened during periods of significant financial stress, the Global Financial Crisis, or the COVID-19 outbreak, and hence seem economically justified. Therefore, they were not removed as outliers for the scope of this analysis.

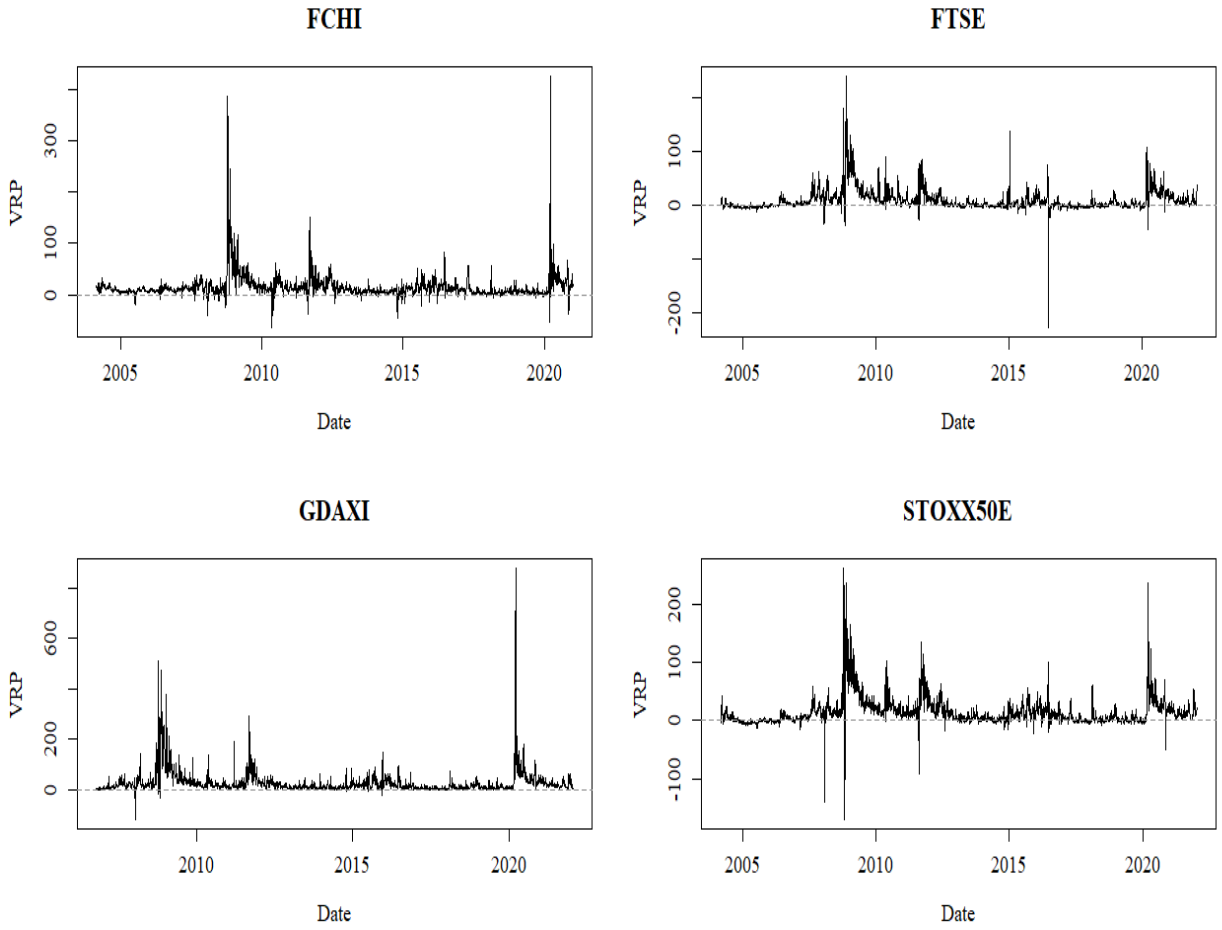

Figure 3: Variance Risk Premium - European Stock Market Indices
Historical evolution of the Variance Risk Premium (VRP) for four large European equity indices in the sample: CAC40 (FCHI), FTSE-100 (FTSE), DAX (GDAXI) and the Stoxx Europe 50 (STOXX50E). The VRP is defined as in Equation 18. Graphs for all other stock market indices can be found in Appendix A.2.

5 Stock market premium

After having defined the VRP for each equity market, I now turn to the question of the predictability of stock returns. For this, I rely on end-of-month observations for the return data and the variance risk premium which is a common choice in the literature on the VRP (see for example Bekaert and Hoerova, 2014 or Bollerslev et al., 2014). Hence, I use the observation on the last trading day of each month to compute returns and the VRP. Then, I follow the approach by Bollerslev et al. (2014) and estimate the return

predictability regression for stock market index i as

$$r^i_{t,t+h} = a^i_h + b^i_h VRP^i_t + u^i_{t,t+h}, \tag{19}$$

where $r^i_{t,t+h}$ is the annualized log return in excess of the 3-month T-Bill yield in stock market i over the h month return horizon. I use the 3-month T-Bill rate across all markets, independent of their geographic location, in order to increase comparability across markets.[7] Also, for some of the markets I analyzed, it was not clear whether the domestic short-term government bond yield could be seen as completely risk-free. a^i_h and b^i_h are the regression coefficients for stock market i over return horizon h, where the latter is the coefficient of main interest for this analysis. Finally, VRP^i_t is the variance risk premium for stock market i as computed at the end of month t. In line with Bollerslev et al. (2014) I estimate 1 to 12 months return regressions. To address the issue of overlapping regressions and autocorrelation in the data, I compute Newey-West standard errors as in Newey and West (1987). For this, I follow Bekaert and Hoerova (2014) and estimate a relatively large number of Newey-West lags using their heuristic of $NW = max\{3, 2 \times h\}$, where NW is the number of Newey-West lags and h is the return horizon as in Equation 19. Table 4 shows a summary of the regression coefficients and standard errors for all 16 stock market indices. Appendix A.3 presents detailed output tables for all individual equity markets.

The results in Table 4 reveal several interesting findings. First, we can observe that the VRP is a significant predictor of stock returns for all four US market indices used in my sample, which is in line with existing research on the US equity market (Londono, 2015). For the DJI (Dow Jones Index) and the IXIC (NASDAQ-100) the VRP significantly predicts returns across most of the return horizons. For the RUT (Russel 100) and the SPXCASH (S&P-500) the null hypothesis of no predictability cannot be rejected for the short-term return horizons. RUT and IXIC also seem to display the "hump" pattern as observed by Bollerslev et al. (2009) and Bekaert and Hoerova (2014), which describes lower short- and long-term return coefficients and the highest regression coefficient for medium-term returns. In terms of economic significance, the findings illustrate that for the IXIC, for example, a 1% (100 bps) increase in the monthly VRP leads approximately

[7] Similarly, I use simple returns of the respective index to increase comparability across indices as total return data series were not available for all markets.

Table 4: Variance Risk Premium Regressions across markets
Regression results estimated as in Equation 19. The regressions were based on end-of-month data and were estimated for the 1 to 12-month return horizon. The dependent variable is the annualized log return in percentages over the h return horizon in excess of the 3-month US T-bill rate. The table contains only the regression coefficient and the Newey-West standard errors in parentheses. The Newey-West standard errors were determined as $NW = max\{3, 2 \times h\}$, where h is the forecast horizon in months. For the sake of readability, not all return horizons are displayed in the table. Detailed information on all individual regressions can be found in Appendix A.3.*$p < .05$;**$p < .01$.

	Horizon							
	1	2	3	4	5	6	9	12
AEX	0.21 (0.28)	0.14 (0.25)	0.11 (0.2)	0.08 (0.17)	0.13 (0.13)	0.17 (0.1)	0.25** (0.07)	0.24** (0.06)
AORD	-0.28 (0.22)	-0.2 (0.18)	-0.14 (0.16)	-0.11 (0.15)	-0.08 (0.14)	-0.01 (0.13)	0.00 (0.11)	0.02 (0.08)
BVSP	0.38** (0.07)	0.33** (0.05)	0.29** (0.04)	0.28** (0.04)	0.24** (0.03)	0.2** (0.03)	0.17** (0.03)	0.14** (0.03)
DJI	0.64 (0.39)	0.42 (0.25)	0.41* (0.18)	0.44** (0.12)	0.44** (0.11)	0.42** (0.1)	0.35** (0.09)	0.31** (0.08)
FCHI	0.77 (0.52)	0.45 (0.38)	0.19 (0.32)	0.22 (0.25)	0.38* (0.17)	0.46** (0.15)	0.42** (0.14)	0.35** (0.12)
FTSE	0.00 (0.31)	0.03 (0.21)	0.06 (0.16)	0.11 (0.14)	0.13 (0.13)	0.13 (0.13)	0.17 (0.12)	0.17 (0.1)
GDAXI	0.4 (0.42)	0.38 (0.27)	0.35 (0.19)	0.38** (0.13)	0.38** (0.11)	0.38** (0.09)	0.31** (0.08)	0.31** (0.06)
HSI	0.01 (0.12)	0.00 (0.1)	0.00 (0.09)	0.01 (0.09)	0.03 (0.09)	0.04 (0.08)	0.03 (0.08)	0.01 (0.07)
IXIC	0.29 (0.55)	0.65** (0.2)	0.67** (0.14)	0.64** (0.16)	0.6** (0.15)	0.57** (0.12)	0.44** (0.09)	0.4** (0.07)
KS11	1.04* (0.42)	0.92** (0.31)	0.89** (0.27)	0.86** (0.23)	0.81** (0.23)	0.71** (0.24)	0.53* (0.25)	0.41* (0.19)
MXX	-0.07 (0.13)	-0.1 (0.12)	-0.08 (0.13)	0.00 (0.13)	0.04 (0.11)	0.07 (0.11)	0.08 (0.09)	0.08 (0.08)
N225	0.09 (0.12)	0.08 (0.1)	0.03 (0.09)	0.01 (0.08)	0.04 (0.06)	0.06 (0.05)	0.04 (0.04)	0.02 (0.04)
NSEI	0.03 (0.06)	0.01 (0.04)	0.00 (0.05)	0.01 (0.04)	0.02 (0.04)	0.02 (0.03)	0.03 (0.02)	0.03** (0.01)
RUT	0.29 (0.26)	0.23 (0.21)	0.23 (0.19)	0.24 (0.16)	0.26* (0.13)	0.27** (0.1)	0.23** (0.09)	0.19** (0.07)
SPXCASH	0.3 (0.31)	0.22 (0.25)	0.2 (0.21)	0.21 (0.18)	0.22 (0.14)	0.22* (0.1)	0.18** (0.06)	0.16** (0.05)
STOXX50E	0.78* (0.37)	0.54* (0.23)	0.38* (0.16)	0.35* (0.14)	0.33* (0.14)	0.29* (0.14)	0.21 (0.14)	0.19 (0.12)

to a 0.65% increase in excess returns over the following two months. Figure 4 illustrates these findings for the US markets. Second, a similar pattern emerges for the European

markets. Predictability is more pronounced for longer-term return horizons in the case of the AEX (Euronext Amsterdam), the FCHI (CAC40), and the GDAXI (German DAX Index). Here, the Euro Stoxx50 seems to be the exception where predictability is higher in the short run and then decreases steadily until it ceases to be statistically significant for return horizons over 9 months. Within the European markets the FCHI and the GDAXI can also be seen as expressing the "hump" pattern if we disregard the non-significant coefficients. Finally, for the FTSE we cannot reject the null hypothesis at the 5% level for any return horizon.

The R^2 differs substantially across markets, which is a pattern also reported by Bollerslev et al. (2014). For the US indices, the IXIC shows the highest R^2 with a value of 22.9%, whereas the R^2 for the SPXCASH is the lowest. Interestingly, the latter is significantly lower than observed by Bollerslev et al. (2014). For the European markets the R^2 ranges from 4.94% for the STOXX50E (Stoxx 50 Europe) to 13% for the GDAXI.[8]

My findings for the US and the European markets differ in some aspects from the evidence on the international VRP presented by Bollerslev et al. (2014). First, Bollerslev et al. (2014) find that the VRP does not consistently predict stock market returns for most European markets, whereas I find predictability for several markets, even if not for all of them. Second, they report the magnitude of the regression coefficient to follow the above-mentioned "hump" pattern. Yet, I only find this pattern for a few selected stock indices. Among the European markets, the FCHI and the GDAXI can also be seen as expressing the "hump" pattern if we disregard the non-significant coefficients. This difference in results could be due to two reasons: on the one hand, this thesis looks at a significantly longer sample period including the Covid-19 outbreak - an event with pronounced stock and option market volatility. On the other hand, I am using more complex variance forecasting techniques as well as different techniques across markets depending on the models' forecasting performance.

For the non-European and non-US indices, the results are more heterogeneous. The BVSP (Brazil) and the KS11 (South Korea) show a consistent pattern of predictability: statistically significant coefficients that are decreasing in magnitude over time. The BVSP also shows the highest levels of R^2 of all indices tested. The remaining four indices, however, show significantly smaller regression coefficients that largely fail to be statistically

[8]See Appendix A.3 for the R^2 of all individual regressions.

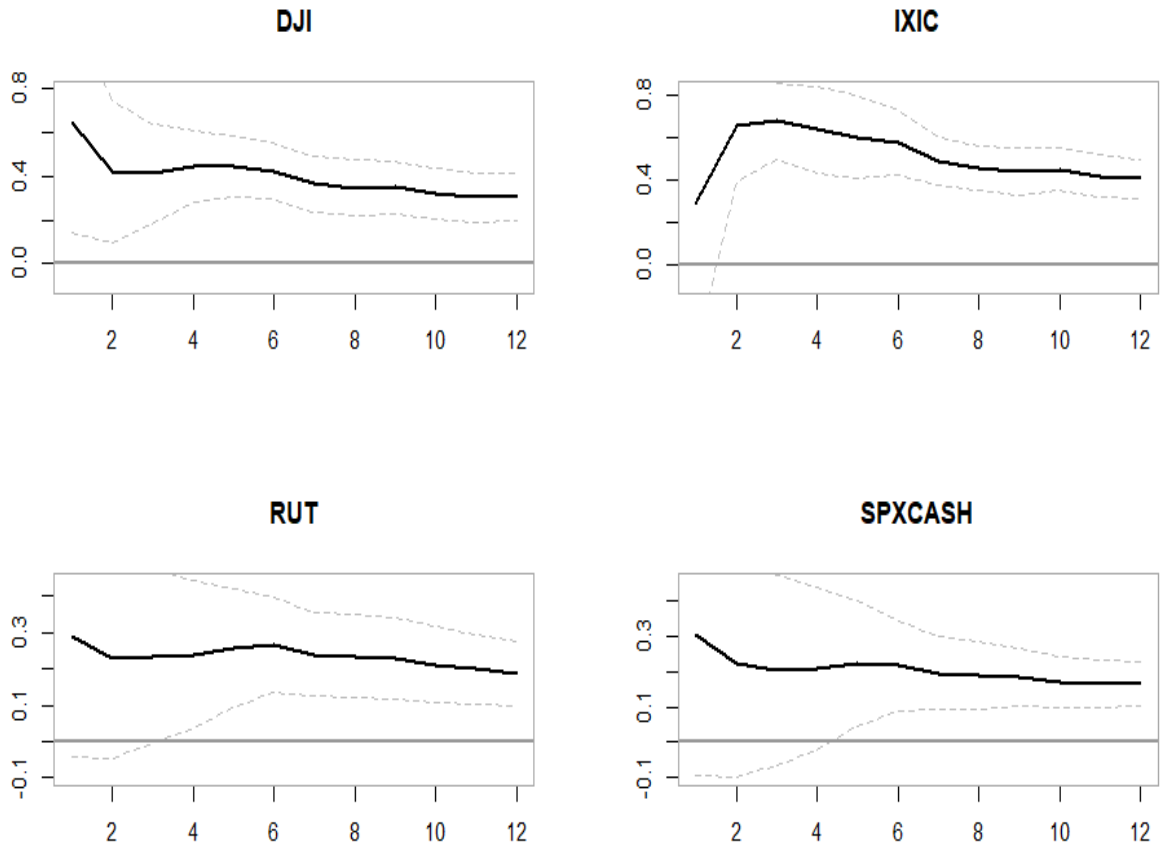

Figure 4: Variance Risk Premium - US Stock Market Indices
The figure illustrates the regression coefficient, as estimated in Equation 4, for the four US stock market indices in the sample: Dow Jones Index (DJI), NASDAQ-100 (IXIC), Russel 2000 (RUT) and the S&P500 Cash index (SPXCASH). The dashed lines display the 90% confidence interval around the estimated coefficient. The horizontal axis displays the $h = 1, \ldots, 12$ return horizon that the respective regression was based on.

significant. For the AORD (All Ordinaries, Australia) and the MXX (IPC Mexico) the sign of the coefficient even turns negative for several return horizons, contradicting prior research and economic intuition on the role of the VRP. For the N225 (Nikkei 225) I fail to reject the null hypothesis for a single return horizon and for the NSEI (Nifty 50, India) only the 12-month return coefficient, although relatively small, can be seen as statistically different from zero. These findings, however, are in line with prior research regarding the weaker predictive power of the VRP in non-US indices (see for example Bollerslev et al., 2014 or Londono, 2015). Yet, Londono (2015) also finds that the US' VRP, in this case, the VRP based on the S&P-500, significantly outperforms the local VRP in predicting stock market returns. Hence, as a final step of the analysis, I will investigate

whether the US VRP carries predictive power for those markets where the local VRP fails to predict returns: the AORD (Australia), the FTSE (UK), the HSI (Hongkong), the MXX (Mexico), and the N225 (Japan). Table 5 summarizes the result of the resulting regressions.

The US VRP seems to predict returns for the FTSE for return horizons above 7 months. The coefficients for the 6 and 7-month return horizons are significant at the 10% level. For all other indices, however, the US VRP cannot significantly predict future returns. The only exception is the MXX, for which, even though not displayed in the table, the null hypothesis of no predictability can be rejected at the 10% level for all return horizons greater than 5 months. The lack of predictability for the remaining non-European equity markets is somewhat in line with Londono (2015), who reports that the US VRP significantly predicts European markets whereas the evidence for predicting Japanese stock returns is mixed. Hence, there seems to be a difference in the effectiveness of return prediction of the US VRP between different markets, particularly when considering non-European markets. Overall, the lack of return predictability of the domestic or US VRP for 6 out of 16 markets, raises the question of why the VRP does not demand a risk premium across all markets and remains somewhat of a puzzle.

6 Robustness tests

In this section, I will investigate the robustness of the results presented in Section 5 to different modifications of the estimation approach. As noted in Section 5, the regression in Equation 19 was based on end-of-month data which is a common choice in the literature (see for example Bollerslev et al., 2014). In the following section, I will explore whether the results of my analysis are robust to using monthly data from the beginning, the middle of each month, or, more generally, using daily instead of monthly data. Furthermore, I will analyze whether the number of lags used to compute the Newey-West standard errors impacts the results of my analysis. Finally, in line with Johnson (2019), I will investigate whether the results are robust to an estimation using weighted least squares.

Table 5: US VRP's predictive power for international stock returns
Regression results based on Equation 19 with the difference that the predictor in each of the regression was the VRP of the S&P-500. The regressions were based on end-of-month data and were estimated for the 1 to 12-month return horizon. The dependent variable is the annualized log return in percentages over the h return horizon in excess of the 3-month US T-bill rate in the markets displayed in the first column. The equity markets displayed are the Australian All-Ordinaries (AORD), the FTSE-100 (FTSE), the Hongkong Hang Seng Index (HSI), the IPC Mexico (MXX), the Nikkei 225 (N225), and the Nifty 50 (NSEI). The table contains only the regression coefficient and the Newey-West standard errors in parentheses. The Newey-West standard errors were determined as $NW = max\{3, 2 \times h\}$, where h is the return horizon in months. For the sake of readability, not all return horizons are displayed in the table.*$p < .05$;**$p < .01$.

	Horizon							
	1	2	3	4	5	6	9	12
AORD	0.19 (0.22)	0.13 (0.17)	0.11 (0.15)	0.11 (0.13)	0.12 (0.11)	0.12 (0.1)	0.12 (0.09)	0.1 (0.07)
FTSE	0.29 (0.19)	0.21 (0.15)	0.17 (0.13)	0.15 (0.11)	0.15 (0.1)	0.15 (0.08)	0.16* (0.07)	0.16** (0.06)
HSI	0.36 (0.29)	0.25 (0.26)	0.27 (0.22)	0.29 (0.19)	0.28 (0.17)	0.26 (0.18)	0.22 (0.15)	0.17 (0.12)
MXX	0.3 (0.25)	0.25 (0.2)	0.23 (0.18)	0.22 (0.17)	0.23 (0.15)	0.25 (0.14)	0.25 (0.13)	0.22 (0.12)
N225	0.13 (0.25)	0.14 (0.22)	0.15 (0.18)	0.13 (0.14)	0.16 (0.11)	0.16 (0.1)	0.1 (0.09)	0.07 (0.08)
NSEI	0.39 (0.33)	0.32 (0.28)	0.32 (0.24)	0.35 (0.21)	0.34 (0.19)	0.32 (0.19)	0.25 (0.17)	0.19 (0.14)

6.1 Beginning of the month data

The regressions in Section 5 were estimated using end-of-month data. Table 6 presents the results of the same analysis using beginning-of-the-month data. Hence, I restrict my sample to the first trading day of each month. Given the results in Table 6 we can observe that the direction of the coefficients and the general conclusion from the analysis do not change when using beginning-of-the-month data. Yet, the magnitude of the regression coefficients changes significantly for several equity markets. A common pattern seems to be an increase in shorter-term coefficients, especially in the 1-month return prediction, and a decrease in the longer-term horizon. This is particularly salient for FCHI, FTSE, and IXIC. In terms of the statistical significance of the coefficients, there is only one important change: for the FTSE we can now reject the null hypothesis for several return horizons whereas this was not the case when using end-of-month data. Yet, for the other

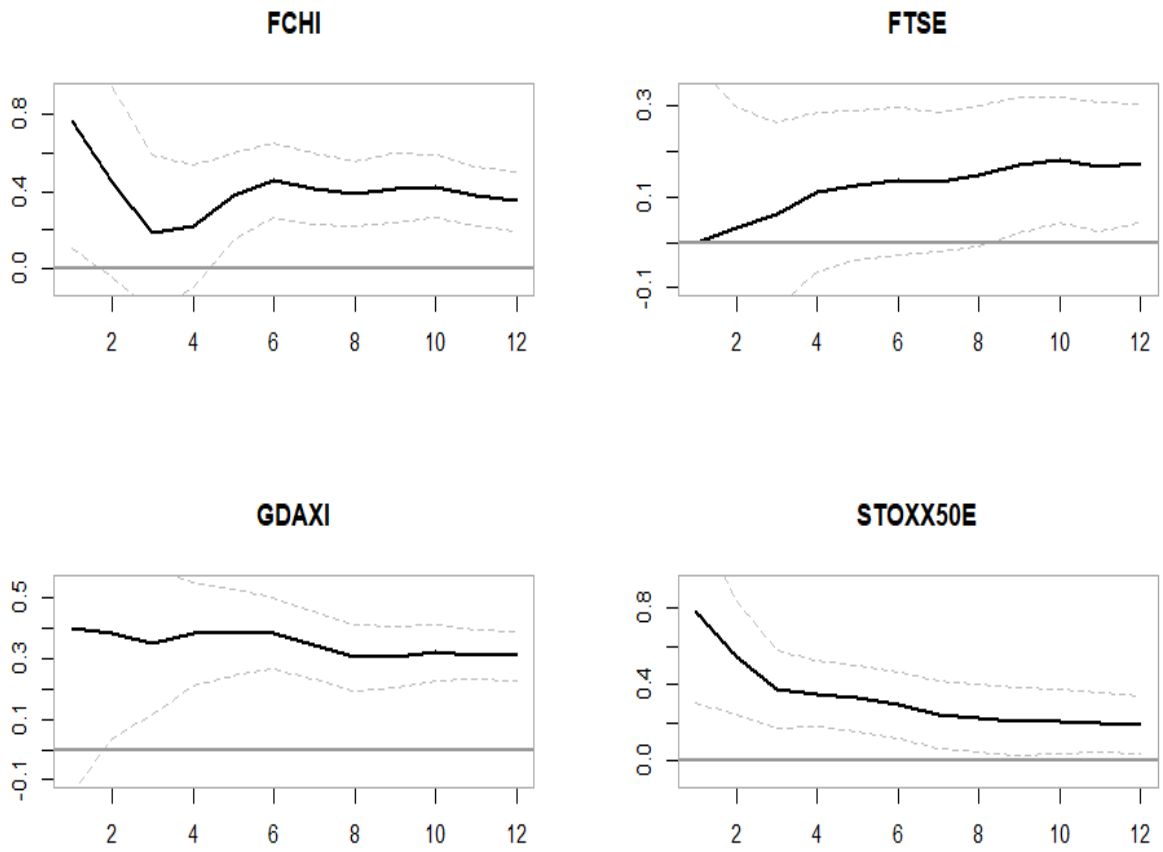

Figure 5: Variance Risk Premium - European Stock Market Indices
The figure illustrates the regression coefficient, as estimated in Equation 4, for the four US stock market indices in the sample: CAC40 (FCHI), FTSE-100 (FTSE), German DAX (GDAXI) and the Stoxx 50 Europe (STOXX50E). The dashed lines display the 90% confidence interval around the estimated coefficient. The horizontal axis displays the $h = 1, \ldots, 12$ return horizon that the respective regression was based on.

indices for which returns could not be explained by their domestic VRP, this does not change when applying regressions based on beginning-of-the-month data.

6.2 Middle of the month data

For this section, I will use data from the middle of the month to compute monthly returns for the different return horizons. I define the middle of the month to be the day that evenly splits all working days in a month. Table 7 presents the results from the regression based on the middle-of-the-month sample. Here, we still observe patterns of predictability across the different equity markets, yet, the results change notably for several markets.

Table 6: Variance Risk Premium - Beginning-of-the-month data
Results of the regression based on Equation 19. The regressions were based on beginning-of-the-month data and were estimated for the 1 to 12-month return horizon. The dependent variable is the annualized log return in percentages over the h return horizon in excess of the 3-month US T-bill rate. The table contains only the regression coefficient and the Newey-West standard errors in parentheses. The Newey-West standard errors were determined as $NW = max\{3, 2 \times h\}$, where h is the return horizon in months. For the sake of readability, not all return horizons are displayed in the table.*$p < .05$;**$p < .01$.

	Horizon							
	1	2	3	4	5	6	9	12
AEX	0.29 (0.26)	0.27 (0.22)	0.17 (0.2)	0.14 (0.15)	0.18 (0.1)	0.2** (0.07)	0.25** (0.05)	0.23** (0.05)
AORD	-0.1 (0.32)	-0.07 (0.24)	-0.04 (0.21)	-0.03 (0.2)	0.00 (0.17)	0.04 (0.16)	0.04 (0.13)	0.06 (0.08)
BVSP	0.4** (0.06)	0.35** (0.04)	0.31** (0.04)	0.28** (0.04)	0.23** (0.03)	0.19** (0.03)	0.16** (0.02)	0.13** (0.02)
DJI	0.96** (0.27)	0.5 (0.26)	0.37 (0.25)	0.4** (0.15)	0.4** (0.12)	0.37** (0.09)	0.3** (0.08)	0.27** (0.07)
FCHI	1* (0.5)	0.73 (0.4)	0.45 (0.35)	0.37 (0.24)	0.44** (0.15)	0.49** (0.12)	0.45** (0.12)	0.39** (0.11)
FTSE	0.59* (0.25)	0.31 (0.24)	0.24 (0.2)	0.23 (0.13)	0.21* (0.1)	0.23* (0.09)	0.23** (0.08)	0.22** (0.07)
GDAXI	0.65** (0.19)	0.47 (0.27)	0.45* (0.22)	0.39** (0.14)	0.37** (0.1)	0.3** (0.07)	0.27** (0.06)	0.27** (0.05)
HSI	0.03 (0.12)	0.04 (0.1)	0.02 (0.1)	0.03 (0.09)	0.05 (0.08)	0.06 (0.08)	0.04 (0.07)	0.03 (0.06)
IXIC	1.42** (0.49)	1.04** (0.37)	1** (0.22)	0.87** (0.21)	0.8** (0.18)	0.58** (0.1)	0.49** (0.09)	0.43** (0.07)
KS11	1.25** (0.35)	1.22** (0.23)	1.1** (0.21)	0.97** (0.21)	0.87** (0.23)	0.73** (0.23)	0.6* (0.26)	0.45* (0.18)
MXX	0.01 (0.19)	-0.05 (0.15)	-0.05 (0.16)	0.04 (0.14)	0.08 (0.12)	0.11 (0.11)	0.11 (0.08)	0.11 (0.07)
N225	0.06 (0.18)	0.1 (0.15)	0.04 (0.14)	0.04 (0.11)	0.08 (0.08)	0.09 (0.06)	0.06 (0.05)	0.04 (0.04)
NSEI	-0.01 (0.07)	0.00 (0.05)	-0.01 (0.05)	0.01 (0.04)	0.02 (0.03)	0.03 (0.03)	0.02 (0.02)	0.03** (0.01)
RUT	0.44* (0.2)	0.31 (0.18)	0.24 (0.17)	0.25 (0.14)	0.25* (0.11)	0.25** (0.08)	0.21** (0.07)	0.17** (0.06)
SPXCASH	0.47* (0.2)	0.28 (0.21)	0.2 (0.19)	0.21 (0.14)	0.22* (0.11)	0.21** (0.07)	0.17** (0.05)	0.15** (0.03)
STOXX50E	0.68* (0.33)	0.36 (0.36)	0.33 (0.26)	0.31 (0.16)	0.27* (0.12)	0.24* (0.1)	0.22* (0.09)	0.21* (0.08)

Interestingly, these changes mainly affect the markets where predictability was relatively strong based on end-of-month data. Now, however, we see that for several equity indices, we can only reject the null hypothesis for return horizons greater than 9 months. The

decrease in predictability is particularly striking for the US markets, with the exception of the IXIC, where the coefficients increase. Overall, the results from Table 7 show that the conclusions of my analysis are sensitive to the day of the month that we use to compute returns, especially when it comes to the short and medium-term return horizons.

6.3 Daily Data

Given the differences in my results using monthly data, based on whether I use data from the beginning, middle, or end of the month, I use this section to estimate Equation 19 based on daily data. Again, I estimate return predictability over the 1 to 12-month horizon. Hence, return horizon h in Equation 19 will vary from 21 for 1-month predictions to 252 for the 12-month prediction. The results of the analysis are displayed in Table 8. They reveal that the analysis using end-of-month data is relatively robust to using daily data. Predictability improves for the four US markets, whereas it decreases slightly for the STOXX50E. The most notable difference is probably that the regression coefficients on the FTSE are now different from zero for most of the return horizons, whereas we could not reject the null hypothesis at 5% for a single one of them using end-of-month data. Yet, we saw that the predictability for the FTSE improved significantly when using beginning or middle-of-the-month data. Thus, it seems like using daily data is a more efficient way to capture the predictability of the FTSE across different return horizons. Again, when it comes to those equity markets where the returns could not be explained by the local VRP, using daily data does not change the results based on the monthly data. Here, the only exceptions are two relatively small, but statistically significant, terms for the N225 and the MXX, which do not seem economically significant against the backdrop of the monthly analysis.

6.4 Number of Newey-West lags

[!htb] Existing research on the VRP mainly uses Newey-West standard errors to account for the heteroskedasticity and autocorrelation in the error term. Yet, to compute Newey-West standard errors, the researcher has to specify how many lags to include. Here, we observe some discrepancies between different publications. In the main analysis in Section 5, I followed Bekaert and Hoerova (2014) and computed the lags of the standard errors

Table 7: Variance Risk Premium - Middle-of-the-month data
Results of the regression based on Equation 19. The regressions were based on middle-of-the-month data and were estimated for the 1 to 12-month return horizon. The dependent variable is the log return in percentages over the h return horizon over the 3-month US T-bill rate. The table contains only the regression coefficient and the Newey-West standard errors in parentheses. The Newey-West standard errors were determined as $NW = max\{3, 2 \times h\}$, where h is the return horizon in months. For the sake of readability, not all return horizons are displayed in the table.*$p < .05$;**$p < .01$.

	Horizon							
	1	2	3	4	5	6	9	12
AEX	0.3 (0.19)	0.13 (0.12)	0.11 (0.12)	0.1 (0.1)	0.07 (0.08)	0.1* (0.05)	0.13** (0.02)	0.13** (0.02)
AORD	-0.06 (0.15)	-0.02 (0.12)	-0.07 (0.11)	-0.07 (0.12)	-0.01 (0.11)	0.01 (0.1)	0.03 (0.09)	0.04 (0.07)
BVSP	0.16 (0.1)	0.17* (0.08)	0.22** (0.06)	0.21** (0.04)	0.18** (0.04)	0.17** (0.03)	0.14** (0.02)	0.11** (0.02)
DJI	0.52 (0.4)	0.35 (0.36)	0.29 (0.26)	0.38* (0.16)	0.38** (0.13)	0.34** (0.12)	0.28** (0.1)	0.22** (0.08)
FCHI	0.38 (0.32)	0.1 (0.18)	0.11 (0.21)	0.1 (0.19)	0.08 (0.14)	0.14 (0.11)	0.16* (0.07)	0.17** (0.05)
FTSE	0.54* (0.27)	0.29 (0.2)	0.18 (0.15)	0.16 (0.14)	0.15 (0.11)	0.17 (0.09)	0.19* (0.07)	0.19** (0.06)
GDAXI	0.18 (0.13)	0.05 (0.08)	0.08 (0.09)	0.08 (0.08)	0.06 (0.06)	0.08 (0.04)	0.07** (0.02)	0.08** (0.01)
HSI	-0.02 (0.08)	0.00 (0.07)	-0.02 (0.07)	-0.01 (0.08)	0.01 (0.07)	0.03 (0.07)	0.02 (0.07)	0.02 (0.05)
IXIC	0.92* (0.43)	0.98** (0.33)	0.89* (0.34)	0.76* (0.32)	0.64* (0.26)	0.62* (0.24)	0.44** (0.14)	0.41** (0.11)
KS11	1.04** (0.38)	1.06** (0.37)	0.92* (0.4)	0.83* (0.37)	0.81* (0.34)	0.71* (0.33)	0.51 (0.34)	0.42 (0.25)
MXX	-0.04 (0.1)	-0.03 (0.11)	0.03 (0.11)	0.06 (0.11)	0.09 (0.11)	0.09 (0.11)	0.09 (0.1)	0.09 (0.08)
N225	0.15 (0.18)	0.08 (0.12)	0.03 (0.09)	0.03 (0.07)	0.04 (0.05)	0.06 (0.05)	0.03 (0.05)	0.02 (0.04)
NSEI	0.03 (0.06)	0.00 (0.05)	0.00 (0.06)	0.01 (0.05)	0.02 (0.04)	0.02 (0.04)	0.03 (0.02)	0.03** (0.01)
RUT	0.26 (0.19)	0.19 (0.19)	0.18 (0.17)	0.21 (0.15)	0.2 (0.14)	0.22 (0.11)	0.21* (0.09)	0.17** (0.06)
SPXCASH	0.29 (0.2)	0.13 (0.19)	0.1 (0.17)	0.13 (0.14)	0.12 (0.13)	0.14 (0.1)	0.13* (0.07)	0.12* (0.05)
STOXX50E	0.58 (0.39)	0.21 (0.26)	0.19 (0.23)	0.2 (0.19)	0.17 (0.13)	0.18 (0.1)	0.2** (0.07)	0.18** (0.06)

as a function of the return horizon h as $NW = max\{3, 2 \times h\}$. Yet, Johnson (2019), for example, uses only 12 lags in his research on the VRP.[9] Londono (2015), on the

[9]Note that Johnson (2019) uses 12 Newey-West lags only when estimating the OLS regression of his analysis. He compares this regression to the results from a WLS regression, where he computes standard

Table 8: Variance Risk Premium Regressions - Daily data
Regression results estimated as in Equation 19. The regressions were based on daily data and were estimated for the 1 to 12-month return horizon. The dependent variable is the annualized log return in percentages over the h return horizon in excess of the 3-month US T-bill rate. The table contains only the regression coefficient and the Newey-West standard errors in parentheses. The Newey-West standard errors were set to 21, i.e., the number of trading days in a month, as in Johnson (2019). For the sake of readability, not all return horizons are displayed in the table.*$p < .05$;**$p < .01$.

	Horizon							
	1	2	3	4	5	6	9	12
AEX	0.27 (0.16)	0.17 (0.1)	0.12 (0.09)	0.07 (0.09)	0.09 (0.07)	0.11* (0.05)	0.15** (0.04)	0.14** (0.03)
AORD	-0.14 (0.16)	-0.09 (0.11)	-0.07 (0.09)	-0.04 (0.09)	0.00 (0.07)	0.03 (0.07)	0.05 (0.06)	0.05 (0.05)
BVSP	0.27** (0.08)	0.23** (0.07)	0.22** (0.05)	0.2** (0.05)	0.17** (0.04)	0.14** (0.04)	0.12** (0.03)	0.1** (0.02)
DJI	0.57** (0.22)	0.34* (0.17)	0.31* (0.14)	0.34** (0.09)	0.34** (0.08)	0.31** (0.07)	0.27** (0.06)	0.23** (0.05)
FCHI	0.43* (0.22)	0.21 (0.14)	0.11 (0.14)	0.08 (0.13)	0.12 (0.11)	0.17 (0.09)	0.18** (0.07)	0.15** (0.05)
FTSE	0.52** (0.18)	0.32* (0.14)	0.2 (0.12)	0.19* (0.08)	0.2** (0.07)	0.21** (0.06)	0.21** (0.05)	0.19** (0.04)
GDAXI	0.3** (0.09)	0.2* (0.08)	0.17* (0.08)	0.16* (0.06)	0.15** (0.05)	0.15** (0.03)	0.13** (0.03)	0.13** (0.02)
HSI	0.03 (0.09)	0.03 (0.06)	0.02 (0.05)	0.02 (0.05)	0.04 (0.04)	0.05 (0.04)	0.03 (0.04)	0.02 (0.03)
IXIC	0.97** (0.2)	0.81** (0.13)	0.75** (0.11)	0.65** (0.09)	0.57** (0.07)	0.49** (0.06)	0.41** (0.06)	0.34** (0.05)
KS11	1.09** (0.27)	1.01** (0.2)	0.92** (0.17)	0.82** (0.13)	0.76** (0.12)	0.64** (0.12)	0.48** (0.14)	0.36** (0.11)
MXX	0.03 (0.22)	0.01 (0.14)	0.03 (0.11)	0.08 (0.08)	0.1 (0.07)	0.1 (0.07)	0.1 (0.06)	0.1* (0.05)
N225	0.15 (0.12)	0.1 (0.09)	0.05 (0.09)	0.06 (0.07)	0.09 (0.05)	0.09* (0.04)	0.05 (0.03)	0.03 (0.03)
NSEI	0.03 (0.04)	0.01 (0.03)	0.00 (0.03)	0.01 (0.03)	0.02 (0.02)	0.02 (0.02)	0.03* (0.01)	0.03** (0.01)
RUT	0.31 (0.17)	0.25* (0.12)	0.22* (0.11)	0.21* (0.1)	0.23** (0.08)	0.23** (0.06)	0.21** (0.04)	0.17** (0.03)
SPXCASH	0.33 (0.17)	0.23 (0.12)	0.18 (0.11)	0.16 (0.1)	0.18* (0.08)	0.18** (0.06)	0.16** (0.04)	0.14** (0.03)
STOXX50E	0.59** (0.2)	0.31 (0.16)	0.22 (0.14)	0.21* (0.11)	0.19* (0.08)	0.19** (0.07)	0.19** (0.05)	0.16** (0.04)

contrary, employs a different approach to computing the number of Newey-West lags: $NW = max\{12, 2 \times h\}$, which applies significantly more lags to the shorter return horizons.

errors differently.

Hence, in this sub-section, I want to investigate whether my estimation results are robust to a lower or higher number of lags. Table 9 summarizes the results if specifying the number of lags as in Londono (2015).

We can observe that for most of the equity indices, the different specifications of the number of lags do not impact the standard errors and hence the significance level of the coefficients significantly. In fact, the few changes we observe, e.g., the GDAXI or the STOXX50E, point towards lower standard errors and hence higher significance levels. Similarly, when fixing the number of Newey-West lags to 12 for all return horizons, as in Johnson (2019), the results for the longer-term return horizons are hardly affected. Table 28 in Appendix A.4 contains the results of this final step. Hence, I conclude that the number of lags does not seem to have a large effect on the results of the analysis as neither increasing nor decreasing them resulted in notable changes to the interpretation of the main results.

6.5 Estimation with WLS

[!htb] As a final robustness test, I compare the analysis in Section 5, where I estimated regressions using OLS, to a regression setting where I apply weighted least squares (WLS). This final step of my analysis is based on the findings by Johnson (2019) who shows that when using WLS instead of OLS, the predictive power of the VRP decreases significantly. For WLS, as described in Section 4, one needs to define a vector of weights to put different weights on each observation in the sample. Here, following Johnson (2019), I use the inverse of the forecasted realized volatility over the next month. Hence, I estimate the WLS regression coefficients as in Johnson (2019):

$$\min_{\beta_0,\beta_1} \sum_{t=1}^{T} \left(\frac{r_{t+h} - \beta_0 - VRP_t \times \beta_1}{\hat{\sigma}_t} \right)^2, \tag{20}$$

where r_{t+h} is the return over the h day forecasting horizon, VRP_t is the estimate of the variance risk premium on day t and $\hat{\sigma}_t$ is the square root of the forecasted realized variance over the next month. Johnson (2019) uses the fitted values of a regression of next month' realized variance on lagged monthly and yearly realized variance as his forecast for $\hat{\sigma}_t$. As the first part of this thesis was focused solely on comparing different variance forecasting techniques, I will take advantage of this by using more sophisticated models to predict

Table 9: Variance Risk Premium Regressions - Robustness to different NW lags
Robustness of the analysis in Section 5 to a change in the number of Newey-West lags. The table presents the regression results based on Equation 19. The regressions were based on end-of-month data and were estimated for the 1 to 12-month return horizon. The dependent variable is the log return in percentages over the h return horizon in excess of the 3-month US T-bill rate. The table contains only the regression coefficient and the Newey-West standard errors in parentheses. The Newey-West standard errors were determined as $NW = max\{12, 2 \times h\}$, as in Londono (2015), where h is the forecast horizon in months. For the sake of readability, not all return horizons are displayed in the table.*$p < .05$;**$p < .01$.

	Horizon							
	1	2	3	4	5	6	9	12
AEX	0.21 (0.26)	0.14 (0.2)	0.1 (0.17)	0.08 (0.16)	0.13 (0.13)	0.17 (0.1)	0.25** (0.07)	0.24** (0.06)
AORD	-0.28 (0.21)	-0.2 (0.15)	-0.14 (0.13)	-0.11 (0.14)	-0.08 (0.13)	-0.01 (0.13)	0.00 (0.11)	0.02 (0.08)
BVSP	0.38** (0.06)	0.33** (0.04)	0.29** (0.04)	0.28** (0.04)	0.24** (0.03)	0.2** (0.03)	0.17** (0.03)	0.14** (0.03)
DJI	0.64 (0.34)	0.42 (0.24)	0.41* (0.18)	0.44** (0.13)	0.44** (0.11)	0.42** (0.1)	0.35** (0.09)	0.31** (0.08)
FCHI	0.77 (0.42)	0.45 (0.28)	0.19 (0.26)	0.22 (0.22)	0.38* (0.17)	0.46** (0.15)	0.42** (0.14)	0.35** (0.12)
FTSE	0.00 (0.28)	0.03 (0.17)	0.06 (0.14)	0.11 (0.13)	0.13 (0.12)	0.13 (0.13)	0.17 (0.12)	0.17 (0.1)
GDAXI	0.4 (0.36)	0.38 (0.22)	0.35* (0.16)	0.38** (0.12)	0.38** (0.11)	0.38** (0.09)	0.31** (0.08)	0.31** (0.06)
HSI	0.01 (0.1)	0.00 (0.08)	0.00 (0.08)	0.01 (0.08)	0.03 (0.08)	0.04 (0.08)	0.03 (0.08)	0.01 (0.07)
IXIC	0.29 (0.38)	0.65** (0.16)	0.67** (0.14)	0.64** (0.15)	0.6** (0.15)	0.57** (0.12)	0.44** (0.09)	0.4** (0.07)
KS11	1.04** (0.34)	0.92** (0.32)	0.89** (0.29)	0.86** (0.24)	0.81** (0.24)	0.71** (0.24)	0.53* (0.25)	0.41* (0.19)
MXX	-0.07 (0.09)	-0.1 (0.1)	-0.08 (0.12)	0.00 (0.11)	0.04 (0.11)	0.07 (0.11)	0.08 (0.09)	0.08 (0.08)
N225	0.09 (0.1)	0.08 (0.08)	0.03 (0.08)	0.01 (0.07)	0.04 (0.06)	0.06 (0.05)	0.04 (0.04)	0.02 (0.04)
NSEI	0.03 (0.05)	0.01 (0.04)	0.00 (0.04)	0.01 (0.04)	0.02 (0.03)	0.02 (0.03)	0.03 (0.02)	0.03** (0.01)
RUT	0.29 (0.26)	0.23 (0.2)	0.23 (0.18)	0.24 (0.16)	0.26* (0.13)	0.27** (0.1)	0.23** (0.09)	0.19** (0.07)
SPXCASH	0.3 (0.3)	0.22 (0.22)	0.2 (0.19)	0.21 (0.17)	0.22 (0.13)	0.22* (0.1)	0.18** (0.06)	0.16** (0.05)
STOXX50E	0.78* (0.32)	0.54** (0.18)	0.38** (0.14)	0.35** (0.13)	0.33* (0.14)	0.29* (0.14)	0.21 (0.14)	0.19 (0.12)

future variance. Hence, as for the computation of the VRP, I use the best-performing forecasting model for each index (see Table 2) to derive an estimate for $\hat{\sigma}_t$, i.e., to define the weight vector for the WLS regression. Furthermore, it is important to emphasize that I base this analysis on daily data in order to compare my findings with the ones reported by Johnson (2019). Table 10 summarizes the results of the WLS regressions for the different return horizons.

Given that I used daily data for the WLS estimation and that the end-of-month results, as reported in Section 5, were relatively robust to using daily data, I will compare the estimates from the WLS regressions to the results of the daily OLS regressions as presented in Table 8. Overall, we observe that the results do not change significantly, when using WLS. Most importantly, when using WLS, I still fail to reject the null hypothesis of no-predictability for the markets mentioned above: AORD, HSI, MXX and N225. Furthermore, the already small coefficients for the NSEI decrease even further and stop being statistically different from zero when using WLS. For the remaining indices, however, the picture does not change significantly when using WLS, which is at odds with the findings reported by Johnson (2019). For several markets the coefficients increase when using WLS (see for example AEX, BVSP and RUT). Another pattern I observe is that for some markets the coefficients for the short-term return predictions increase, while they decrease for the long-term predictions (see for example KS11, SPXCASH and STOXX50E). Yet, in general, the results seem to be robust to applying WLS instead of OLS.

7 Conclusion

In the second part of this thesis, I was able to show that different variance forecasting techniques can be used to derive the VRP for a range of international equity markets. I showed that the VRP could, in turn, be used to predict future stock returns in the US markets and most of the European markets. The predictability for non-US and non-European markets, however, was mixed and varied heavily between markets. For most of the US markets, my results confirm the "hump"-pattern reported by Bollerslev et al. (2014) and Bekaert and Hoerova (2014) where predictability is highest for medium-term returns and lower for short- or long-term returns. The robustness checks in Section 6 revealed that the results are sensitive to the construction of the sample, especially when

Table 10: Variance Risk Premium Regressions - Estimation with WLS
Robustness of the analysis when estimating return regressions using weighted least squares (WLS). The table presents the regression results based on Equation 20. The regressions were based on daily data and were estimated for the 1 to 12-month return horizon, i.e., 21 to 252 days. The dependent variable is the log return in percentages over the h return horizon in excess of the 3-month US T-bill rate. The table contains only the regression coefficient and the Newey-West standard errors in parentheses. The Newey-West standard errors were set to 21, i.e., the number of trading days in a month, as in Johnson (2019). For the sake of readability, not all return horizons are displayed in the table.*$p < .05$;**$p < .01$.

	Horizon							
	1	2	3	4	5	6	9	12
AEX	0.49** (0.19)	0.27* (0.13)	0.18 (0.11)	0.12 (0.1)	0.11 (0.08)	0.12 (0.06)	0.12* (0.05)	0.1* (0.05)
AORD	0.09 (0.19)	0.08 (0.15)	0.09 (0.12)	0.08 (0.11)	0.05 (0.11)	0.05 (0.1)	0.02 (0.09)	0.04 (0.06)
BVSP	0.42** (0.11)	0.37** (0.09)	0.34** (0.07)	0.32** (0.06)	0.27** (0.04)	0.24** (0.03)	0.17** (0.03)	0.14** (0.02)
DJI	0.84** (0.24)	0.51** (0.16)	0.51** (0.11)	0.44** (0.09)	0.37** (0.08)	0.32** (0.08)	0.24** (0.07)	0.19** (0.06)
FCHI	0.84** (0.27)	0.45* (0.19)	0.3 (0.17)	0.24 (0.15)	0.23* (0.12)	0.27** (0.1)	0.2** (0.07)	0.13* (0.06)
FTSE	0.81** (0.21)	0.52** (0.16)	0.37** (0.12)	0.31** (0.1)	0.25** (0.08)	0.24** (0.08)	0.19** (0.07)	0.15* (0.07)
GDAXI	0.54** (0.15)	0.37** (0.12)	0.31** (0.08)	0.27** (0.07)	0.21** (0.06)	0.21** (0.05)	0.17** (0.04)	0.18** (0.03)
HSI	0.06 (0.1)	0.03 (0.08)	0.01 (0.07)	0.01 (0.06)	0.00 (0.06)	-0.01 (0.05)	-0.03 (0.05)	-0.02 (0.03)
IXIC	0.68* (0.31)	0.62** (0.21)	0.61** (0.15)	0.55** (0.12)	0.46** (0.09)	0.43** (0.09)	0.32** (0.08)	0.36** (0.07)
KS11	1.15** (0.38)	1.07** (0.27)	0.96** (0.22)	0.8** (0.18)	0.72** (0.17)	0.63** (0.17)	0.42** (0.14)	0.3* (0.12)
MXX	-0.1 (0.2)	-0.03 (0.17)	0.05 (0.14)	0.08 (0.13)	0.06 (0.12)	0.03 (0.12)	0.02 (0.1)	0.04 (0.07)
N225	0.3 (0.17)	0.18 (0.12)	0.13 (0.09)	0.08 (0.08)	0.06 (0.07)	0.06 (0.07)	0.00 (0.06)	-0.02 (0.05)
NSEI	0.04 (0.04)	0.02 (0.04)	0.01 (0.04)	0.01 (0.03)	0.02 (0.03)	0.01 (0.02)	0.02 (0.02)	0.02 (0.01)
RUT	0.57** (0.17)	0.45** (0.13)	0.46** (0.11)	0.4** (0.09)	0.34** (0.08)	0.31** (0.06)	0.21** (0.05)	0.14** (0.04)
SPXCASH	0.65** (0.17)	0.39** (0.13)	0.35** (0.1)	0.29** (0.09)	0.21** (0.08)	0.16* (0.07)	0.09 (0.06)	0.08 (0.05)
STOXX50E	0.88** (0.22)	0.49** (0.18)	0.33* (0.13)	0.26* (0.1)	0.17 (0.09)	0.15 (0.08)	0.09 (0.07)	0.06 (0.06)

working with monthly data. Using daily data or estimating the forecasting regressions with WLS instead of OLS, however, did not materially change my results. This is an important finding given that most existing research uses end-of-month data to compute monthly returns.

The main contributions of my thesis are threefold: first, I extended the comparison of the predictive power of the VRP to a larger sample of international markets. Second, I extended the variance forecasting exercise of Bekaert and Hoerova (2014) by testing a larger number of potential models, including several machine learning models, and applying them to different equity markets. Third, I conducted several robustness checks regarding the construction of the sample, the computation of standard errors, and the use of WLS. Further research on the VRP should consider the differences between, for example, end-of-month and middle-of-the-month data to secure robust estimation results.

Finally, the results presented in this thesis, suggest several fields of future research. First, one could select the forecasting technique based only on the performance prior to the prediction. This would ensure a truly out-of-sample prediction of realized variance, as I selected the best-performing forecasting model, in line with Bekaert and Hoerova (2014), based on its performance over the entire sample. Second, future research could add other well-known factors from the literature to the analysis to see if this improves the return predictions, especially in those markets where the VRP could not significantly predict returns. Third, and directly related to that, more research seems to be needed on where the differences between stock markets come from. There may be other ways to capture market risk aversion if the VRP cannot explain excess returns because economically one would expect the VRP to be present across international markets. Fourth, I could show that my analysis was relatively robust to using WLS instead of OLS. This, however, stands in contrast to the findings reported by Johnson (2019), who finds that using WLS significantly lowers the predictive power of the VRP. Hence, more research is needed on the effect of the estimation technique on the magnitude of the VRP.

References

Bali, T. G. and Zhou, H. (2016). Risk, uncertainty, and expected returns. *Journal of Financial and Quantitative Analysis*, 51(3):707–735.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, 2(1):1–37.

Bartz, E., Bartz-Beielstein, T., Zaefferer, M., and Mersmann, O. (2023). *Hyperparameter Tuning for Machine and Deep Learning with R: A Practical Guide*. Springer Nature.

Bekaert, G. and Hoerova, M. (2014). The vix, the variance premium and stock market volatility. *Journal of econometrics*, 183(2):181–192.

Bollerslev, T., Marrone, J., Xu, L., and Zhou, H. (2014). Stock return predictability and variance risk premia: Statistical inference and international evidence. *Journal of Financial and Quantitative Analysis*, 49(3):633–661.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18.

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

Bollerslev, T., Xu, L., and Zhou, H. (2015). Stock return and cash flow predictability: The role of volatility risk. *Journal of Econometrics*, 187(2):458–471.

Cheng, I.-H. (2019). The vix premium. *The Review of Financial Studies*, 32(1):180–227.

Christensen, K., Siggaard, M., and Veliyev, B. (2023). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, 21(5):1680–1727.

Clements, A. and Preve, D. P. (2021). A practical guide to harnessing the har volatility model. *Journal of Banking & Finance*, 133:106285.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F. and Reno, R. (2009). Har volatility modelling with heterogeneous leverage and jumps. *Available at SSRN*, 1316953.

Drechsler, I. and Yaron, A. (2011). What's vol got to do with it. *The Review of Financial Studies*, 24(1):1–45.

Fassas, A. P. and Papadamou, S. (2018). Variance risk premium and equity returns. *Research in International Business and Finance*, 46:462–470.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* sage publications, Inc.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101.

Johnson, T. L. (2019). A fresh look at return predictability using a more efficient estimator. *The Review of Asset Pricing Studies*, 9(1):1–46.

Li, S. Z. and Tang, Y. (2022). Automated risk forecasting. *Available at SSRN 3776915*.

Londono, J. M. (2015). The variance risk premium around the world. *Available at SSRN 2517020*.

Luong, C. and Dokuchaev, N. (2018). Forecasting of realised volatility with the random forests algorithm. *Journal of Risk and Financial Management*, 11(4):61.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society*, pages 703–708.

Probst, P., Wright, M. N., and Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, 9(3):e1301.

Qiao, F., Xu, L., Zhang, X., and Zhou, H. (2019). Variance risk premiums in emerging markets. *Available at SSRN 3199522*.

Rahimikia, E. and Poon, S.-H. (2020). Alternative data for realised volatility forecasting: Limit order book and news stories. *Available at SSRN*, 3684040.

Wang, H., Zhou, H., and Zhou, Y. (2013). Credit default swap spreads and variance risk premia. *Journal of Banking & Finance*, 37(10):3733–3746.

Witten, D. and James, G. (2013). *An introduction to statistical learning with applications in R*. springer publication.

Zhou, H. (2018). Variance risk premia, asset predictability puzzles, and macroeconomic uncertainty. *Annual Review of Financial Economics*, 10:481–497.

A Appendix

A.1 Monthly realized variance forecast results

Table 11: MSE - Monthly realized variance forecasts
The table displays the mean squared error (MSE) across nine models for all 17 stock market indices forecasting monthly realized variance. The statistics were divided by the result for the benchmark HAR model for standardization. For each row, the value with the lowest error is printed in bold and the value with the second lowest error is printed in italics. The full names of the indices can be found in Table 2. Results for daily or weekly as well as results based on QLIKE and MAPE loss functions can be found in the first part of this thesis.

	HAR	C-HAR	HAR-J	WLS-HAR	RR-HAR	SVM	RF	AR-1	LSTM
AEX	1.00	*0.954*	0.986	1.129	0.977	1.61	**0.943**	1.11	1.65
AORD	1.00	0.987	0.988	0.964	**0.932**	1.414	*0.934*	0.945	1.164
BVSP	1.00	0.96	0.979	**0.91**	*0.933*	1.764	1.19	1.057	1.771
DJI	*1.00*	1.051	1.01	1.022	**0.939**	1.743	1.154	1.111	1.584
FCHI	1.00	*0.984*	1.011	1.147	0.994	1.634	**0.968**	1.157	1.573
FTSE	1.00	**0.97**	0.994	1.106	*0.987*	1.616	1.013	1.184	1.442
GDAXI	1.00	1.036	*0.999*	1.076	**0.959**	1.603	1.006	1.253	1.644
HSI	1.00	1.021	1.003	1.00	**0.888**	1.885	*0.95*	1.198	1.489
IXIC	1.00	*0.954*	1.004	1.053	0.957	1.446	**0.918**	1.16	1.402
KS11	1.00	1.031	1.015	1.00	**0.88**	1.414	*0.974*	1.053	1.391
MXX	1.00	**0.858**	0.939	0.958	*0.91*	1.548	1.00	1.13	1.309
N225	1.00	*0.993*	1.00	1.06	**0.967**	1.608	1.114	1.145	1.57
NSEI	1.00	*0.865*	1.672	1.083	0.948	1.193	**0.837**	1.215	1.055
RUT	*1.00*	1.03	1.006	1.129	**0.98**	1.836	1.03	1.062	1.657
SPX	*1.00*	1.037	1.012	1.016	**0.924**	2.012	1.099	1.064	2
STOXX50E	1.00	**0.982**	0.989	1.691	0.983	1.467	0.989	1.156	1.521

A.2 Historical evolution of VRP

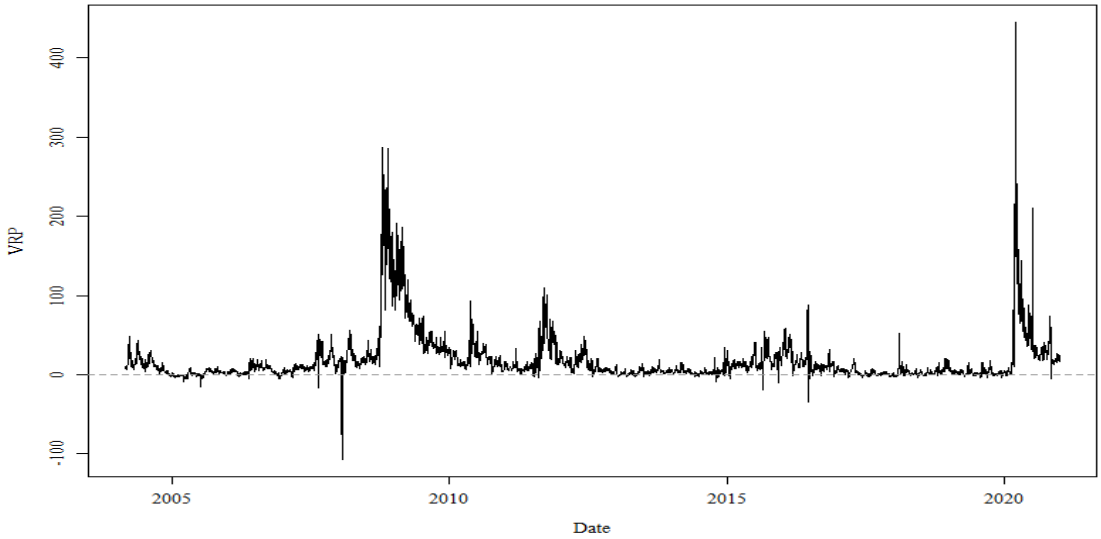

Figure 6: Variance Risk Premium - AEX (Netherlands)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

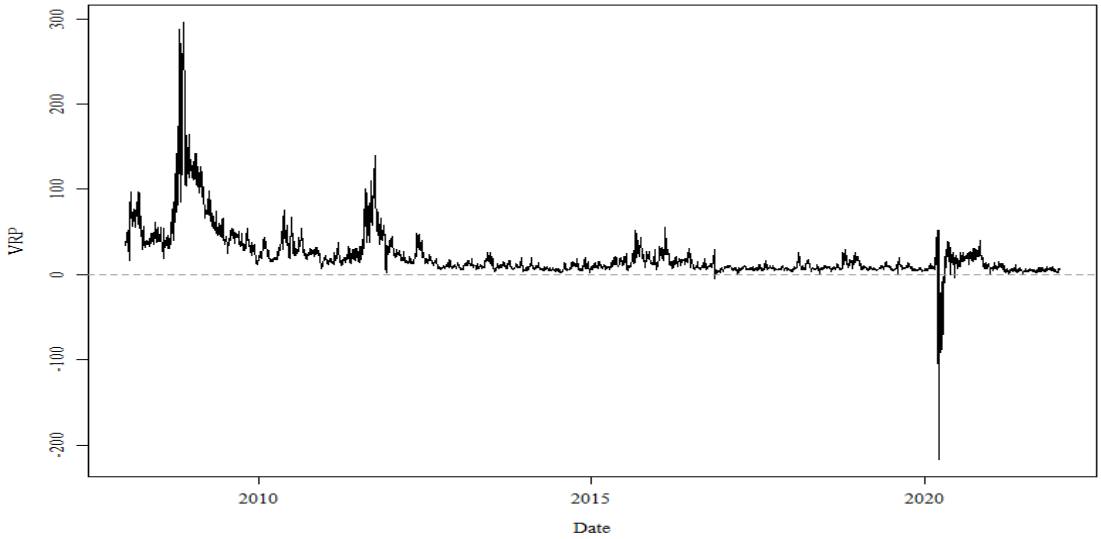

Figure 7: Variance Risk Premium - AORD (Australia)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

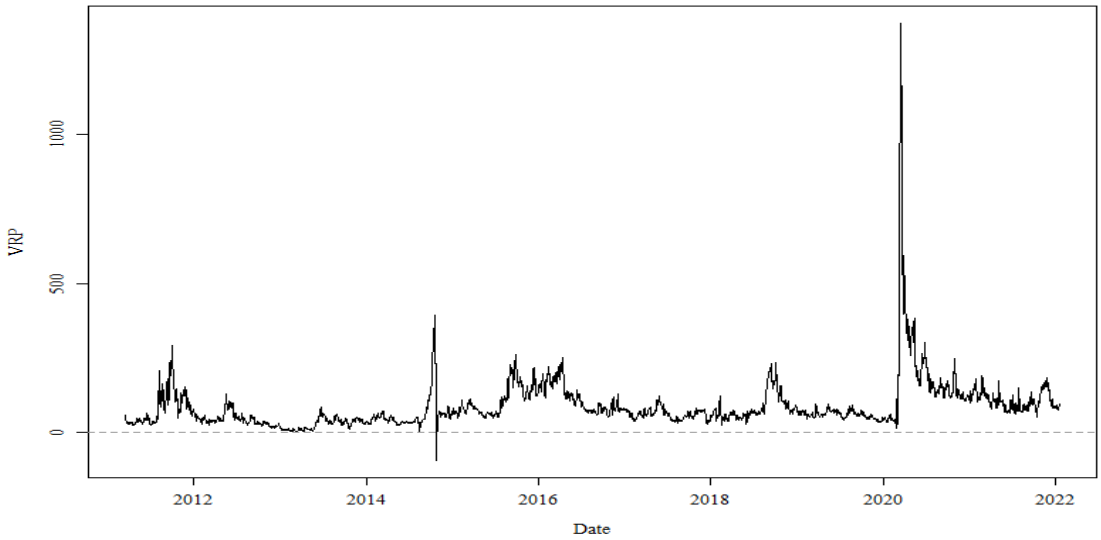

Figure 8: Variance Risk Premium - BVSP (Brazil)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

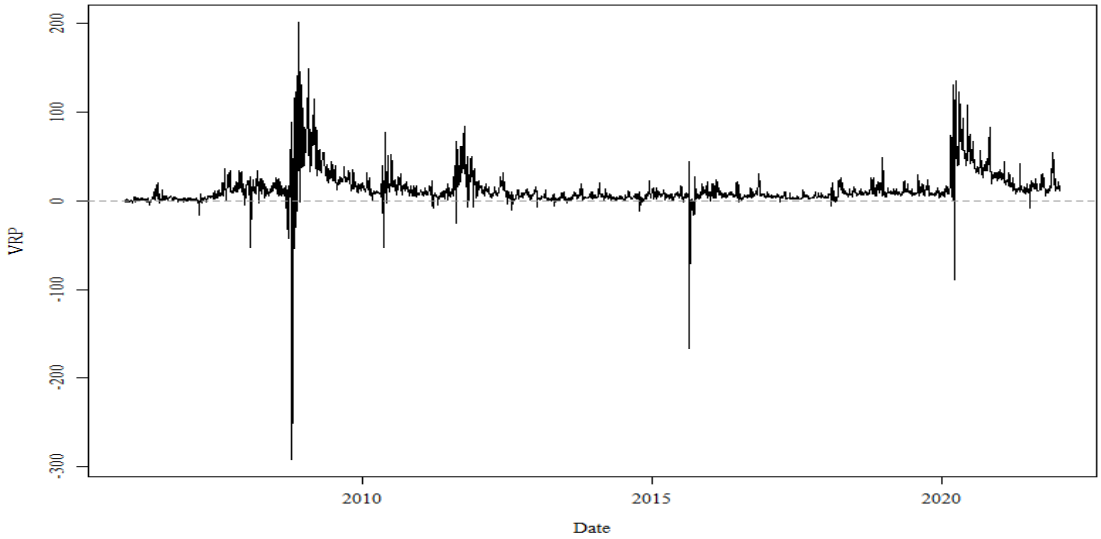

Figure 9: Variance Risk Premium - DJI (USA)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

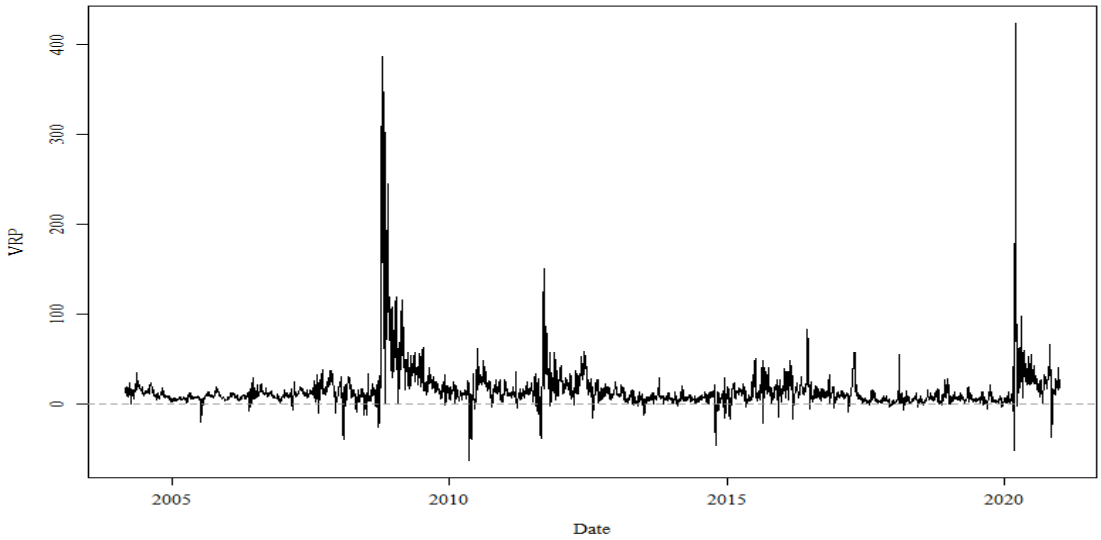

Figure 10: Variance Risk Premium - FCHI (France)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

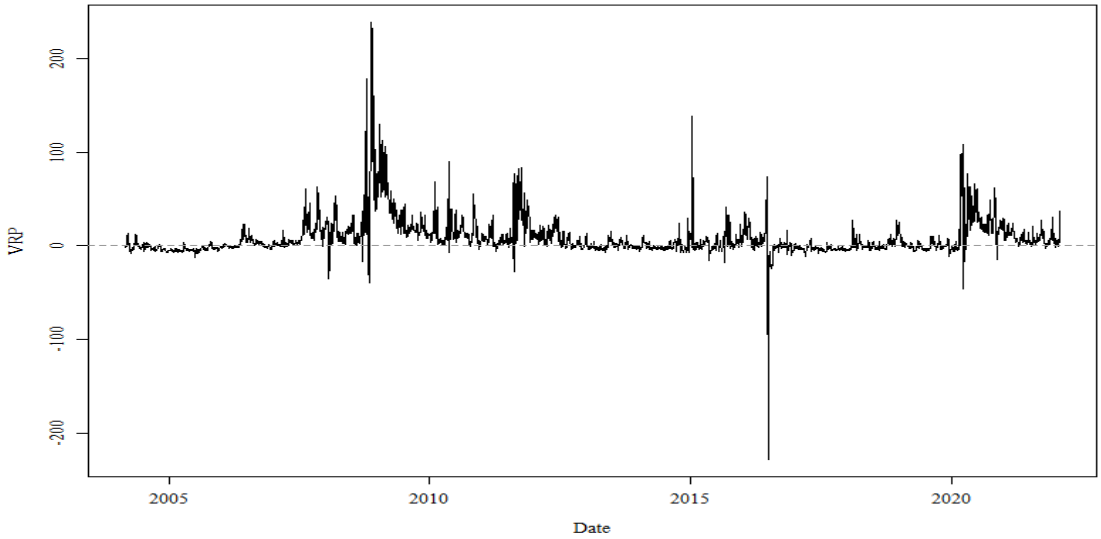

Figure 11: Variance Risk Premium - FTSE (UK)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

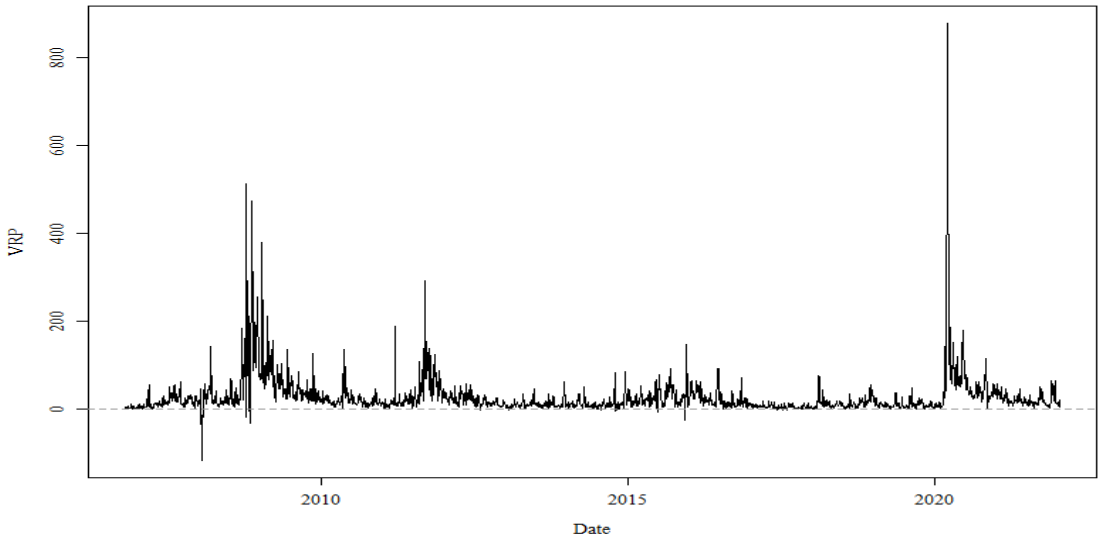

Figure 12: Variance Risk Premium - GDAXI (Germany)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

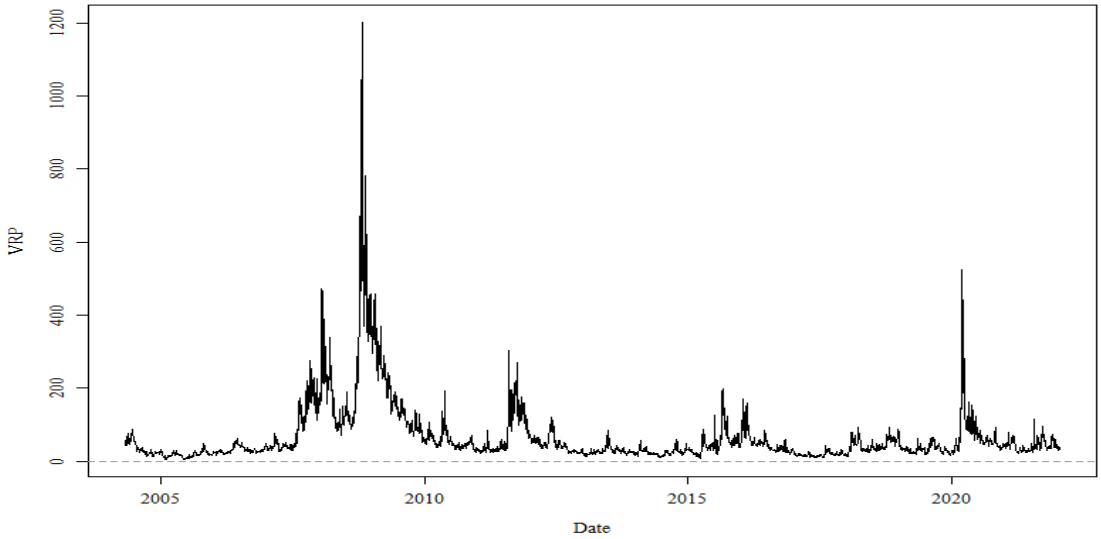

Figure 13: Variance Risk Premium - HSI (Hongkong)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

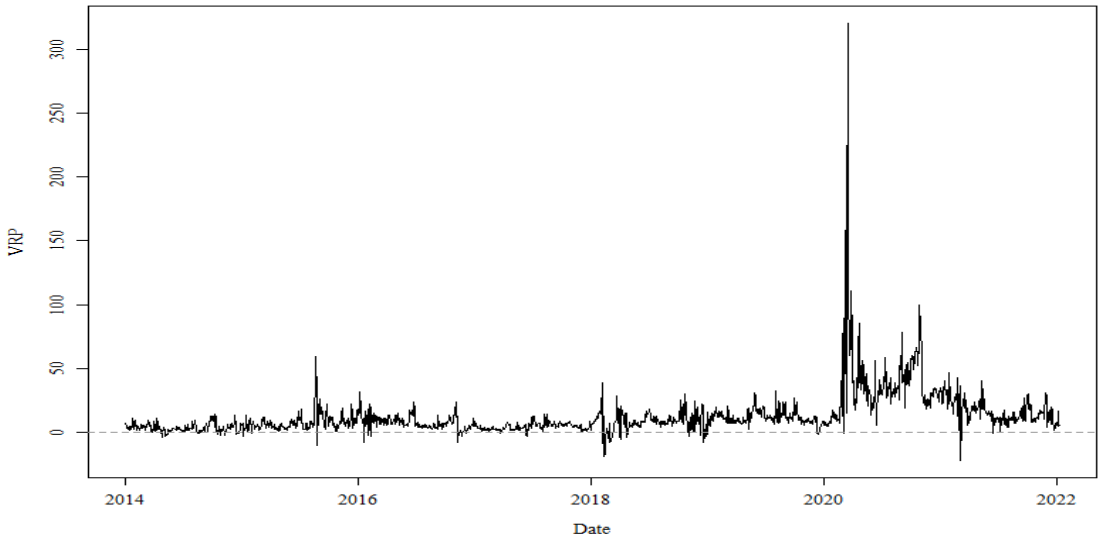

Figure 14: Variance Risk Premium - IXIC (USA)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

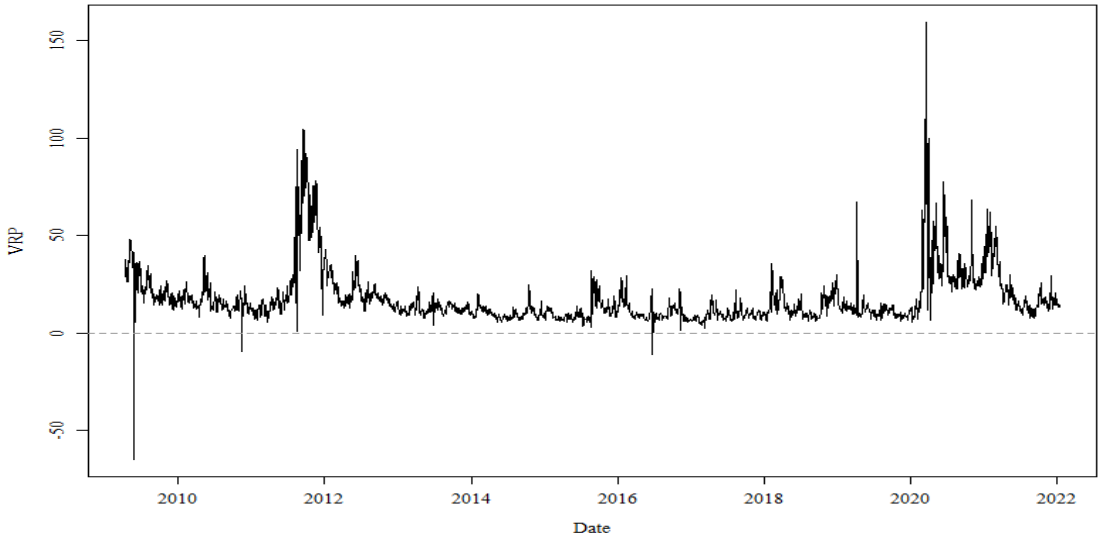

Figure 15: Variance Risk Premium - KS11 (South Korea)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

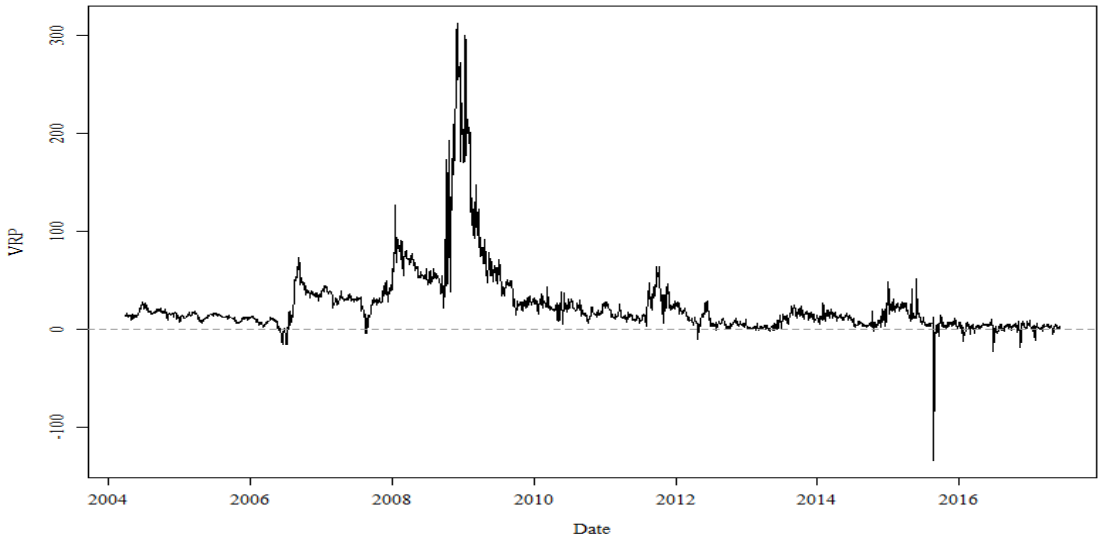

Figure 16: Variance Risk Premium - MXX (Mexico)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

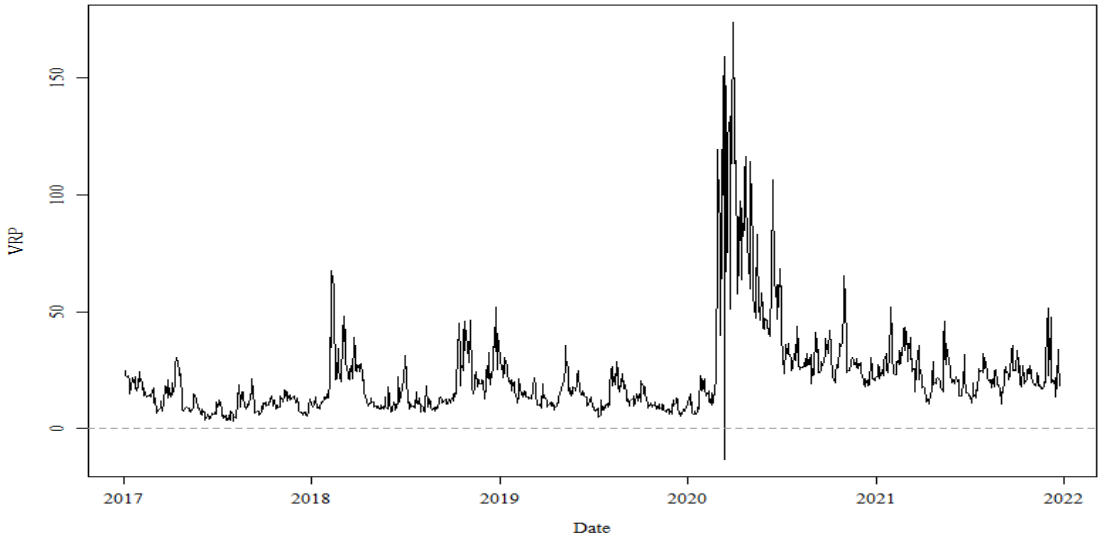

Figure 17: Variance Risk Premium - N225 (Japan)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

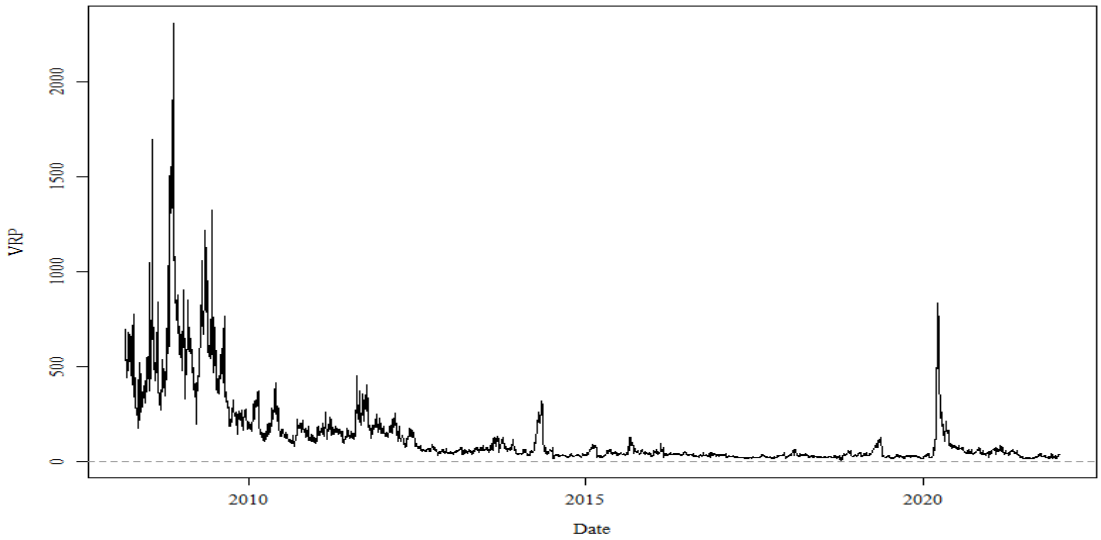

Figure 18: Variance Risk Premium - NSEI (India)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

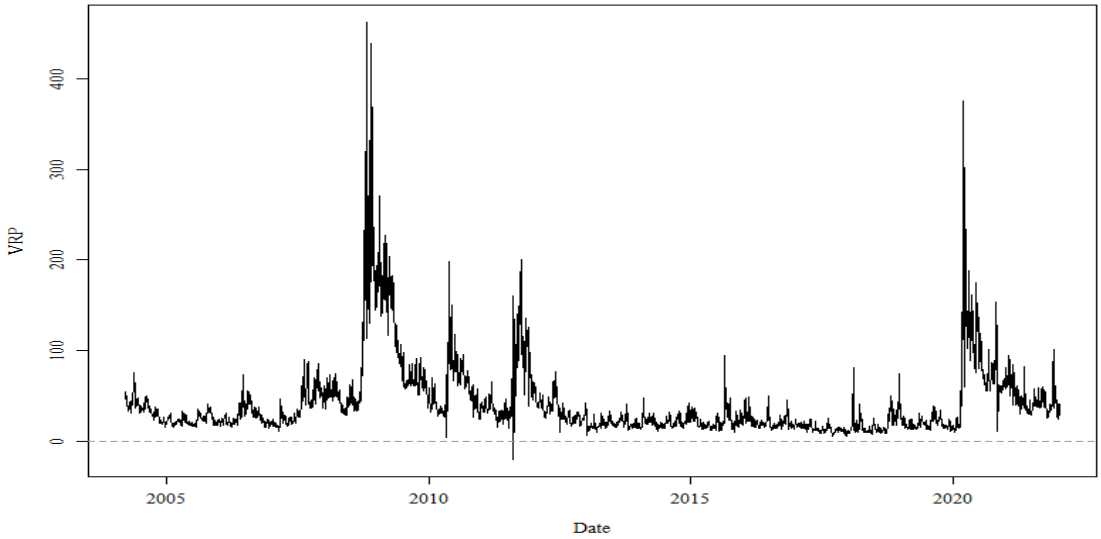

Figure 19: Variance Risk Premium - RUT (USA)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

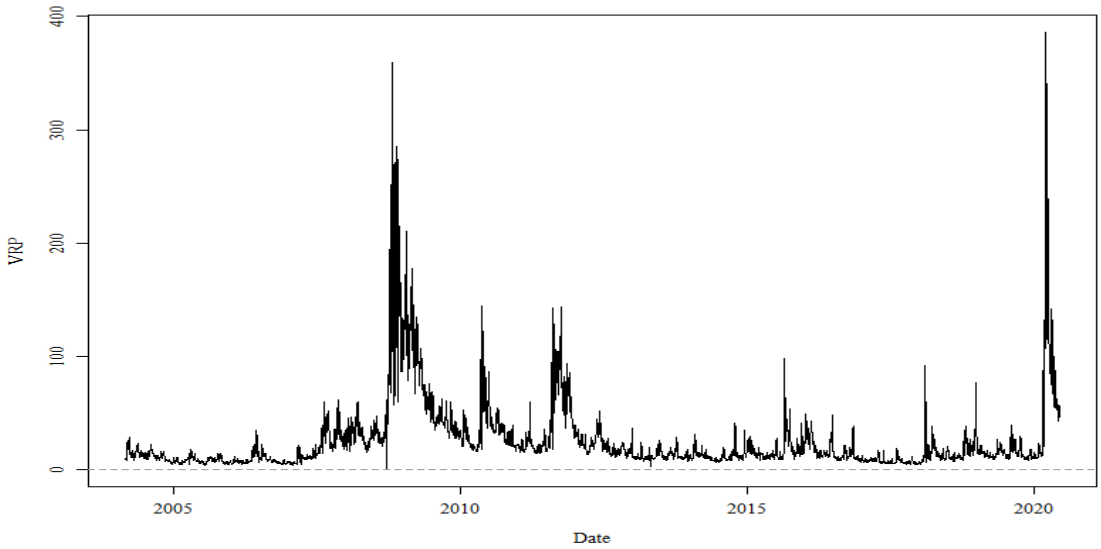

Figure 20: Variance Risk Premium - SPX (USA)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

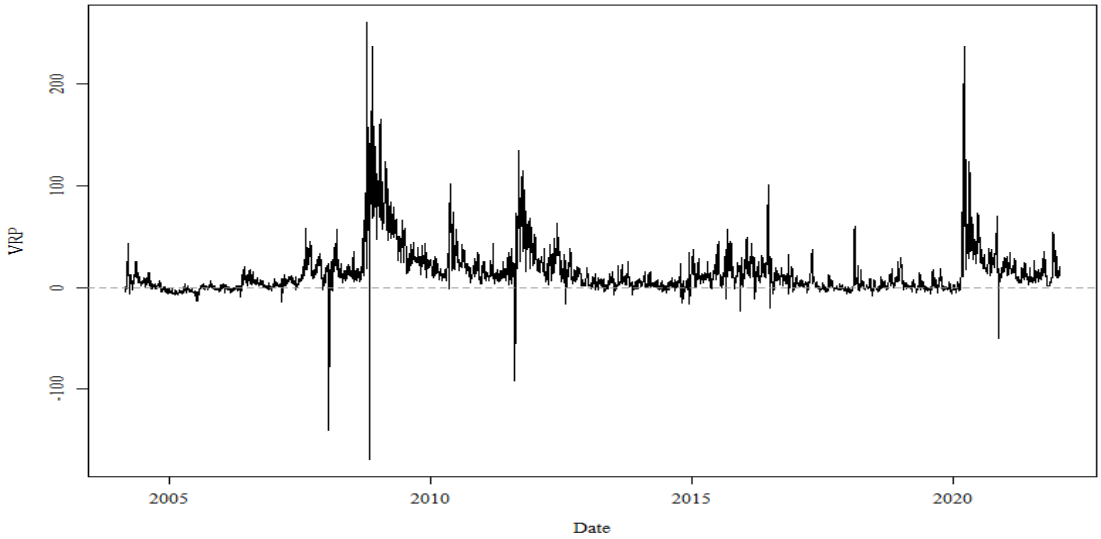

Figure 21: Variance Risk Premium - STOXX50E (Europe)
The figure illustrates the historical evolution of the variance risk premium. The variance risk premium is defined as the difference between the implied volatility taken from the volatility index for the respective market and the realized variance forecast generated by the best-performing model for the same market. The full name of the stock market index as well as the best-performing forecasting model for each index can be found in Table 2.

A.3 Detailed regression results

The following tables display the results of the regression analysis based on Equation 19 for all the $h = 1, ..., 12$ return horizons in months. The summary of these regressions was presented in section 5. This Appendix displays all return horizons, as well as the R^2, the magnitude of the constant, and the standard error of the constant. All regressions were based on end-of-month data. The dependent variable is the annualized log return in percentages over the h return horizon in excess of the 3-month US T-bill rate. I computed Newey-West standard errors, where the number of lags was determined as $NW = max\{3, 2 \times h\}$, precisely as in the main text. Note, also, that for all regressions:*$p <$ .05;**$p <$ .01

Table 12: AEX Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-15.57**	5.59	0.21	0.28	0.67
2	-6.64	4.90	0.14	0.25	0.62
3	-3.07	4.44	0.1	0.20	0.43
4	-1.24	4.13	0.08	0.17	0.32
5	-1.21	4.36	0.13	0.13	1.02
6	-1.47	4.69	0.17	0.10	2.20
7	-1.18	4.75	0.19*	0.08	2.84
8	-1.32	4.92	0.21**	0.07	4.31
9	-1.74	5.16	0.25**	0.07	6.53
10	-1.57	5.13	0.26**	0.06	7.47
11	-1.26	5.05	0.25**	0.06	7.58
12	-0.94	4.92	0.24**	0.06	7.41

Table 13: AORD Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	1.01	4.73	-0.28	0.22	1.96
2	2.70	3.81	-0.20	0.18	2.00
3	2.74	3.34	-0.14	0.16	1.51
4	2.66	3.13	-0.11	0.15	1.17
5	2.27	3.01	-0.08	0.14	0.71
6	1.08	2.99	-0.01	0.13	0.03
7	0.87	2.91	0.01	0.12	0.01
8	1.32	2.87	-0.01	0.12	0.01
9	1.26	2.78	0.00	0.11	0.00
10	1.28	2.65	0.01	0.10	0.05
11	1.40	2.49	0.02	0.09	0.14
12	1.69	2.28	0.02	0.08	0.13

Table 14: BVSP Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-34.49**	8.60	0.38**	0.07	9.13
2	-26.25**	7.83	0.33**	0.05	12.21
3	-21.84**	6.74	0.29**	0.04	15.83
4	-19.74**	5.53	0.28**	0.04	20.80
5	-15.93**	4.89	0.24**	0.03	22.10
6	-12.48*	4.91	0.2**	0.03	21.23
7	-11.3*	4.72	0.19**	0.03	23.54
8	-9.92*	4.61	0.18**	0.03	23.88
9	-8.99*	4.40	0.17**	0.03	25.22
10	-8.03*	3.98	0.16**	0.03	27.01
11	-7	3.92	0.14**	0.03	27.01
12	-6.46	4.10	0.14**	0.03	26.82

Table 15: DJI Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-14.63*	5.95	0.64	0.39	3.81
2	-5.16	4.19	0.42	0.25	3.32
3	-2.85	3.76	0.41*	0.18	5.11
4	-2.25	3.80	0.44**	0.12	7.89
5	-1.45	4.06	0.44**	0.11	9.44
6	-0.65	4.11	0.42**	0.10	10.07
7	0.39	4.18	0.36**	0.10	8.61
8	0.86	4.29	0.35**	0.10	8.71
9	1.03	4.40	0.35**	0.09	9.72
10	1.53	4.34	0.32**	0.09	9.21
11	1.91	4.26	0.3**	0.09	9.11
12	1.89	4.24	0.31**	0.08	10.17

Table 16: FCHI Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-23.72**	6.75	0.77	0.52	2.72
2	-11.4*	5.37	0.45	0.38	1.81
3	-5.01	5.00	0.19	0.32	0.48
4	-4.1	4.66	0.22	0.25	0.88
5	-5.39	4.85	0.38*	0.17	3.13
6	-5.91	5.10	0.46**	0.15	5.33
7	-4.88	5.04	0.42**	0.14	5.05
8	-4.17	5.08	0.39**	0.13	4.93
9	-4.32	5.23	0.42**	0.14	6.34
10	-4.26	5.13	0.43**	0.13	7.28
11	-3.42	4.91	0.38**	0.12	6.30
12	-2.9	4.80	0.35**	0.12	5.92

Table 17: FTSE Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-12.12**	3.87	-0.00	0.31	0.00
2	-5.06	3.03	0.03	0.21	0.03
3	-2.86	2.97	0.06	0.16	0.15
4	-2.06	3.01	0.11	0.14	0.68
5	-1.3	3.01	0.13	0.13	1.03
6	-0.85	3.16	0.13	0.13	1.38
7	-0.56	3.16	0.13	0.12	1.56
8	-0.32	3.22	0.15	0.12	2.07
9	-0.4	3.25	0.17	0.12	3.05
10	-0.37	3.23	0.18	0.11	3.87
11	-0.14	3.17	0.17	0.11	3.63
12	-0.13	3.11	0.17	0.10	4.38

Table 18: GDAXI Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-13.14	8.69	0.4	0.42	1.61
2	-7.84	6.16	0.38	0.27	2.82
3	-5.40	5.54	0.35	0.19	3.84
4	-5.22	5.33	0.38**	0.13	6.15
5	-4.79	5.43	0.38**	0.11	7.65
6	-4.39	5.56	0.38**	0.09	8.97
7	-3.41	5.59	0.34**	0.09	8.88
8	-2.42	5.54	0.3**	0.09	8.06
9	-2.46	5.39	0.31**	0.08	9.38
10	-2.66	5.33	0.32**	0.07	11.25
11	-2.45	5.07	0.31**	0.06	11.96
12	-2.35	4.93	0.31**	0.06	13.00

Table 19: HSI Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-11.21	6.56	0.01	0.12	0.01
2	-3.52	5.19	0.00	0.10	0.00
3	-1.18	4.97	0.00	0.09	0.00
4	-0.80	4.94	0.01	0.09	0.06
5	-0.86	5.11	0.03	0.09	0.29
6	-1.32	5.21	0.04	0.08	0.86
7	-1.06	5.36	0.04	0.08	1.06
8	-0.42	5.24	0.04	0.08	0.91
9	0.30	5.17	0.03	0.08	0.70
10	0.99	5.01	0.02	0.07	0.43
11	1.46	4.92	0.02	0.07	0.31
12	1.83	4.77	0.01	0.07	0.23

Table 20: IXIC Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	3.16	7.33	0.29	0.55	0.52
2	2.12	4.71	0.65**	0.20	5.94
3	3.88	3.96	0.67**	0.14	11.09
4	5.58	3.81	0.64**	0.16	15.65
5	6.72	3.89	0.6**	0.15	19.99
6	7.46*	3.65	0.57**	0.12	23.04
7	8.93*	3.62	0.49**	0.09	20.56
8	9.53**	3.55	0.45**	0.08	19.87
9	9.94**	3.33	0.44**	0.09	19.86
10	9.85**	3.09	0.45**	0.08	22.96
11	10.38**	3.03	0.42**	0.08	21.23
12	10.6**	2.84	0.4**	0.07	21.71

Table 21: KS11 Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-18.85*	7.50	1.04*	0.42	6.24
2	-13.95*	5.90	0.92**	0.31	10.51
3	-12.38*	5.43	0.89**	0.27	15.23
4	-11.53*	5.08	0.86**	0.23	18.89
5	-10.56*	4.98	0.81**	0.23	21.39
6	-8.9	4.88	0.71**	0.24	19.30
7	-6.86	4.75	0.59*	0.24	15.77
8	-5.87	4.59	0.54*	0.24	14.15
9	-5.96	4.73	0.53*	0.25	14.88
10	-5.24	4.57	0.49*	0.24	13.63
11	-4.46	4.20	0.44*	0.22	12.32
12	-3.98	3.79	0.41*	0.19	11.34

Table 22: MXX Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-1.21	5.16	-0.07	0.13	0.18
2	7.26	3.88	-0.10	0.12	0.75
3	9.27**	3.48	-0.08	0.13	0.72
4	8.59*	3.36	-0.00	0.13	0.00
5	8.28*	3.70	0.04	0.11	0.22
6	7.85	4.11	0.07	0.11	0.82
7	8.04	4.32	0.07	0.10	1.00
8	7.97	4.56	0.08	0.10	1.42
9	8.01	4.70	0.08	0.09	1.53
10	7.85	4.86	0.08	0.09	1.93
11	7.93	4.96	0.08	0.08	1.94
12	7.9	5.07	0.08	0.08	2.11

Table 23: N225 Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-12.28*	6.05	0.09	0.12	0.31
2	-4.76	5.47	0.08	0.10	0.47
3	-0.7	5.27	0.03	0.09	0.09
4	1.17	5.16	0.01	0.08	0.01
5	1	5.25	0.04	0.06	0.28
6	0.93	5.44	0.06	0.05	0.74
7	1.5	5.53	0.06	0.05	0.73
8	2.11	5.60	0.05	0.05	0.55
9	2.31	5.67	0.04	0.04	0.60
10	2.74	5.70	0.04	0.04	0.45
11	3.21	5.72	0.03	0.04	0.29
12	3.64	5.72	0.02	0.04	0.15

Table 24: NSEI Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-0.04	6.55	0.03	0.06	0.42
2	4.72	5.03	0.01	0.04	0.12
3	7.15	4.85	0.00	0.05	0.00
4	7.22	4.71	0.01	0.04	0.09
5	6.28	4.62	0.02	0.04	0.85
6	5.82	4.59	0.02	0.03	1.90
7	5.7	4.50	0.03	0.03	3.10
8	6.22	4.16	0.03	0.02	3.97
9	6.17	3.89	0.03	0.02	6.19
10	6.39	3.70	0.03*	0.01	7.76
11	6.7	3.45	0.03**	0.01	8.93
12	7.07*	3.27	0.03**	0.01	10.20

Table 25: RUT Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-19.53*	9.26	0.29	0.26	2.09
2	-9.57	7.35	0.23	0.21	2.56
3	-7.18	6.29	0.23	0.19	4.20
4	-6.06	5.29	0.24	0.16	6.03
5	-6.3	4.74	0.26*	0.13	8.34
6	-6.37	4.65	0.27**	0.10	10.59
7	-5.03	4.52	0.24**	0.09	10.40
8	-4.66	4.72	0.23**	0.09	11.50
9	-4.5	4.81	0.23**	0.09	12.27
10	-3.63	4.69	0.21**	0.08	11.92
11	-3.07	4.50	0.2**	0.08	11.84
12	-2.61	4.41	0.19**	0.07	11.52

Table 26: SPXCASH Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-17.03*	6.94	0.3	0.31	2.07
2	-7	5.13	0.22	0.25	2.18
3	-3.64	4.11	0.2	0.21	2.72
4	-2.48	3.66	0.21	0.18	3.85
5	-2.05	3.53	0.22	0.14	5.23
6	-1.28	3.53	0.22*	0.10	5.84
7	-0.27	3.58	0.19*	0.08	5.62
8	0.27	3.71	0.19*	0.07	5.93
9	0.61	3.82	0.18**	0.06	6.39
10	1.19	3.83	0.17**	0.06	6.07
11	1.5	3.75	0.17**	0.05	6.32
12	1.74	3.75	0.16**	0.05	6.66

Table 27: STOXX50E Regression - All horizons

Return horizon h	Constant	Std.Err (Constant)	VRP	Std.Err(VRP)	R^2
1	-22.5**	5.53	0.78*	0.37	4.94
2	-12.22**	4.28	0.54*	0.23	4.94
3	-7.68	4.19	0.38*	0.16	3.77
4	-6.47	4.62	0.35*	0.14	4.22
5	-5.35	4.90	0.33*	0.14	4.29
6	-4.42	5.11	0.29*	0.14	3.95
7	-3.6	5.11	0.24	0.14	3.12
8	-2.92	5.18	0.22	0.14	3.03
9	-2.78	5.10	0.21	0.14	2.98
10	-2.7	5.09	0.21	0.13	3.25
11	-2.7	4.92	0.2	0.12	3.27
12	-2.51	4.84	0.19	0.12	3.17

A.4 Robustness to Newey-West lags

Table 28: Variance Risk Premium regressions - robustness to different NW lags
Table 28 investigates the robustness of the analysis in section 5 to a change in the number of Newey-West lags. The table presents the regression results based on Equation 19. The regressions were based on end-of-month data and were estimated for the 1 to 12-month return horizon. The dependent variable is the annualized log return in percentages over the h return horizon in excess of the 3-month US T-bill rate. The table contains only the regression coefficient and the Newey-West standard errors in parentheses. The number of Newey-West lags was set to 12, as in Johnson (2019). For the sake of readability, not all return horizons are displayed in the Table.*$p < .05$;**$p < .01$.

	Horizon							
	1	2	3	4	5	6	9	12
AEX	0.21	0.14	0.1	0.08	0.13	0.17	0.25**	0.24**
	(0.26)	(0.2)	(0.17)	(0.16)	(0.13)	(0.1)	(0.07)	(0.06)
AORD	-0.28	-0.2	-0.14	-0.11	-0.08	-0.01	0.00	0.02
	(0.21)	(0.15)	(0.13)	(0.14)	(0.13)	(0.13)	(0.13)	(0.11)
BVSP	0.38**	0.33**	0.29**	0.28**	0.24**	0.2**	0.17**	0.14**
	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DJI	0.64	0.42	0.41*	0.44**	0.44**	0.42**	0.35**	0.31**
	(0.34)	(0.24)	(0.18)	(0.13)	(0.11)	(0.1)	(0.09)	(0.08)
FCHI	0.77	0.45	0.19	0.22	0.38*	0.46**	0.42**	0.35**
	(0.42)	(0.28)	(0.26)	(0.22)	(0.17)	(0.15)	(0.13)	(0.11)
FTSE	0.00	0.03	0.06	0.11	0.13	0.13	0.17	0.17
	(0.28)	(0.17)	(0.14)	(0.13)	(0.12)	(0.13)	(0.12)	(0.11)
GDAXI	0.4	0.38	0.35*	0.38**	0.38**	0.38**	0.31**	0.31**
	(0.36)	(0.22)	(0.16)	(0.12)	(0.11)	(0.09)	(0.08)	(0.06)
HSI	0.01	0.00	0.00	0.01	0.03	0.04	0.03	0.01
	(0.1)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
IXIC	0.29	0.65**	0.67**	0.64**	0.6**	0.57**	0.44**	0.4**
	(0.38)	(0.16)	(0.14)	(0.15)	(0.15)	(0.12)	(0.1)	(0.09)
KS11	1.04**	0.92**	0.89**	0.86**	0.81**	0.71**	0.53*	0.41*
	(0.34)	(0.32)	(0.29)	(0.24)	(0.24)	(0.24)	(0.26)	(0.21)
MXX	-0.07	-0.1	-0.08	0.00	0.04	0.07	0.08	0.08
	(0.09)	(0.1)	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)	(0.09)
N225	0.09	0.08	0.03	0.01	0.04	0.06	0.04	0.02
	(0.1)	(0.08)	(0.08)	(0.07)	(0.06)	(0.05)	(0.05)	(0.04)
NSEI	0.03	0.01	0.00	0.01	0.02	0.02	0.03	0.03*
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.01)
RUT	0.29	0.23	0.23	0.24	0.26*	0.27**	0.23**	0.19**
	(0.26)	(0.2)	(0.18)	(0.16)	(0.13)	(0.1)	(0.08)	(0.07)
SPXCASH	0.3	0.22	0.2	0.21	0.22	0.22*	0.18**	0.16**
	(0.3)	(0.22)	(0.19)	(0.17)	(0.13)	(0.1)	(0.07)	(0.05)
STOXX50E	0.78*	0.54**	0.38**	0.35**	0.33*	0.29*	0.21	0.19
	(0.32)	(0.18)	(0.14)	(0.13)	(0.14)	(0.14)	(0.13)	(0.11)

Realized volatility forecasting: a comparison of HAR and ML models across international stock indices

Master Thesis

Department of Finance

Università Bocconi

Supervisor: Prof. Massimo Guidolin

Submitted by: Luca Mauritz (3210676)

Abstract

This thesis investigates the performance of eight forecasting models for realized volatility. The set of models includes different heterogeneous autoregressive (HAR) models as well as machine-learning (ML) models. The models' performance is tested across 31 international stock market indices for daily, weekly, and monthly horizons. The estimation is based on recursive backtesting with a monthly re-estimation of model parameters. The results of this thesis reveal robust performance of different HAR model extensions, especially for the daily and weekly forecasting horizon. Over the monthly forecasting horizon, the performance of some of the HAR models decreases whereas the machine learning models, especially the random forest, show improved performance. A final robustness test indicates that the performance of the long short-term memory (LSTM) model depends crucially on its pre-specified hyperparameters and may therefore be difficult to compare across different markets.

Contents

1 Introduction **1**

2 Literature Review **3**

3 Data **9**

4 HAR Models **12**

4.1 Benchmark HAR Model 13

4.2 HAR-J Model 14

4.3 Continuous HAR Model 15

4.4 Weighted Least Squares HAR Model 15

4.5 Robust Regression HAR Model 16

5 Machine Learning Models **18**

5.1 Random Forests 19

5.2 Support Vector Machine 21

5.3 Long Short-Term Memory Model 23

6 Empirical Analysis **27**

6.1 Estimation Approach 27

6.2 Loss functions 28

6.3 Diebold-Mariano Test 29

6.4 Daily Volatility Forecast 30

6.5 Weekly Volatility Forecast 36

6.6 Monthly Volatility Forecast 39

7 Robustness Check - LSTM Architecture **42**

8 Conclusion **45**

References **47**

A Appendix **50**

A.1 Additional Information on stock indices 50

A.2 Realized Volatility Graphs 51
A.3 QLIKE and MAPE results 67
A.4 Individual Diebold-Mariano tests 71

1 Introduction

Volatility forecasting is crucial for most areas of modern finance such as asset pricing, portfolio management, or risk management. In derivatives pricing, for example, volatility is a key input to price derivatives accurately. Similarly, in risk management, estimating future volatility is essential to derive the value at risk or expected shortfall of a given portfolio. Over the last decade, a range of new forecasting techniques has been proposed in the financial literature, each of them aiming to improve the accuracy of volatility forecasts. Due to better intra-day return data availability and improved computing power, most of the latest volatility forecasting models use realized volatility from intra-day stock returns to estimate future volatility. This thesis will add to the rich literature on realized volatility forecasting by comparing the performance of a large number of forecasting techniques over a sample of 31 international stock market indices.

In the current literature on volatility forecasting, the heterogeneous autoregressive model (HAR), as proposed by Corsi (2009), arguably remains the single most important forecasting technique, given its simple estimation and good forecasting accuracy. The HAR aims to predict future volatility from lagged daily, weekly, and monthly realized volatility. In the years after its publication, a range of extensions to the standard HAR model have been proposed such as the C-HAR and the HAR-J (Corsi, 2009) or the HAR-Q (Bollerslev et al., 2016), which can all be estimated by ordinary least squares (OLS). More recently, extensions of the HAR model that use different estimation techniques have been proposed, such as the weighted least squares HAR (WLS-HAR) or the robust regression HAR (RR-HAR), both introduced by Clements and Preve (2021). In addition to the extensive research on HAR models, a second field of research has emerged over the last years, which applies different machine learning techniques to model the non-linear and time-dependent relationships of realized volatilities. Several classes of machine learning models have been used, the most important being regularization (see for example Audrino and Knaus, 2016), random forests (for example Christensen et al., 2023), and neural networks (for example Li and Tang, 2022). Within the large field of neural networks, the long short-term memory (LSTM) model has been found to be particularly effective at forecasting realized volatility (Bucci, 2020).

Given the extensive research on different volatility forecasting techniques, my thesis

compares the performance of eight different volatility forecasting techniques across 31 international stock markets. The models include five variations of the HAR model: standard HAR, HAR-J, C-HAR, WLS-HAR, and RR-HAR, as well as three machine-learning models: Support Vector Machines, random forests, and the LSTM model. For this analysis, I apply recursive backtesting, using an expanding window approach and monthly re-estimation of the model parameters. My thesis adds to the existing literature on volatility forecasting in three important ways: first, most of the existing research has focused on US markets, whereas I will expand my analysis to 31 international stock markets to investigate market-specific differences in forecasting accuracy. Second, this thesis compares a number of machine-learning models to the most recent extensions of the HAR model which allows for a comparison of the latest models for volatility forecasting. In particular, the WLS-HAR and the RR-HAR have, to my best knowledge, not been tested against machine learning models in prior studies. Third, I apply a recursive backtesting scheme that is different from the ones used in previous studies by using an expanding window with monthly re-estimation of parameters, which strikes a good balance between real-world applicability and computational efficiency. On the one hand, I use an expanding window while most existing research uses a rolling window approach. On the other hand, by estimating the model parameters every month, I am able to compare a large number of indices while ensuring a realistic application of the forecasting models allowing the model parameters to react to new data being published.

Across all 31 stock market indices, my results indicate that the different HAR extensions emerge as the best-performing models for the daily and weekly forecasting horizons. Here, the WLS-HAR and RR-HAR models show superior forecasting performance. Yet, depending on which loss function is applied to measure forecasting performance, the C-HAR and the HAR-J also demonstrate solid forecasting accuracy. Furthermore, my results illustrate that the benchmark HAR was outperformed by its extension for most indices and across different loss functions. The machine learning models, however, show relatively poor performance for the daily and weekly forecasts and tend to be less accurate than even the benchmark HAR. The picture is different when forecasting monthly volatility: here, the random forest and the LSTM model are among the better forecasting techniques. Yet, the results vary depending on which loss function is used to measure and quantify forecasting performance. The WLS-HAR and RR-HAR perform well even for the monthly

forecasting horizon whereas the forecasting accuracy of the other HAR models decreases. The results reveal strong performance of the HAR extensions, which is in line with prior studies (Clements and Preve, 2021; Rahimikia and Poon, 2020). Yet, the performance of the machine learning models, especially that of the LSTM model, is weaker than found in prior studies (Bucci, 2020; Christensen et al., 2023). This inferior performance may be due to the characteristics of the backtesting scheme or the hyperparameter selection. A final robustness check indicates that the LSTM model's results are vulnerable to changes in the hyperparameters, but a comprehensive re-estimation for each stock index would probably be needed to significantly lower its forecasting error.

The main part of this thesis is structured as follows: Section 2 presents a detailed overview of the literature on volatility forecasting, while Section 3 describes the data used for this thesis and presents summary statistics. Section 4 introduces the five HAR models and Section 5 defines the three machine learning models tested in this thesis. Section 6 explains the estimation approach, introduces the loss functions, and presents the results for the different forecasting horizons. Section 7 contains a robustness check of the LSTM architecture and, finally, Section 8 concludes the thesis and points out potential future research.

2 Literature Review

There exists a large body of literature on volatility forecasting with different forecasting techniques being developed over time. In this section, I will first present a brief overview of the evolution of volatility forecasting techniques over the last decades. Then, I will describe in more detail the studies on the forecasting methods that will be employed in this thesis to give an overview of the current research frontier on stock market volatility forecasting.

Early attempts at forecasting volatility used past data to predict future volatility. For this, researchers and practitioners looked at different weighting schemes as well as different lag orders. Simple techniques, such as historical moving averages or exponential smoothing methods, in which more emphasis is put on more recent observations, were commonly employed. To address the strong auto-correlation in asset price volatility, different models of past observations have been used, including auto-regressive moving average (ARMA)

models (Poon and Granger, 2003).

Subsequently, a class of more complex models was introduced, when Engle (1982) published his work on auto-regressive conditional heteroskedasticity (ARCH) models. ARCH models allow for direct modeling of the conditional variance from past returns and were widely used in research as well as among practitioners. An important extension of the ARCH model came with the introduction of the generalized auto-regressive conditional heteroskedasticity (GARCH) models, introduced by Bollerslev (1986), which allowed today's conditional variance to depend on past returns and past conditional variances. Building on these results, a large number of extensions to the GARCH model have been proposed (see Poon and Granger (2003) for a comprehensive overview), which led to Engle (2002) calling it the "alphabet soup of volatility models". The exponential GARCH (EGARCH) model proposed by Nelson (1991), for example, has been regarded as one of the most significant extensions of the original GARCH model. The EGARCH model defines the conditional variance as a logarithmic variable. Hence, it ensures positive variances, but it also accounts for the so-called *leverage effect*: the empirical observation that negative returns lead to more volatility than positive returns of the same magnitude (Poon and Granger, 2003).

In the mid-2000s a new class of models appeared, based on realized volatility computed from intraday returns (see Section 3 for a detailed definition of realized volatility), pioneered by Corsi (2009) who proposed the heterogeneous auto-regressive (HAR) model. In this model class, today's volatility is regressed on lagged daily, weekly, and monthly volatility. The forecasting performance of the HAR model has been remarkable, which is why it was widely adopted as a benchmark model to test new volatility forecasting models. There have also been several important extensions of the HAR, such as the HAR-J (Corsi and Reno, 2009), accounting for jumps in the realized volatility, or the HAR-Q proposed by Bollerslev et al. (2016), which considers the impact of measurement errors on forecasting realized volatility. Similarly, Clements and Preve (2021) tested several adjustments to the original HAR model to improve its performance and robustness.

Finally, the most recent development in volatility forecasting has been the application of different machine learning techniques. Christensen et al. (2023), for example, employ random forests and other regression trees to forecast realized volatility. Furthermore, well-known regularization techniques, such as the least absolute shrinkage and selection

operator (LASSO) or Ridge regression, are used to forecast volatility (see for example, Audrino and Knaus, 2016 and Carr et al., 2019). Another approach for forecasting stock price volatility has been proposed by Gong et al. (2019) who are using support vector machines. In addition to these models, a number of deep-learning techniques have also been used. Several researchers have used neural networks to forecast volatility (see, among others, Christensen et al., 2023) where a special focus has been on the use of the long short-term memory (LSTM) model as, for example, applied by Bucci (2020).

After this brief overview of the evolution of volatility forecasting, I will now describe in more detail the existing research that is most relevant to this thesis. Most of the existing literature compares the forecasting performance of a new model of interest to the performance of benchmark models such as HAR or GARCH (see, for example, Li and Tang, 2022 or Bucci, 2020). The contribution of this thesis is to compare different machine learning models to the standard HAR model, but also to a range of extensions of the HAR model, that have been found to be particularly powerful at forecasting volatility. Furthermore, I compare the performance of the models across 31 different countries, which, to my knowledge, has not yet been done in other studies.

As mentioned above, the HAR model has been extended several times. Corsi and Reno (2009), for example, introduced the HAR with jumps (HAR-J) and the continuous HAR (C-HAR) which use a jump component and time series averages of bipower variation (see Section 4 for a detailed definition). A recent study by Rahimikia and Poon (2020) shows that the C-HAR outperformed all other HAR versions tested. Another well-known extension of the standard HAR model is the HAR-Q model as proposed by Bollerslev et al. (2016) that exploits *realized quarticity*, the fourth moment of intra-day returns, to control for the measurement error in the HAR estimation. Yet, realized quarticity data is not as readily available as data on realized variance, and a recent study by Clements and Preve (2021) finds that simple adjustments to the benchmark HAR model outperform the HAR-Q model. The authors find that estimating the HAR model using robust regression or weighted least squares regression instead of OLS significantly improves its performance compared to the HAR-Q or the benchmark HAR model. Hence, given the empirical findings from the studies listed in this paragraph, the HAR models tested within this thesis consist of the benchmark HAR, the HAR-J, the C-HAR, the robust regression HAR, and the weighted least squares HAR.

The objective of this thesis is to compare the performance of different HAR models to several machine learning models that have been suggested in the literature. These machine learning models can be categorized broadly into three groups: regression techniques, decision trees, and neural networks. What I define as regression techniques, covers regularization techniques, such as Ridge regression and LASSO, but also other regression problems like Support Vector Machines (SVM). Christensen et al. (2023), for example, employ Ridge regression and LASSO to forecast stock market volatility and find that the regularization techniques perform broadly in line with the HAR models for one-day forecasts. This result, however, only holds when the training data is restricted to the HAR input variables and no additional variables are used. According to Christensen et al. (2023), this shows that the lagged values of realized volatility are strong predictors that can hardly be improved. Similarly, Audrino and Knaus (2016) find no significant difference between LASSO and the standard HAR model for their out-of-sample performance. The authors attribute this to the fact that both models try to reduce the number of relevant lags to forecast volatility in a parsimonious manner. Another regression algorithm that has been used frequently in the literature is the SVM (see Section 5 for a detailed explanation). Gong et al. (2019) find that SVM, if using a least squares loss function, shows strong performance in volatility forecasting. Similarly, Liu (2019) shows that volatility forecasts generated with SVM can perform in line with the well-known long short-term memory (LSTM) model and both algorithms outperformed the GARCH model.

Another class of machine learning models that has gained significant attention in the literature are decision trees and, especially, the random forest algorithm (see Section 5). Luong and Dokuchaev (2018), for example, show that using the random forest algorithm can significantly improve the forecasting ability of the HAR model. Moreover, Christensen et al. (2023) finds that random forests outperform the HAR model for the one-day forecast horizon when a wide range of variables are used, not just the main variables from the HAR model. For a one-week horizon, the random forest produces a smaller error than the HAR model, even if using only the core HAR variables. Li and Tang (2022) also report that the random forest outperforms their comprehensive OLS model for the monthly and quarterly forecasting horizon.

More recently, the literature on volatility forecasting has focused on the use of neural networks. Li and Tang (2022), for example, report that for daily, weekly, and monthly

forecasts, a simple neural network shows the best performance across all machine learning models applied in their study. Similarly, Christensen et al. (2023) show that most neural networks outperform the benchmark HAR model for daily, weekly, and monthly forecast horizons. They find that this also holds for relatively simple neural networks. One type of neural network that has received a lot of attention is the long short-term memory (LSTM) Model (see Section 5). Bucci (2020), for example, shows that the LSTM model outperforms most conventional models for the one-day forecasting horizon, but not for the five-day horizon. Rahimikia and Poon (2023) confirm that the LSTM model outperforms a range of HAR models on days where volatility was not extreme.

Given this overview of the recent literature on the use of machine learning in volatility forecasting, this study will test the performance of one model from each of the three classes described above: support vector machines (SVM), random forest, and the long short-term memory (LSTM) model.

In addition to using different models to forecast volatility, another part of the financial literature focuses on the effects of forecasting volatility for different horizons, i.e., the difference between tomorrow's volatility or volatility over the next week. The frequency of the data can also be varied by using daily, monthly, or quarterly volatility. Different frequencies can have implications for the forecasting performance. Clements and Preve (2021), for example, show that the WLS and RR HAR are consistently outperforming the standard HAR model for longer forecasting horizons. Similarly, Christensen et al. (2023) report that machine learning models outperform particularly over longer time horizons, i.e., weekly and monthly. The authors find regression trees to be particularly effective in forecasting long-term volatility. Bollerslev et al. (2016), in contrast, find that the HAR-J and C-HAR model fail to outperform the standard HAR model for the weekly and monthly forecasting horizon. Within a range of machine learning models, Li and Tang (2022) report that neural networks perform best for forecasting daily volatility, whereas regularization techniques perform better over weekly and monthly forecasting horizons. The authors also point out that the ensemble model they identify as the best-performing model, which combines several machine learning techniques, does not outperform all individual models for all forecasting horizons.

The frequency of the realized volatility data also varies across studies. The publications that focus on the HAR model and its extensions usually use daily frequency (see for

example Bollerslev et al., 2016 or Clements and Preve, 2021). Some publications that focus more on machine learning models apply different frequency schemes. Bucci (2020) demonstrates supportive evidence for the LSTM model, utilizing monthly realized variance as the primary variable of interest. Echoing these findings, Filipović and Khalilzadeh (2021) also report strong performance of the LSTM model, employing monthly volatility data in their analysis. Christensen et al. (2023), on the other hand, use daily realized variances when comparing linear and machine learning models, but estimates the performance of the neural network based on a fixed window approach, whereas Bucci (2020) and Filipović and Khalilzadeh (2021) employ a rolling-window scheme. The results presented by Christensen et al. (2023) underline the strength of machine learning models, but their estimation approach does not easily allow daily out-of-sample predictions as it is based on a fixed window estimation. Similarly, Rahimikia and Poon (2023) find that machine learning models outperform different HAR models when using a fixed window approach.

Finally, a large share of the literature on realized volatility forecasting has been focusing on US stock data, such as the S&P-500 (see for example Bollerslev et al., 2016, or Clements and Preve, 2021), the NASDAQ (Rahimikia and Poon, 2023) or the Dow Jones (Christensen et al., 2023). Yet, there is some evidence, that the performance of different models differs between markets. Bollerslev et al. (2016), for example, report differing results between forecasting index volatility versus individual large-cap stock volatility. Another study by Kourtis et al. (2016) compares a range of linear forecasting models across ten international stock markets and concludes that the performance of models is relatively constant across markets but changes with the forecasting horizon. Furthermore, research by Degiannakis (2018) on combining different forecasting techniques, reveals differences between the major European stock markets. Horpestad et al. (2019) provide evidence that Asian markets differ from European and US markets in the performance of certain realized volatility models. Building on this relatively new line of research, this thesis will investigate the performance of volatility forecasting models across a sample of 31 international stock market indices.

In conclusion, this thesis adds to the large body of literature on volatility forecasting in three distinct ways: first, it compares a large number of powerful HAR extensions to a number of machine learning models. The weighted least squares and robust regression HAR models proposed by Clements and Preve (2021), in particular, have not been compared

in their performance to machine learning models. Second, this thesis combines daily and monthly frequency by re-estimating the model parameters monthly, but forecasting daily volatility (see Section 6 for a detailed explanation of the estimation approach). This approach seems to strike a good balance between the relevance of real-world risk management provided by daily volatility forecasts and the more computationally efficient strategy of re-estimating the more complex models once per month. Third, this thesis compares the performance of the different forecasting techniques across a sample of 31 international stock market indices and, hence, investigates whether the results found in prior research, mainly focusing on the US, also hold in international markets.

3 Data

The forecasted variable in this thesis is *realized volatility*, which is the most commonly used real-world measure of the integrated variance of stock price movements. Assume that the price of stock A follows the stochastic differential equation

$$dlog(P_t^A) = \mu_t dt + \sigma_t dW_t, \tag{1}$$

where P_t^A is the price of stock A on day t, μ is the drift variable, σ is the volatility, and W_t is the Brownian motion. The integrated variance of this process between day $t-1$ and day t can be written as

$$IV = \int_{t-1}^{t} \sigma_s^2 \, ds. \tag{2}$$

However, the integrated variance cannot be observed and is hence commonly proxied with the realized volatility

$$RV_t^d = \sum_{i=1}^{M} = r_{t-j\Delta}^2, \tag{3}$$

in which, following Corsi (2009), I define $M = h/\Delta$ as the number of intra-day returns taken into account. Hence, Δ is the frequency of intra-day returns, i.e., five minutes for my analysis and h is the number of trading hours in the respective market.[1] The intra-day

[1]For 6.5 trading hours and five-minute sampling frequency, for example, $M = 6.5 * 12 = 78$

returns can be expressed as

$$r_{t-j\Delta} = log(P_{(t-j\Delta)}) - log(P_{t-(j+1)\Delta}), \tag{4}$$

where t indexes the day and j the time during the day. Realized variance is a consistent estimator of integrated variance when the intra-day sampling frequency goes towards 0, $\Delta \to 0$, and hence, the number of intra-day returns becomes very large. This analysis is based on five-minute intra-day returns, a common choice in the literature on volatility forecasting (see for example Bollerslev et al. (2016) or Clements and Preve (2021)).

The indices used in this thesis cover developed and emerging markets and thus compare a large number of different stock indices including the world's most closely watched stock markets. For most of the indices, the data was retrieved from the Oxford-Man Institute's library on realized volatility which provided realized volatility data to the public until 2022. Only the indices for the S&P-500 were constructed by me in order to investigate possible differences between the cash and the futures index. Hence, they are termed SPXCASH (for the cash index) and SPXFUT (for the futures index) throughout this thesis. Table 1 presents summary statistics for all 31 stock markets covered in the analysis. Overall, the data covers the period from January 2000 until February 2020. Yet, the exact number of observations differs between indices (see Appendix A.1 for a more detailed description of all individual stock indices). In Table 1 we can observe a large extent of heterogeneity between stock markets, which underlines the relevance of this thesis. The maximum volatility and particularly the kurtosis differ significantly across the individual indices.[2]

Figure 1 illustrates the pattern of realized variance for the whole sample period across four large stock market indices included in the analysis: the Dow-Jones, the Nikkei-225, the S&P-500, and the Stoxx50 Europe (see Appendix A.2 for graphs on all 31 stock market indices). The graphs in 1 display a well-known stylized fact about volatilities: we observe long periods of low volatility as well as prolonged periods of very high volatility around certain events (Kim and Shin, 2023). This finding is known as volatility clustering (Mandelbrot and Mandelbrot, 1963). Another important stylized fact about asset price

[2]Note that for the sake of readability, Table 1 does not display the indices' full names, but only their Bloomberg tickers. Appendix A.1 contains a list of all stock market indices with detailed information on names, index composition, and the precise days covered in the study period.

Table 1: Summary Statistics - Stock Market Indices
Summary statistics on the daily realized volatility for all 31 stock market indices included in the analysis. The mean, median, average standard deviation and maximum values are displayed in percentages. The average standard deviation was computed as the average of the square root of daily realized variances. N describes the number of days that the respective index was present in the sample. The total sample period spans from January 2000 to February 2022, but the length of the individual indices varies. See Appendix A.1 for a list of all stock market indices with detailed information on names, index composition, and the precise days covered in the study period.

	Mean	Median	Avg. SD	Max	Kurt.	Skew.	N
AEX	0.0115	0.0056	0.9053	0.4200	97.6464	7.8908	5650
AORD	0.0053	0.0026	0.6095	0.4776	495.3322	18.0807	5594
BFX	0.0091	0.0050	0.8281	0.3689	143.4391	9.3099	5648
BSESN	0.0145	0.0071	1.0107	1.4898	646.9528	20.7365	5498
BVLG	0.0052	0.0036	0.6566	0.1575	147.8045	9.8937	2384
BVSP	0.0154	0.0094	1.0943	0.5912	115.3666	8.9032	5449
DJI	0.0111	0.0049	0.8534	0.8624	258.5274	12.4884	5549
FCHI	0.0132	0.0072	0.9885	0.5122	132.6006	9.1097	5652
FTMIB	0.0112	0.0069	0.9392	0.2728	70.1624	6.6687	3230
FTSE	0.0120	0.0057	0.9166	1.0600	418.0757	15.8558	5586
GDAXI	0.0158	0.0078	1.0616	0.5883	103.3088	7.7941	5614
GSPTSE	0.0083	0.0028	0.6890	2.9391	2859.9343	47.9415	4953
HSI	0.0096	0.0057	0.8698	0.4373	191.3357	10.7315	5426
IBEX	0.0141	0.0089	1.0485	0.5510	151.2359	9.3237	5615
IXIC	0.0124	0.0055	0.9276	0.5963	129.7255	8.6483	5554
KS11	0.0118	0.0058	0.9226	0.5944	161.6204	9.3120	5450
KSE	0.0101	0.0051	0.8577	0.3696	68.1538	6.1592	5398
MXX	0.0082	0.0043	0.7793	0.5225	294.2412	13.2708	5555
N225	0.0101	0.0059	0.8764	0.3849	131.1969	9.1827	5381
NSEI	0.0126	0.0058	0.9184	1.6725	894.9999	25.6498	5494
OMXC20	0.0125	0.0060	0.9366	1.2159	440.3879	17.0204	4077
OMXHPI	0.0111	0.0045	0.8456	2.2457	1713.1826	35.4348	4113
OMXSPI	0.0098	0.0041	0.8044	1.0535	555.5592	18.5198	4114
OSEAX	0.0132	0.0062	0.9579	1.9696	1328.1877	28.9602	5106
RUT	0.0081	0.0039	0.7527	0.3415	117.5368	9.1533	5550
SMSI	0.0126	0.0072	0.9768	0.6105	241.1251	12.1447	4244
SPXCASH	0.0059	0.0023	0.6082	0.2896	120.0070	9.3105	5174
SPXFUT	0.0090	0.0038	0.7673	0.4236	108.7177	8.7009	5386
SSEC	0.0155	0.0074	1.0544	0.4279	64.5629	6.2380	5348
SSMI	0.0085	0.0040	0.7777	0.5548	225.2164	12.2726	5552
STOXX50E	0.0156	0.0080	1.0558	1.0827	273.9459	12.3116	5636

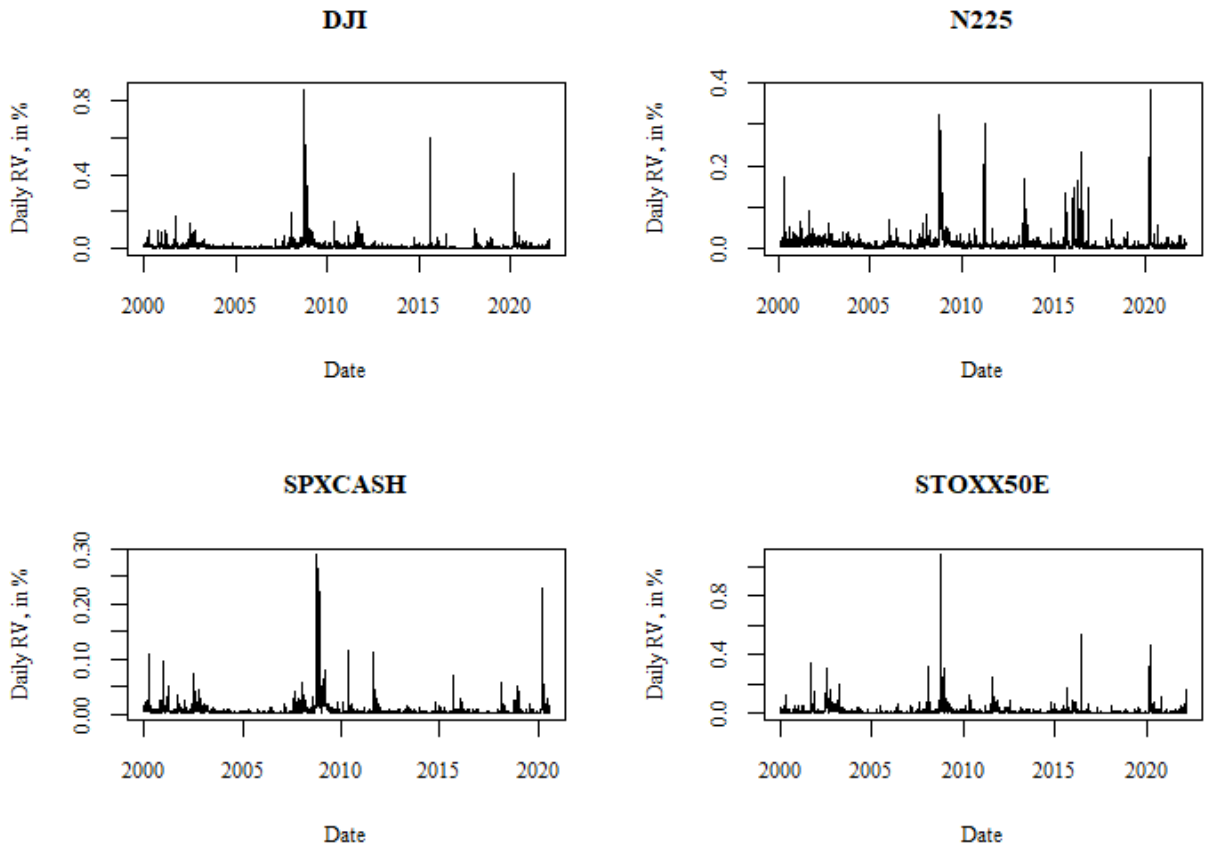

Figure 1: Realized Volatility of selected indices
Historical evolution of daily realized variance for four selected stock market indices: Dow Jones Index, Nikkei 225, S&P500 Cash Index, and the Stoxx 50 Europe. Graphs for all other stock market indices can be found in Appendix A.2.

volatility is the persistent and slowly decreasing autocorrelation (McAleer and Medeiros, 2008). This suggests that forecasting financial volatility from past observations, even over an extended forecasting horizon, can be achieved. Figure 2 illustrates the autocorrelation functions for the same four stock market indices as displayed above. One can observe a similar pattern across all four stock market indices: the autocorrelation is declining but the rate of change is decreasing over time.

4 HAR Models

The first group of models used for this analysis, consists of different variations and extensions of the well-known heterogeneous autoregressive (HAR) model pioneered by Corsi (2009). HAR models successfully capture some of the most common empirical

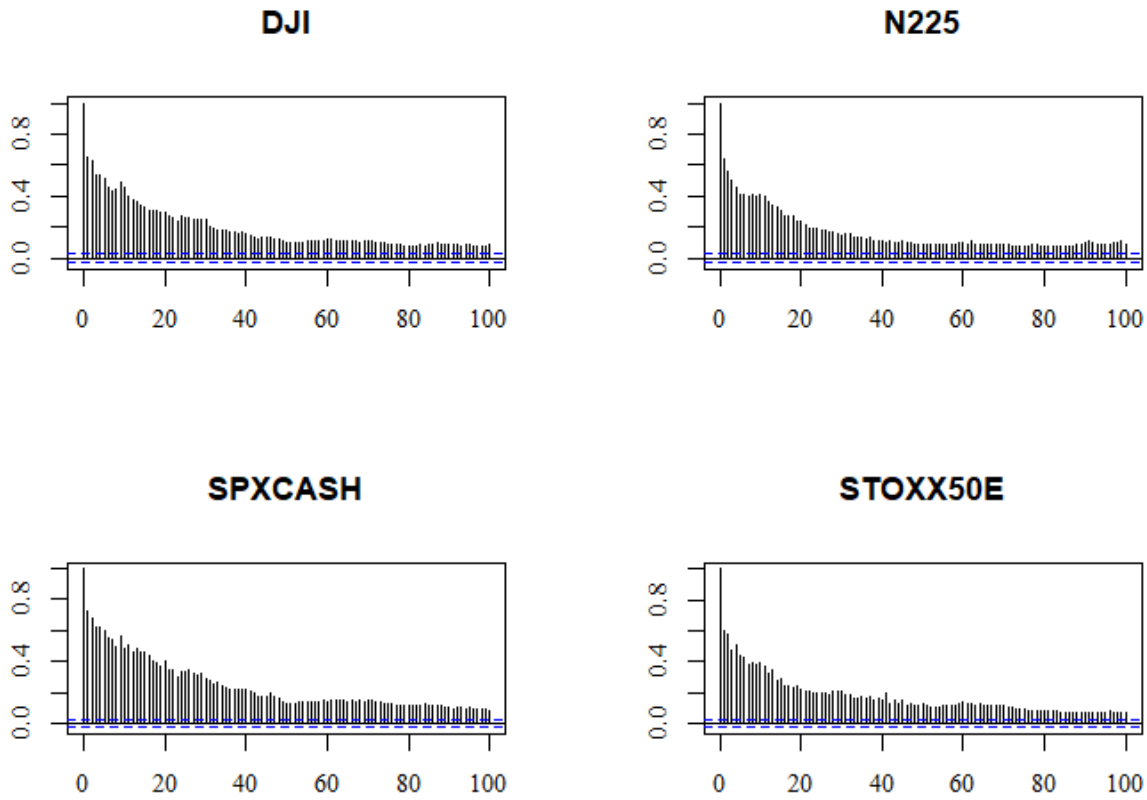

Figure 2: Autocorrelation Function for selected indices
Autocorrelation function for daily realized volatility for four selected stock market indices: Dow Jones Index, Nikkei 225, S&P500 Cash Index, and the Stoxx 50 Europe.

features of stock price volatility such as long memory, high auto-correlation, and kurtosis. In this Section, I will introduce five different HAR models: the standard HAR model, which will serve as a benchmark model, the HAR-J model, the continuous HAR model, a HAR model estimated with weighted least squares regression, and finally a HAR model estimated with Robust Regression.

4.1 Benchmark HAR Model

The benchmark HAR model, as introduced by Corsi (2009), estimates integrated volatility via realized volatility from intra-day returns (see Equation 3). The aim of the HAR model is to estimate future realized variance, based on daily, weekly, and monthly average realized variances that are known today. Daily realized variance was defined in Equation

(3) whereas the weekly average realized variance can simply be expressed as

$$RV_t^w = \frac{1}{5}(RV_t^d + RV_{t-1}^d + \cdots + RV_{t-4}^d) \tag{5}$$

Similarly, the monthly average realized variance is given by

$$RV_t^m = \frac{1}{22}(RV_t^d + RV_{t-1}^d + \cdots + RV_{t-21}^d), \tag{6}$$

assuming 22 trading days in a month. The HAR model can then be estimated as

$$RV_{t+k} = \beta_0 + \beta_1 RV_t^d + \beta_2 RV_t^w + \beta_3 RV_t^m + u_{t+k} \tag{7}$$

where k determines the forecast horizon. For this analysis, I will compare three common forecast horizons. Hence, $k = 1, 5, 22$, depending on whether one intends to forecast, daily, weekly, or monthly volatility. In the standard HAR model, Equation (7) will be estimated using OLS.

4.2 HAR-J Model

The HAR-J model builds on the work by Barndorff-Nielsen and Shephard (2004), who show that bipower variation can be used as an estimator that is robust against rare jumps in the data and, hence, allows us to separate realized volatility from these jumps. Bipower variation on day t is defined as

$$BPV_t = \frac{\pi}{2} \sum_{i=1}^{M-1} |r_{t,i}||r_{t,i+1}|. \tag{8}$$

Here, M is the sampling frequency as in Equation 3 and $r_{t,i}$ is the return on the respective asset on day t in sampling interval i. Bipower variation was then included by Corsi and Reno (2009) to extend the benchmark HAR model allowing the model to capture the effect of jumps. Here, I follow the notation introduced by Rahimikia and Poon (2020) who estimate the HAR-J model as

$$RV_{t+k} = \beta_0 + \beta_1 RV_t^d + \beta_2 RV_t^w + \beta_3 RV_t^m + \beta_J J_t + u_{t+k} \tag{9}$$

where all variables are defined as in Equation (7) and jump J on day t is defined as

$$J_t = max(RV_t - BPV_t, 0). \tag{10}$$

Adding the jump component in Equation (9) permits the model to detect the short-term volatility introduced by jumps in the data (Rahimikia and Poon, 2020).

4.3 Continuous HAR Model

Bipower Variation, as introduced in Equation (8) was also used by Corsi and Reno (2009) to further expand the HAR model. The authors proposed a continuous volatility HAR, which uses the previous values of bipower-variation, instead of the previous values of realized variance, to explain today's realized variance. The continuous HAR (C-HAR) can hence be written as

$$RV_{t+k} = \delta_0 + \delta_1 BPV_t^d + \delta_2 BPV_t^w + \delta_3 BPV_t^m + \epsilon_{t+k}, \tag{11}$$

where BPV_t^w and BPV_t^m are the weekly and monthly averages of the bipower-variation, respectively. Without the jump component in Equation (9), the C-HAR shows better performance at modeling the time persistence and auto-regressive patterns in realized variance (Rahimikia and Poon, 2020).

4.4 Weighted Least Squares HAR Model

As mentioned above, the standard HAR model is estimated via OLS, and hence, all models that were introduced so far, are estimated by OLS. Yet, given the known characteristics of return data and realized volatility, such as outliers, conditional heteroskedasticity and a return distribution that does not follow a normal distribution (Clements and Preve, 2021), one should consider other estimation techniques. This seems particularly important as estimation results from OLS are known to be vulnerable to violations of the assumptions described above. Hence, Clements and Preve (2021) proposed a range of new specifications of the HAR model, that follow the basic concept of HAR models, but use different input variables and different estimation techniques. Within the scope of this thesis, I will introduce two models that Clements and Preve (2021) showed to have superior

performance among a wide range of different models: a HAR model estimated by weighted least squares (hereafter WLS-HAR) and a HAR model estimated via robust regression (hereafter RR-HAR). While OLS minimizes the sum of squared residuals, weighted least squares (WLS) weighs these residuals and minimizes the weighted sum. Therefore, if an adequate weighting scheme is chosen, WLS can be a more efficient estimator compared to OLS which would result in more precise volatility forecasting. One can estimate the WLS-HAR as in Clements and Preve (2021) via the following minimization problem:

$$\min_{\phi_0,\phi_1,\phi_2,\phi_3} \sum_{t=23}^{n} w_t(RV_{t+k} - \phi_0 - \phi_1 RV_t^d - \phi_2 RV_t^w - \phi_3 RV_t^m)^2, \tag{12}$$

where w_t is the weight on every observation t and n is the number of days in the sample. Clements and Preve (2021) test a range of different weighting schemes. The authors identify a model using weights that are inversely proportional to the realized quarticity on that day to be the model with the best performance in forecasting realized variance. Realized quarticity is the fourth moment of intraday returns and can be seen as the volatility of the volatility. Therefore, using these weights, the WLS HAR will put less weight on days on which there is high uncertainty around the volatility measure. However, Clements and Preve (2021) also tested a model using weights $w_t = 1/RV_{t-1}$ and found its performance to be comparable to the model using realized quarticity due to the high correlation between realized variance and realized quarticity. Therefore, they conclude that such a model represents a good alternative for situations in which realized quarticity is not available for all observations. Hence, as realized quarticity was not available for a large share of indices used for this thesis, I will estimate the WLS-HAR using the above-mentioned weights that are inversely proportional to the first lag of realized variance.[3]

4.5 Robust Regression HAR Model

The final HAR model included in this analysis, as mentioned above, is a HAR model that is estimated with Robust Regression (RR). RR was pioneered by Huber (1964) and represents an estimation technique that is more robust to outliers in the data. These

[3]Note that realized quarticity can only be estimated from intraday returns which were not available for the Oxford Man data sets that provided the majority of the data for this thesis. For the sake of comparability, I did not use realized quarticity for the remaining data sets, for which it could be computed.

outliers can pose a severe problem for OLS, which is known to perform poorly under the existence of unusual observations in the data. If we wish to estimate a linear model

$$\begin{aligned} Y_i &= \alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \epsilon_i \\ &= \mathbf{x_i'}\boldsymbol{\beta} + \epsilon_i \end{aligned} \tag{13}$$

as is the case for the HAR estimation, RR applies an objective function $\sum_{i=1}^{n} \rho(\epsilon_i)$ to the residual ϵ_i. Estimators that minimize an objective function are called M estimators. One of the most common M estimators is Tukey's biweight function (Fox, 1997) which I will use for this analysis. This is in line with the work by Clements and Preve (2021). Tukey's biweight function can be defined as

$$\rho(\epsilon) = \begin{cases} \frac{k^2}{6}\{1 - [\,1 - (\frac{\epsilon}{k})^{\,2}\,]^{\,3}\} & \text{for } |\epsilon| \leq k \\ \frac{k^2}{6} & \text{for } |\epsilon| > k, \end{cases} \tag{14}$$

where k is the tuning constant that is used to distinguish between the center and the tails of the distribution of residuals.[4] The tuning parameter k is usually expressed as a function of the scale of the dataset, for example, $k = cS$, where variable S is defined as a quantification of scale (Fox, 1997)). A commonly used and robust choice for S is the *median absolute deviation* (MAD), which is defined as

$$MAD = \text{median}|y_i - \hat{\mu}|, \tag{15}$$

where $\hat{\mu}$ is an estimate of the sample mean. As we can see in Equation (14), the objective function flattens out for residuals above k. Hence, a smaller tuning parameter can be used to adjust the robustness of the RR model.[5] Finally, *iterative weighted least squares* (IWLS), can be used to compute the M estimators. For this, we define a weight function $w(\epsilon) = \frac{\rho'(\epsilon)}{\epsilon}$, where $\rho'(\epsilon)$ is the first derivative of the objective function in (14) with respect

[4]Note that for the sake of readability, I suppressed the subscripts of the residuals as defined in Equation (13)

[5]I used R's "rlm" function from the "MASS" package and its default value for c, when applying Tukey's biweight function and using the MAD as a measure of scale.

to ϵ. The weight function can therefore be written as

$$w(\epsilon) = \begin{cases} [\,1 - (\frac{\epsilon}{k})^{\,2}\,]^{\,2} & \text{for } |\epsilon| \leq k \\ 0 & \text{for } |\epsilon| > k. \end{cases} \tag{16}$$

The weight function illustrates the fact that no weight will be put on observations with residuals that exceed the tuning parameter k. For regression analysis, the M estimator is applied to the linear model (see Fox, 1997, for all intermediate steps), and, using the weight function, the estimation equation becomes

$$\sum_{i=1}^{n} w_i(y_i - \mathbf{x_i'}\boldsymbol{b})\mathbf{x_i} = \mathbf{0}, \tag{17}$$

which minimizes the weighted sum of squares. Yet, the weights depend on the residuals, the residuals depend on the coefficients, and those depend on the weights again (Fox, 1997). Hence, this minimization problem must be solved with an iterative method. IWLS sets initial regression coefficients and uses those to back out the residuals, which allows us to retrieve the scale S and thereafter the weights w_i. Then, IWLS proceeds at each iteration by solving equation 17 with the current weights to derive new regression coefficients $\mathbf{b}$. The algorithm follows this iteration process until the changes in the regression coefficients become insignificant and $\mathbf{b}^{(I)} - \mathbf{b}^{(I-1)} \approx 0$, with I being the iteration counter (Fox, 1997).

5 Machine Learning Models

In addition to the different HAR models introduced above, this thesis also makes use of three machine learning models and contrasts their performance in forecasting volatility with the linear models presented above. These three machine learning models are random forests, support vector machines (SVM), and the long short-term memory (LSTM) Model. All three models have been shown to be highly effective in forecasting asset price volatility.

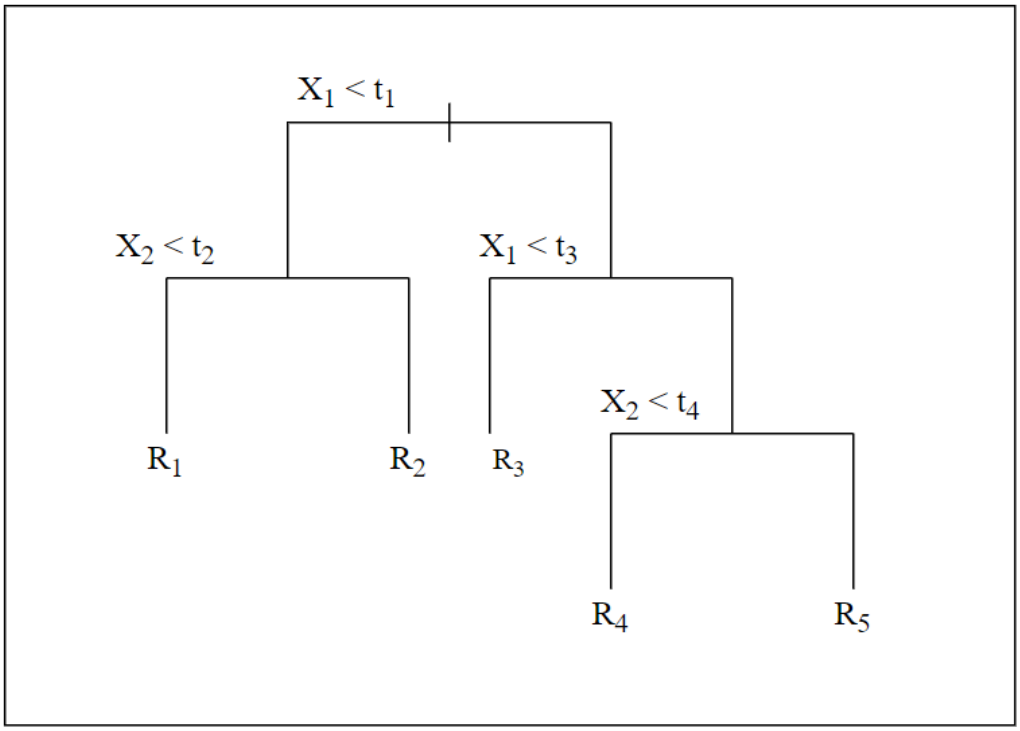

Figure 3: Illustration of a simple regression tree
The figure displays the structure of a simple regression tree using two predictors, X_1, X_2 and four internal nodes $t_1, \ldots, t_4$, resulting in five different external nodes $R_1, \ldots, R_5$. The figure is displayed as in Hastie et al. (2009).

5.1 Random Forests

Random forests can be described as a machine learning algorithm that generates and combines a large number of decision trees to reduce the noise embedded in individual decision trees. Decision trees can be used for classification and regression problems, yet, this explanation will focus solely on so-called regression trees given the numerical nature of forecasting volatility. Regression trees apply a set of splitting rules to each observation of the data set. Figure (3) shows the shape of a simple regression tree using two predictors, X_1, X_2. A regression tree is characterized by its *internal nodes* and its *terminal nodes* or *leaves* (Witten and James, 2013). Internal nodes are the points inside the tree where the individual observation is assigned to the left or right branch of the tree, depending on a certain parameter. On the top of the tree in Figure (3), for example, we see that an observation is assigned to the left part of the tree if $X_1 < t_1$, where t_1 is the first parameter used to separate the data. External nodes, also called *regions*, are the different outcomes of the regression trees: five in the case of Figure (3).

In order to construct the tree, all variables are scanned to find the combination of splitting variables and splitting points that split the data set most efficiently. This procedure is repeated at each internal node using the remaining observations in the respective branch of the tree (Hastie et al., 2009). If regression trees are grown sufficiently deep, meaning that they have a high number of internal nodes, they tend to fit the data well and show little bias. Yet, regression trees are known to produce noisy estimates (Hastie et al., 2009). Therefore, several machine learning algorithms have been proposed to solve this problem, random forests being one of them.

The concept behind random forests is to generate a large amount of uncorrelated trees and compute the average of their predictions. The main characteristic of random forests is, that in order to obtain uncorrelated trees, a subset of $m \leq p$ of the predictors is drawn before generating a new tree, where p is the total number of independent variables in the data set. Usually, the number of predictors used per tree is selected as $m = \sqrt{p}$ (Hastie et al., 2009). Furthermore, the individual trees are trained on a sub-sample of the data set which means that not all available observations are used for each tree. Hence, each new tree does not see all of the existing data nor all of the independent variables, which leads to a lower correlation between trees. For regression problems, the final prediction of the random forest will be the average of all generated trees. As the data sets generated from the sub-sampling are identically distributed, the expectation of their averages will be identical. Therefore, the only possibility to improve the performance of the forecast is by reducing the variance (Hastie et al., 2009). The variance of the average can be written as

$$\rho\sigma^2 + \frac{1-\rho}{N}\sigma^2, \tag{18}$$

where ρ is the correlation between trees, N is the number of trees generated, and σ^2 is the variance of the individual tree. Equation 18 shows that the term on the right side converges to zero as N increases and the term on the left side decreases with a lower correlation between trees (Hastie et al., 2009). The size of the sub-sample, the number of predictors in each sub-sample, and the number of trees must be selected by the researcher in a way that minimizes the variance of the prediction. The selection of hyperparameters is often carried out via grid search or comparing the performances of different model. This approach, however, was not feasible for this thesis, given the large number of stock indices and the monthly re-estimation of the models. Hence, hyperparameters were selected

based on well-known heuristics or in line with the literature on volatility forecasting. For the random forests, the number of trees was set to 500 and each tree was trained on a sub-sample of 50% of the observations as in Li and Tang (2022). The number of trees should be sufficiently large as this generally improves results (Probst et al., 2019) if it does not significantly impair computation time. The number of predictors, m, that each individual tree was trained on, was selected based on the heuristic $m = \sqrt{p}$ which is a common choice in the literature and was found to yield similar results compared to other heuristics (Bartz et al., 2023).

5.2 Support Vector Machine

The main idea for support vector machines (SVM), sometimes also referred to as *support vector regressions* is to explain an output variable y_t with a set of p predictors $x_{1,t}, \ldots, x_{p,t}$ accepting a deviation of ϵ (Smola and Schölkopf, 2004). All losses above this threshold should be penalized with a cost parameter C, losses below this threshold are not penalized. Put differently, with support vector regression we accept certain deviations from the fitted function as long as they do not exceed a certain threshold. Hence, for a linear optimization problem, the function we want to fit becomes

$$f(x) = g(w, x) + b, \tag{19}$$

where $g(w, x)$ is the best model from a set of available models that tries to minimize prediction loss. w describes the vector of parameters that has to be parametrized to fit the function. Variable b is a bias term that captures random noise (Cherkassky and Ma, 2004). Following Smola and Schölkopf (2004), the minimization problem can therefore be expressed as:

$$\begin{aligned} \min \quad & \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{s.t.} \quad & y_i - g(w, x_i) - b \leq \epsilon + \xi_i \\ & g(w.x_i) + b - y_i \leq \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0. \end{aligned} \tag{20}$$

Here, C is the cost parameter that penalizes deviations above ϵ and ξ_i, ξ_i^* are the so-called *slack parameters* that capture the penalty above ϵ. These two parameters capture the

trade-off between fitting a complex function while at the same time allowing for some errors above ϵ as a function without such deviations probably does not exist for complex problems (Cherkassky and Ma, 2004). The slack parameters can be defined as

$$|\xi|(\epsilon) := \begin{cases} 0 & \text{if } |\xi| \leq \epsilon \\ |\xi| - \epsilon & \text{if } |\xi| > \epsilon. \end{cases} \tag{21}$$

To estimate a support vector regression, the researcher must specify four hyperparameters: the cost parameter C, the acceptable deviation ϵ, the kernel function, and *gamma* which influences the kernel function. Given that I re-estimate the support vector regression every month and compare 31 different stock market indices, the hyperparameters had to be set using well-established heuristics as applying grid search was not feasible for the large number of models that had to be estimated. First, cost parameter C and ϵ were estimated following Cherkassky and Ma (2004) as their heuristic represents a common approach for hyperparameter selection (Bartz et al., 2023). Hence, C was computed as

$$C = max(|\overline{y} + 3\sigma_y|, |\overline{y} - 3\sigma_y|), \tag{22}$$

where $\overline{y}$ is the sample mean of the dependent variable, daily realized variance in the case of this thesis and σ_y is its standard deviation. Both, mean and standard deviation, are computed from the training data (Cherkassky and Ma, 2004). Second, ϵ should depend on the size of the training data and the noise in the input data (Cherkassky and Ma, 2004). Thus, ϵ can be determined as

$$\epsilon = 3\sigma\sqrt{\frac{ln\ n}{n}}, \tag{23}$$

where σ is the square root of the noise variance. To estimate the noise variance, Cherkassky and Ma (2004) propose estimating a complex, but low-bias model and extracting the variance of the residuals. For this, the authors estimate a *k-nearest neighbor* model. This model takes the local average of the k closest variables as a prediction of the dependent variable. Building on Cherkassky and Ma (2004) one can compute the noise variance as

$$\hat{\sigma}^2 = \frac{n^{1/5}k}{n^{1/5}k - 1}\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \tag{24}$$

Here, n is the number of observations in the training data, $y_i - \hat{y}_i$ is the prediction error, and k is the number of variables for the k-nearest neighbor model. Following Cherkassky and Ma (2004), I used $k = 3$ for all models. The square root of the noise variance can then be plugged into Equation 23 to compute ϵ. The third hyperparameter γ is usually computed as $\gamma = 1/F$, where F is the total number of features in the training data. This approach to determining γ is a common default in the literature (Bartz et al., 2023) which yields good performance for different applications. Finally, for the selection of the kernel function, Bartz et al. (2023) report that for most applications, the *radial basis* kernel shows good performance and was therefore also used for all models in this thesis.

5.3 Long Short-Term Memory Model

Finally, this thesis applies a neural network to forecast realized volatility, more specifically, a long short-term memory (LSTM) model. For this, I will first introduce neural networks in general, before explaining recurrent networks and the LSTM model in more detail.

In its most general form, a neural network takes a vector of inputs X_j and transforms those via a non-linear function to predict a set of outputs (Witten and James, 2013). Figure 4 depicts the structure of a simple neural network. The *input layer*, on the left, consists of the four input features $X_1, \ldots, X_4$ that feed into the *hidden layer* - the second layer of the neural network. A neural network can have multiple hidden layers, but here I am presenting only one, as this paper will focus on neural networks with a single hidden layer in line with Bucci (2020). Within a hidden layer, there are several K *hidden units* which have to be specified by the researcher, five in the case of the simple example in Figure 4.

The first part of the neural network model is constructed via the K *activations* that are computed from the input vector such that

$$A_k = g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j), \tag{25}$$

where A_k are the K *activations*, X_j are the p input values, w_{kj} are the weights of these inputs and $g(Z)$ is a nonlinear *activation function* which has to be specified in advance

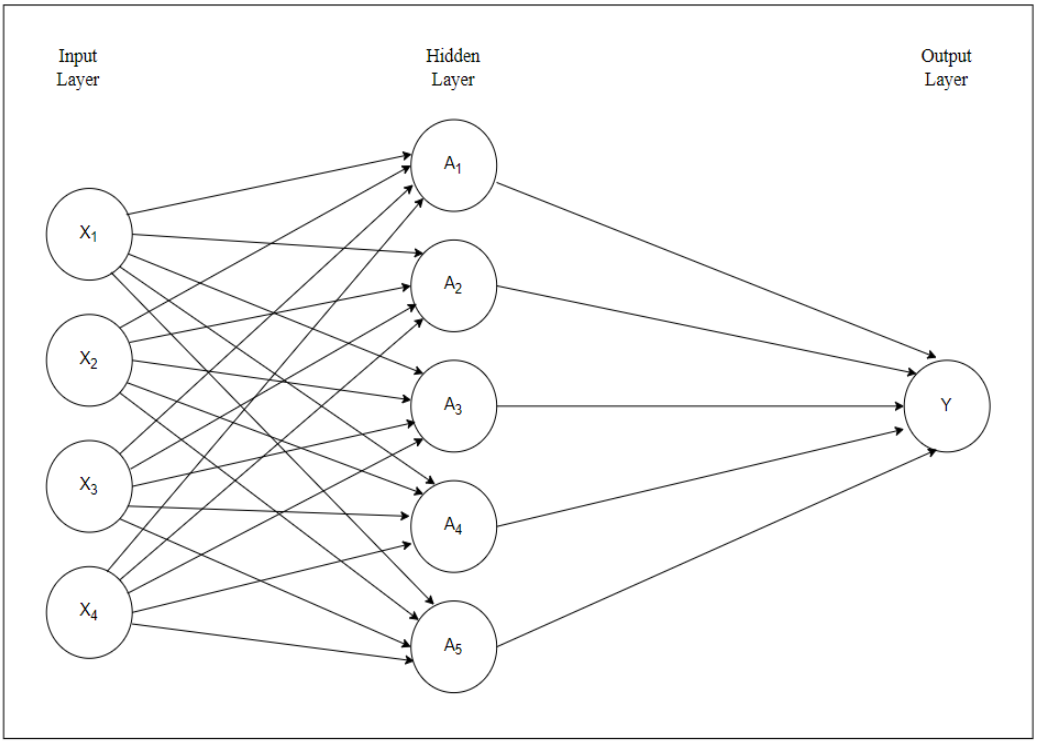

Figure 4: Illustration of a simple regression treee
Structure of a simple regression tree using two predictors, X_1, X_2 and four internal nodes $t_1, \ldots, t_4$, resulting in five different external nodes $R_1, \ldots R_5$. The figure is displayed as in Witten and James (2013).

by the researcher. A popular choices for $g(Z)$ is the *sigmoid activation function* (Bucci, 2020):

$$g(z) = \frac{1}{1 + e^{-z}}, \tag{26}$$

which is the same function as in logistic regression. The *sigmoid* activation function allows the neural network to model nonlinear structures in the data. The hidden units, or *neurons*, show a high activation when A_k is close to one and practically no activation when A_k is close to zero (Witten and James, 2013). As a final step, the neural network feeds the activation from the hidden layer into the output layer via a linear model, which can be expressed as:

$$f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k A_k. \tag{27}$$

Hence, the output of the general neural network is linear in the activations from the hidden layer (Witten and James, 2013). The complete model of the neural network can

therefore be written as:

$$f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k \, g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j), \tag{28}$$

where all parameters $\beta_0, \ldots, \beta_K$ and $w_{10}, \ldots w_{Kp}$ have to be estimated from the data. Note that for the neural network described above, one assumes that information flows forward from the input layer, through the hidden layer into the output layer. This type of neural network is called *Feed Forward Network* (FNN) or *static* models as it predicts an output variable at a fixed time t. FNNs are unable to model dependencies across time in the data and have no memory function to capture the sequential nature of certain data sets (Bucci, 2020). Hence, for the main analysis of this thesis, I apply a different type of neural network, which is better suited for time series data: *recurrent neural networks.*

In a recurrent neural network (RNN), the input consists of sequential data, e.g., a time series of stock prices, and the network capitalizes on the sequential characteristics of the data (Witten and James, 2013). An RNN allows feedback from later to earlier layers, which gives the network the ability to save certain information in its memory (Bucci, 2020). Such a model can be *unfolded* to look at each period t individually. Then, for each timestamp, the model consists of an input value, an output value, and a hidden unit. The sequential information in the time series data is used to train the model. After training, the same weighting function is used for all timestamps (Aggarwal et al., 2018).

The LSTM model is a well-known RNN that has been shown to be particularly powerful in forecasting volatility (see chapter 2). The main novelty of the LSTM is that it uses a *cell state* that determines how much of the information from prior layers of the model is kept and how much is forgotten. This allows the model to accurately learn sequential information embedded in the data (Aggarwal et al., 2018). The *cell state* is being updated with the help of three variables: the *input gate* i_t, the *output gate* o_t, and the *forget gate* f_t. The input gate adds new data to the cell state c_t. Separately, the forget gate determines how much of the information from prior layers and of the new information is kept. Finally, the output gate computes the outcome of the LSTM cell by using all the information in the cell state (Bucci, 2020). Figure 5 illustrates the dynamics within a sample LSTM cell.

Mathematically, following the notation suggested by Bucci (2020), the procedure

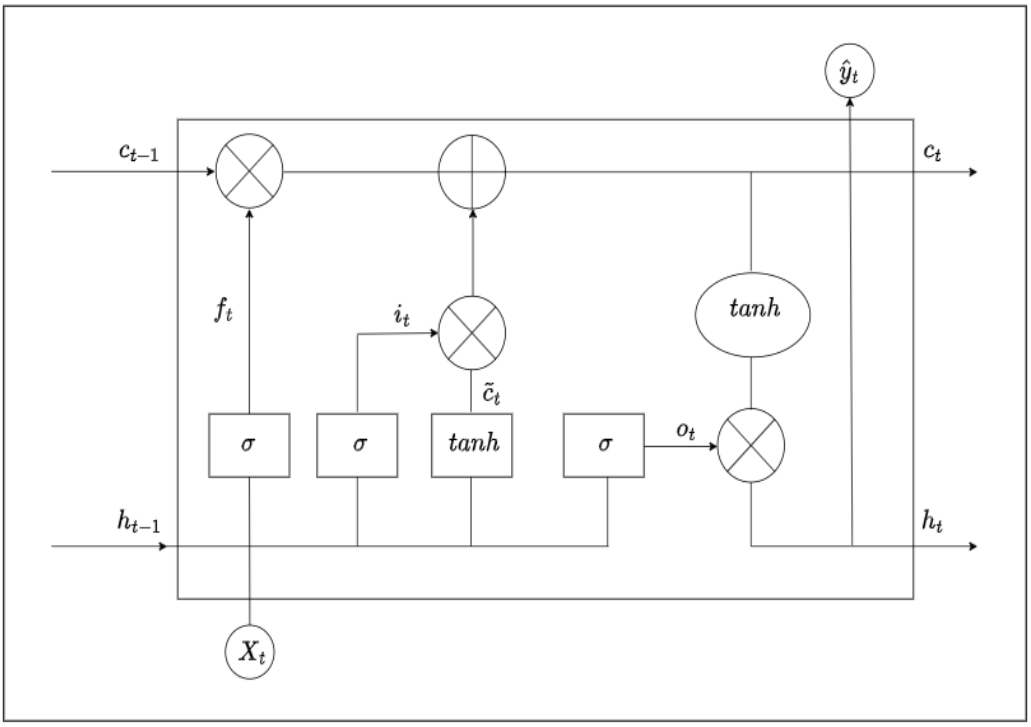

Figure 5: Illustration of an individual LSTM Cell
Structure of a single long short-term memory Cell. The illustration shows how information within the cell is passed on from the input gate to the output gate, using the forget gate to delete redundant information. $\oplus$ describes element-wise addition, whereas $\otimes$ stands for element-wise multiplication. The figure is displayed as in Bucci (2020).

within a single LSTM cell can be expressed as:

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \tag{29}$$

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \tag{30}$$

$$\tilde{c}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \tag{31}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{32}$$

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + V_o c_t + b_o) \tag{33}$$

$$h_t = o_t \odot \tanh(c_t) \tag{34}$$

$$\hat{y}_t = h_t, \tag{35}$$

where $\tilde{c}_t$ is a vector of potential new values of the cell state, h_t is the hidden state of the cell, and $\hat{y}_t$ is the final output of the model. Furthermore, within the LSTM cell,

two functions are used: the tanh-function, and the logistic function from Equation 26, defined as σ. W_f, W_i, W_c, W_o, U_f, U_i, U_c, U_o denote the matrices containing the model weights for the forget gate, input gate, cell state, and output gate, respectively. Finally, $\odot$ describes the element-wise product of two matrix objects.

For the main analysis of this thesis I will deploy an LSTM network with one hidden layer and 50 units as suggested by Bucci (2020). Furthermore, I used the *Adam* optimization algorithm as proposed by Kingma and Ba (2014). The Adam algorithm is based on stochastic gradient descent and is a common choice in the literature on neural networks. It is also used by Rahimikia and Poon (2023) in their study on the use of the LSTM for variance forecasting. Furthermore, I used the mean absolute error as the loss function for the LSTM. Hyperparameters have to be determined by the researcher. Section 7, however, investigates how robust the forecasting performance of the LSTM is to changes in some of the parameters. The LSTM was trained on the first 22 lags of daily realized variance to facilitate a direct comparison with the HAR models.

6 Empirical Analysis

6.1 Estimation Approach

To compare the performance of the different models across the 31 stock market indices, I employ an expanding window approach with monthly re-estimation of the models. In order to investigate the forecasting performance of each model I use a quasi-out-of-sample backtesting approach. Initially, each model was trained on the first 1000 trading days available for the respective stock market index. After that, the estimated model is used to compute out-of-sample predictions of realized variance for the next month.[6] The estimation takes into account that new data becomes available during the out-of-sample period. For example, on day five of the out-of-sample period, the HAR model would use the realized variance of the past four days for its prediction. Yet, it was not trained on these observations and its parameters are held constant for the entire month. After the one-month out-of-sample prediction, the model is re-estimated on an expanded window of $N = 1000 + T$ observations, where T is the number of trading days in a month. Hence,

[6]Here, I defined a month as having 20 trading days which largely facilitated computations for the LSTM model.

the sample that each model is trained on, increases over time until the model uses almost all available data for the final out-of-sample predictions. Overall, this approach seems to strike a good balance between the computational efficiency of re-estimating the models every month and the real-world applicability of using all data available on a given day t.

The models are tested based on their performance for each of the 31 stock market indices. For this, stock market volatility was forecasted for three different time horizons: daily, weekly, and monthly. The performance of the individual models was then assessed using common loss functions such as the mean squared error (MSE), the quasi-likelihood error (QLIKE), and the mean absolute percentage error (MAPE), as well as pairwise statistical testing with the use of Diebold-Mariano tests. The following sections introduce the loss functions and the Diebold-Mariano test before turning to the results of the daily, weekly, and monthly forecasts.

6.2 Loss functions

Three different loss functions were applied to evaluate the performance of the different models. The MSE is defined as

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2. \tag{36}$$

Here, T is the total number of days in the sample, y_t is the true, observed realized variance on day t and $\hat{y}_t$ is the realized variance forecasted by the respective model.[7] The QLIKE, which describes the relative volatility forecasting error, can be written as

$$\text{QLIKE} = \frac{1}{T}\sum_{t=1}^{T}(\frac{y_t}{\hat{y}_t} - log(\frac{y_t}{\hat{y}_t}) - 1), \tag{37}$$

where all variables are defined as in Equation 36. Finally, MAPE describes the absolute error committed in the forecasting process, relative to the forecasted quantity, which I define as

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100. \tag{38}$$

Again, all variables are defined as introduced in Equation 36.

[7] For the sake of simplicity, I define the loss functions only for the forecasting of daily volatility. Yet, the loss functions are defined accordingly for the forecasting of weekly and monthly volatility.

6.3 Diebold-Mariano Test

To compare the different forecasting models within each country, I will apply the Diebold-Mariano test, which is commonly used in the literature on volatility forecasting (see for example Christensen et al., 2023). The test was introduced by Diebold and Mariano (2002) in order to compare the forecasting accuracy of different models. Following the original notation by Diebold and Mariano (2002), I compute the pairwise mean loss differential $\bar{d}$ for all forecasting models, where d is defined as

$$d = e_j - e_i. \tag{39}$$

Here, e is the squared, daily prediction error of model j and model i, respectively. The test statistic for the Diebold-Mariano test can be written as

$$\text{DM} = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}}, \tag{40}$$

where T is the number of days in the sample and $\hat{f}_d(0)$ is a consistent estimator of the spectral density of the loss differential at frequency 0 that is commonly expressed as the weighted sum of the sample autocovariances:

$$2\pi \hat{f}_d(0) = \sum_{\tau=-(T-1)}^{T-1} 1\ (\frac{\tau}{S(T)})\ \hat{\gamma}_d(\tau), \tag{41}$$

with $\hat{\gamma}_d(\tau)$ being the autocovariance of the loss differential between the two forecasting models at lag τ (see Diebold and Mariano (2002) for a more detailed derivation of the test statistic). The null hypothesis of the Diebold Mariano test states that both forecasts have the same forecasting accuracy:

$$H_0 : E(d_t) = 0\ \forall t. \tag{42}$$

Under the null hypothesis, the test statistic in Equation 40 is asymptotically $N(0, 1)$ distributed. For this application, I will test the pairwise one-sided alternative hypothesis

$$H_1 : E(d_t) > 0, \tag{43}$$

for all possible combinations of models within a country. The Diebold-Mariano test will only be applied on a within-country basis, comparing all models for one specific country, for all three forecasting horizons. I am not using the Diebold-Mariano test to compare models across countries as the forecast error relates to the actual volatility in the respective country and comparing forecasting errors across countries would introduce substantial uncertainty regarding the interpretation of the test results.

6.4 Daily Volatility Forecast

The first part of the analysis focuses on forecasting realized volatility at the daily forecasting horizon. As described above, all forecasting models were estimated for all 31 stock market indices. I also added the results from a standard AR-1 model as an additional benchmark model. Table 2 presents the MSE for all models and across all markets. The results were divided by the MSE of the standard HAR model to facilitate the comparison between models. Hence, values below 1 indicate that the respective model outperforms the standard HAR. For each row, the table shows the result of the best-performing model printed in bold and the result of the second-best-performing model printed in italics. The first finding from Table 2 is that the C-HAR seems to be the best-performing model for forecasting daily volatility internationally. The C-HAR was the best or second-best performing model in 19 out of 31 indices. The forecasting performance of HAR-J and WLS-HAR is also noticeable and underlines the power of popular HAR extensions. It is important to note that the WLS-HAR performs particularly well when forecasting volatility for some of the main US indices, such as the S&P-500 (SPXCASH and SPXFUT) and the NASDAQ (IXIC). The results in Table 2 also illustrate that the machine learning models performed relatively poorly and were mostly outperformed by the benchmark HAR model. The poor performance of the LSTM model is particularly surprising, given its strong performance in prior studies. It fails to be the best or second-best model for a single index and only outperforms the standard HAR model for two indices.

Table 3 illustrates the performance of all models using the QLIKE as a loss function. Overall, one can observe clear deviations from the results based on the MSE. Here, the WLS-HAR emerges as the best-performing model across most markets. Again, the model repeats to show the best performance of the US indices including good accuracy for the Russel 2000 (RUT) and the Dow Jones Index (DJI). C-HAR and HAR-J perform similarly

Table 2: MSE - Daily Realized Variance Forecast
The table displays the MSE across nine models for all 31 stock market indices. The models are the benchmark HAR model, the continuous HAR, the HAR with Jumps, the weighted least squares HAR, the robust regression HAR, the support vector machine, random forest, AR-1, and the long short-term memory model. The statistics were divided by the result for the benchmark HAR model for standardization. For each row, the value with the lowest error is printed in bold and the value with the second lowest error is printed in italics. The last two rows count the total number of times that each model was the best or second best model in one of the 31 markets.

	HAR	C-HAR	HAR-J	WLS-HAR	RR-HAR	SVM	RF	AR-1	LSTM
AEX	1.00	*0.986*	**0.976**	0.998	1.02	2.184	1.121	1.117	2.072
AORD	1.00	0.967	*0.934*	**0.933**	1.008	1.72	1.008	1.087	1.565
BFX	1.00	*0.928*	**0.926**	0.983	1.005	1.947	1.062	1.105	1.706
BSESN	1.00	**0.917**	0.959	1.02	0.989	1.166	*0.951*	1.065	1.165
BVLG	1.00	0.996	*0.99*	**0.985**	1.129	3.184	1.972	1.056	2.962
BVSP	*1.00*	**0.976**	1.007	1.015	1.066	2.477	1.174	1.157	2.364
DJI	1.00	*0.943*	**0.94**	0.976	1.031	1.906	1.182	1.142	1.83
FCHI	1.00	*0.943*	**0.931**	0.994	0.99	1.897	1.064	1.122	1.673
FTMIB	**1.00**	1.048	*1.024*	1.078	1.057	2.218	1.119	1.08	2.317
FTSE	1.00	1.002	**0.988**	**0.988**	1.022	1.487	1.088	1.168	1.464
GDAXI	1.00	**0.937**	*0.946*	0.996	0.996	1.845	1.095	1.096	1.848
GSPTSE	1.00	**0.838**	1.084	0.959	*0.952*	1.118	1.021	1.052	1.091
HSI	1.00	*0.979*	**0.977**	0.986	1.036	1.832	1.164	1.141	1.614
IBEX	1.00	**0.979**	*0.99*	1.00	1.004	1.712	1.042	1.071	1.538
IXIC	1.00	*0.988*	0.995	**0.982**	0.992	1.949	1.017	1.147	1.83
KS11	1.00	1.047	1.021	**0.876**	*0.891*	1.895	1.335	1.053	1.833
KSE	**1.00**	1.044	*1.006*	1.011	1.04	1.566	1.007	1.144	1.452
MXX	1.00	**0.934**	*0.955*	0.982	1.033	1.367	1.02	1.14	1.319
N225	1.00	1.006	*0.99*	0.999	**0.988**	1.737	1.254	1.088	1.634
NSEI	1.00	**0.917**	0.96	1.021	1.005	1.197	*0.948*	1.035	1.176
OMXC20	1.00	1.232	1.203	0.997	0.984	*0.937*	**0.93**	1.171	0.979
OMXHPI	1.00	1.095	1.088	1.013	0.966	**0.902**	*0.904*	1.174	0.926
OMXSPI	1.00	**0.895**	*0.913*	1.014	0.987	0.982	0.921	1.196	1.036
OSEAX	1.00	**0.95**	*0.959*	0.991	1.014	1.29	0.999	1.259	1.256
RUT	1.00	*0.979*	**0.977**	0.99	1.038	2.356	1.209	1.115	1.767
SMSI	1.00	**0.905**	*0.908*	1.036	0.983	1.311	0.922	1.095	1.281
SPXCASH	1.00	0.992	1.01	**0.95**	*0.985*	2.381	1.175	1.136	2.232
SPXFUT	1.00	*0.987*	0.998	**0.962**	0.998	2.405	1.108	1.136	2.172
SSEC	*1.00*	1.019	1.003	**0.997**	1.056	2.014	1.096	1.127	1.599
SSMI	**1.00**	1.077	1.018	*1.001*	1.002	1.837	1.102	1.138	1.438
STOXX50E	1.00	**0.925**	*0.949*	1.197	0.973	1.474	0.988	1.13	1.408
Best Model	3	11	6	8	1	1	1	0	0
Second best	2	8	13	1	3	1	3	0	0

well, but their accuracy decreases compared to the MSE results. The RR-HAR, however, shows an improved performance based on the QLIKE loss function and is the second-best performing model for almost half of the covered indices. Furthermore, we can conclude from Table 3 that the performance of the standard HAR model is weaker based on the

QLIKE loss function. Finally, while LSTM and SVM perform poorly, we observe that the random forest improved its performance and seems to outperform the benchmark HAR model for several markets.

Table 3: QLIKE - Daily Realized Variance Forecast
The table displays the QLIKE across nine models for all 31 stock market indices. The models are the benchmark HAR model, the continuous HAR, the HAR with Jumps, the weighted least squares HAR, the robust regression HAR, the support vector machine, random forest, AR-1, and the long short-term memory model. The statistics were divided by the result for the benchmark HAR model for standardization. For each row, the value with the lowest error is printed in bold and the value with the second lowest error is printed in italic. The last two rows count the total number of times that each model was the best or second best model in one of the 31 markets.

	HAR	C-HAR	HAR-J	WLS-HAR	RR-HAR	SVM	RF	AR-1	LSTM
AEX	1.00	0.998	0.989	**0.94**	*0.966*	2.433	1.027	1.412	2.912
AORD	1.00	**0.982**	*0.997*	1.00	1.127	3.961	1.008	1.287	2.184
BFX	1.00	*0.969*	0.98	**0.933**	0.985	2.879	0.993	1.455	2.175
BSESN	1.00	0.847	0.881	**0.705**	0.76	1.489	*0.754*	1.563	1.637
BVLG	1.00	*0.989*	0.993	**0.934**	0.998	4.791	1.081	1.469	2.75
BVSP	1.00	*0.996*	**0.982**	1.009	1.012	2.518	1.094	1.223	2.479
DJI	1.00	0.991	*0.975*	**0.945**	1.01	2.414	0.992	1.441	2.521
FCHI	1.00	1.013	0.998	**0.919**	*0.953*	2.29	1.004	1.457	1.989
FTMIB	1.00	1.004	*0.988*	1.008	**0.971**	2.134	1.052	1.419	2.342
FTSE	1.00	1.017	0.99	**0.936**	0.99	2.019	*0.944*	1.508	2.652
GDAXI	1.00	1.008	0.99	**0.907**	*0.921*	1.88	0.984	1.599	2.775
GSPTSE	1.00	4.437	6.332	**0.72**	0.868	3.636	*0.747*	1.668	2.307
HSI	1.00	*0.996*	**0.99**	1.00	1.027	2.347	1.034	1.384	1.873
IBEX	1.00	1.001	*0.992*	**0.956**	1.00	2.712	1.007	1.389	2.079
IXIC	1.00	1.109	1.021	**0.878**	*0.908*	2.037	0.968	1.664	2.238
KS11	1.00	1.082	1.00	**0.889**	*0.912*	2.129	0.989	1.675	1.965
KSE	1.00	1.047	0.998	**0.945**	*0.991*	1.814	0.998	1.322	2.064
MXX	1.00	*0.982*	**0.958**	0.986	1.131	2.177	0.983	1.3	1.941
N225	1.00	1.063	1.006	**0.969**	*0.991*	2.051	1.053	1.261	2.092
NSEI	1.00	0.847	1.389	**0.787**	0.868	2.131	*0.805*	1.477	2.061
OMXC20	1.00	0.996	1.013	1.102	*0.975*	1.219	**0.96**	1.371	1.87
OMXHPI	1.00	0.928	0.945	0.797	*0.654*	1.271	**0.648**	1.694	1.238
OMXSPI	1.00	*0.761*	0.772	0.87	**0.741**	1.301	0.762	1.826	1.818
OSEAX	1.00	**0.949**	*0.958*	1.004	1.059	2.545	1.008	1.303	3.041
RUT	**1.00**	1.006	1.036	*1.004*	1.06	4.246	1.044	1.175	2.492
SMSI	1.00	1.107	1.019	**0.897**	*0.963*	2.303	0.969	1.323	2.137
SPXCASH	1.00	1.055	1.016	**0.945**	*0.947*	3.19	1.036	1.497	2.614
SPXFUT	1.00	1.04	1.018	**0.931**	*0.946*	2.609	0.99	1.516	2.287
SSEC	*1.00*	1.009	1.002	**0.966**	1.075	3.438	1.001	1.413	2.242
SSMI	1.00	1.03	1.00	**0.946**	*0.967*	3.239	1.037	1.438	4.386
STOXX50E	1.00	**0.943**	*0.948*	1.084	0.961	1.858	0.999	1.409	2.115
Best Model	1	3	3	20	2	0	2	0	0
Second best	1	6	6	1	13	0	4	0	0

Finally, Table 4 presents the performance of all models based on the MAPE. Here, a

different picture emerges, compared to the MSE and QLIKE results shown above. First, the RR-HAR is the best-performing model, showing the lowest MAPE for 28 out of 31 indices and being the second-best-performing model for all three remaining indices. Second, the WLS-HAR can be seen as the second-best overall model as it is the best or second-best performing model for 24 out of 31 models. Third, we can observe that the HAR,C-HAR and HAR-J all show quite poor performance which is probably due to the fact that all three are estimated via OLS which minimizes squared residuals whereas the MAPE is based on absolute deviations. Fourth, the MAPE is the only loss function for which the LSTM model outperforms the HAR model for most indices and is the best or second-best performing model for 4 out of 31 indices. Yet, it is worth noting that the LSTM performs well for the major stock market indices of all three Nordic EU members: the OMXC20 (OMX Copenhagen), the OMXHPI (OMX Helsinki), and the OMXSPI (OMX Stockholm). The very good performance of the LSTM model in all three markets can probably be, at least partially, explained by the likely correlation among these markets given their geographical and structural similarities. Finally, the random forest model, which performed relatively well based on QLIKE, shows a poor performance based on the MAPE.

As a final step in the analysis of daily forecasts, as described above, one-sided Diebold-Mariano tests were computed between all forecasting models for a respective index. Table 5 and Table 6 summarize the findings of the Diebold-Mariano tests by counting all Diebold-Mariano tests that were significant at the 5% or 10 % level, respectively. A count of 2 in the second row and first column of Table 5, for example, indicates that the C-HAR model was significantly more accurate than the benchmark HAR model for 2 out of the total of 31 indices. Hence, high numbers in a row signal strong model performance, whereas high numbers in a column point to poor model performance. The Diebold-Mariano test results provide several interesting additions to the findings derived from the loss functions above. First of all, very few models can significantly outperform the benchmark HAR at the 5 % level. This suggests that the differences in model performance presented above, are relatively small and often fail to be statistically significant. One additional reason for the solid performance of the benchmark HAR may be that the Diebold-Mariano test is built on squared deviations and the standard HAR model performed relatively well when measured by the MSE. Yet, at the 10 % level, the HAR is outperformed for

Table 4: MAPE - Daily Realized Variance Forecast
The table displays the MAPE across nine models for all 31 stock market indices. The models are the benchmark HAR model, the continuous HAR, the HAR with Jumps, the weighted least squares HAR, the robust regression HAR, the support vector machine, random forest, AR-1, and the long short-term memory model. The statistics were divided by the result for the benchmark HAR model for standardization. For each row, the value with the lowest error is printed in bold and the value with the second lowest error is printed in italic. The last two rows count the total number of times that each model was the best or second best model in one of the 31 markets.

	HAR	C-HAR	HAR-J	WLS-HAR	RR-HAR	SVM	RF	AR-1	LSTM
AEX	1.00	0.984	0.983	*0.825*	**0.8**	1.429	0.971	1.529	1.009
AORD	1.00	1.01	1.009	0.887	**0.762**	0.891	1.024	1.262	*0.831*
BFX	1.00	0.944	0.958	*0.826*	**0.792**	1.338	0.954	1.555	0.968
BSESN	1.00	0.818	0.83	*0.59*	**0.562**	1.011	0.652	1.718	0.599
BVLG	1.00	0.985	0.996	*0.906*	**0.857**	1.809	0.991	1.486	1.002
BVSP	1.00	*0.958*	0.965	0.979	**0.875**	1.237	1.138	1.348	1.141
DJI	1.00	0.985	0.977	*0.819*	**0.801**	1.361	0.955	1.695	0.889
FCHI	1.00	1.002	0.976	*0.801*	**0.791**	1.392	0.932	1.575	0.939
FTMIB	1.00	1.019	0.991	*0.941*	**0.838**	1.482	1.062	1.529	0.957
FTSE	1.00	0.989	0.981	*0.776*	**0.705**	1.074	0.859	1.742	0.809
GDAXI	1.00	1.035	1.003	*0.792*	**0.759**	1.491	0.946	1.795	0.895
GSPTSE	1.00	0.577	0.671	*0.518*	**0.501**	2.022	0.649	1.902	0.563
HSI	1.00	1.033	1.006	*0.921*	**0.86**	1.378	1.014	1.582	0.937
IBEX	1.00	0.997	0.984	*0.845*	**0.813**	1.447	0.953	1.485	0.974
IXIC	1.00	1.153	1.033	**0.748**	*0.777*	1.51	0.923	1.878	0.896
KS11	1.00	1.118	1.002	**0.796**	*0.798*	1.597	0.898	1.799	0.812
KSE	1.00	1.002	0.996	*0.725*	**0.689**	1.445	2.096	1.405	2.684
MXX	1.00	0.849	0.906	0.848	**0.674**	*0.737*	0.977	1.444	0.811
N225	1.00	1.154	1.034	*0.874*	**0.827**	1.504	1.062	1.459	0.942
NSEI	1.00	0.783	0.799	*0.579*	**0.566**	0.997	0.69	1.573	0.626
OMXC20	1.00	0.965	1.012	0.927	**0.606**	0.958	0.919	1.766	*0.617*
OMXHPI	1.00	0.987	0.994	0.633	*0.452*	1.165	0.548	2.035	**0.421**
OMXSPI	1.00	0.712	0.781	0.696	**0.553**	1.075	0.704	2.248	*0.553*
OSEAX	1.00	0.894	0.925	*0.878*	**0.751**	1.064	1.028	1.491	0.883
RUT	1.00	0.987	0.994	1.024	**0.807**	*0.968*	1.044	1.172	1.048
SMSI	1.00	1.127	1.04	*0.801*	**0.735**	0.919	0.945	1.397	0.961
SPXCASH	1.00	1.106	1.029	*0.824*	**0.776**	1.609	1.044	1.799	0.945
SPXFUT	1.00	1.071	1.027	*0.847*	**0.824**	1.562	0.952	1.656	0.922
SSEC	1.00	0.993	0.992	*0.877*	**0.743**	1.081	0.951	1.496	1.04
SSMI	1.00	0.983	1.001	*0.839*	**0.774**	1.292	0.956	1.547	1.05
STOXX50E	1.00	0.882	0.878	*0.782*	**0.75**	1.194	0.87	1.512	0.868
Best Model	0	0	0	2	28	0	0	0	1
Second best	0	1	0	22	3	2	0	0	3

several indices by C-HAR, WLS-HAR, HAR-J, and the random forest. Second, one can observe that these four models seem to provide a comparable level of forecasting accuracy as the null hypothesis is rejected for only a small number of pairs between these four models. The RR-HAR model, on the other hand, is less precise than the other HAR

Table 5: Diebold-Mariano Test Results, Daily Forecasts - 5% Level
The table displays a summary of all one-sided Diebold-Mariano tests carried out between all nine models and across all 31 indices. The table illustrates for how many indices, out of the total of 31, the model listed in the row was significantly more accurate than the model listed in the column at the 5% level. Appendix A.4 provides statistics on all Diebold-Mariano tests at the index level.

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0	0	10	0	6	17	5	6
C-HAR	2	-	2	11	0	5	17	4	6
WLS	2	2	-	9	1	7	18	5	6
RR	0	0	1	-	0	2	20	2	4
HAR-J	4	0	2	11	-	6	17	3	6
AR-1	0	1	0	0	0	-	17	1	4
SVM	0	0	0	0	0	0	-	0	0
RF	0	0	1	1	0	0	21	-	3
LSTM	0	0	0	0	0	0	0	0	-

Table 6: Diebold-Mariano Test Results, Daily Forecasts - 10% Level
The table displays a summary of all one-sided Diebold-Mariano tests carried out between all nine models and across all 31 indices. The table illustrates for how many indices, out of the total of 31, the model listed in the row was significantly more accurate than the model listed in the column at the 10% level. Appendix A.4 provides statistics on all Diebold-Mariano tests at the index level.

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	3	3	18	1	12	26	12	9
C-HAR	5	-	4	15	2	12	26	13	9
WLS	4	2	-	18	1	11	24	13	9
RR	0	1	1	-	0	3	27	7	9
HAR-J	7	2	4	21	-	17	26	15	9
AR-1	1	1	1	0	0	-	24	2	8
SVM	0	0	0	0	0	0	-	0	0
RF	2	1	2	4	0	1	27	-	6
LSTM	0	0	0	0	0	0	0	0	-

extensions for a larger number of markets. This, again, may be due to the fact that the RR-HAR shows inferior performance when measured by the MSE compared to the other loss functions and the Diebold-Mariano test also uses squared deviations from the true value. With respect to the machine learning models, we can observe that the SVM shows the weakest performance, whereas the random forest, at least at the 5% level, is solely being significantly outperformed in a relatively small number of markets. Finally, the Diebold-Mariano test results soften the picture of the performance of the LSTM model. From Table 5 we can see that the HAR model, at the 5% level, can only be seen as significantly more accurate than the LSTM for 6 out of 31 indices.

6.5 Weekly Volatility Forecast

Next, we turn to forecasting weekly volatility. Following the notation of the HAR model and the approach by Li and Tang (2022), I define weekly volatility as the average volatility over the next week, defined on day t:

$$RV_t^w = \frac{1}{5}(RV_{t+1}^d + RV_{t+2}^d \cdots + RV_{t+5}^d). \tag{44}$$

Here, for the sake of brevity, I will only display the results based on the MSE. The respective tables for the QLIKE and MAPE can be found in Appendix A.3. Table 7 presents the MSE for forecasting weekly realized variance for all 31 indices. The results again show a solid performance of the different HAR extensions. The HAR-J, in particular, emerges as the most potent model, being the best or second-best performing model for 21 out of 31 markets. Similarly, the WLS- and RR-HAR perform well across a range of markets. The machine learning models, however, show poor performance for most markets and fail to outperform the benchmark HAR for the majority of indices. The weekly forecasting performance seems to be slightly poorer than the daily forecasts for the machine learning models. It is worth noticing that the C-HAR's performance also declines in comparison to the daily forecasts when measured using the MSE. Furthermore, we can observe that there seems to be a small difference between the S&P-500 future and cash index as the best-performing model differs between the two markets. This is the only case across all loss functions and all forecasting horizons, where the best-performing model for the cash index is not equal to the best-performing model of the futures index.

The forecasting errors based on the QLIKE, which can be found in Appendix A.3, show a clear dominance of the WLS- and the RR-HAR models. The WLS-HAR model, in particular, outperforms all other models for 21 out of 31 indices including the major US indices. The HAR and C-HAR models perform relatively poorly and the performance of the HAR-J also deviates significantly compared to the MSE-based results in Table 7. Most of the machine learning models also perform relatively poorly. The random forest, however, emerges as the notable exception, being the second-best model for four indices and outperforming the standard HAR model for several indices.

Finally, the results as measured by the MAPE reveal a similar picture as for the daily forecasts. The RR-HAR is the best-performing model for 28 out of 31 indices. The

Table 7: MSE - Weekly Realized Variance Forecast
The table displays the MSE across the nine models for all 31 stock market indices forecasting weekly realized variance. For each row, the value with the lowest error is printed in bold and the value with the second lowest error is printed in italic. The last two rows count the total number of times that each model was the best or second best model in one of the 31 markets. Comparable tables for the QLIKE and the MAPE loss functions can be found in Appendix A.3.

	HAR	C-HAR	HAR-J	WLS-HAR	RR-HAR	SVM	RF	AR-1	LSTM
AEX	**1.00**	1.034	*1.006*	1.007	1.018	3.572	1.366	1.208	3.464
AORD	1.00	0.957	**0.954**	*0.956*	1.027	3.424	1.496	1.278	2.828
BFX	1.00	*0.959*	**0.957**	1.002	1.02	3.415	1.422	1.182	3.024
BSESN	1.00	**0.904**	*0.948*	1.024	0.972	1.601	1.005	1.175	1.645
BVLG	*1.00*	1.013	**0.996**	1.004	1.044	4.104	1.903	1.047	3.82
BVSP	1.00	*0.977*	**0.968**	0.991	1.037	3.967	1.533	1.241	3.612
DJI	1.00	0.991	**0.962**	*0.978*	*0.978*	3.255	1.637	1.195	3.093
FCHI	1.00	1.02	*0.997*	1.014	**0.996**	3.21	1.349	1.218	3.079
FTMIB	**1.00**	1.08	*1.011*	1.052	1.019	3.054	1.284	1.078	3.288
FTSE	1.00	1.07	*0.994*	**0.979**	1.005	2.837	1.526	1.283	2.707
GDAXI	1.00	0.976	*0.969*	1.007	**0.964**	2.628	1.145	1.179	2.604
GSPTSE	1.00	**0.682**	*0.825*	0.98	0.92	1.615	1.207	1.189	1.534
HSI	1.00	0.99	**0.971**	0.977	*0.972*	3.177	1.48	1.254	2.534
IBEX	1.00	**0.981**	*0.983*	1.035	0.997	2.56	1.109	1.15	2.312
IXIC	1.00	1.033	1.00	*0.983*	**0.961**	3.056	1.175	1.194	3.089
KS11	1.00	1.098	1.031	*0.854*	**0.804**	2.735	1.53	1.03	3.153
KSE	*1.00*	1.178	**0.995**	1.467	1.017	2.815	1.111	1.31	2.492
MXX	1.00	*0.939*	**0.93**	0.955	0.987	2.339	1.253	1.226	2.139
N225	1.00	1.048	0.995	*0.966*	**0.935**	2.626	1.415	1.137	2.556
NSEI	1.00	**0.875**	2.783	0.973	*0.947*	1.613	0.984	1.147	1.56
OMXC20	1.00	1.762	1.304	0.97	*0.942*	1.027	**0.877**	1.429	1.196
OMXHPI	1.00	1.277	1.166	0.994	*0.927*	0.928	**0.825**	1.47	1.041
OMXSPI	1.00	0.965	**0.823**	0.997	*0.895*	1.269	0.964	1.522	1.41
OSEAX	1.00	*0.892*	**0.888**	0.952	0.995	2.514	1.434	1.551	2.406
RUT	1.00	1.007	*0.987*	**0.978**	1.001	3.87	1.679	1.125	3.522
SMSI	1.00	1.024	**0.922**	1.076	*0.99*	2.056	0.995	1.19	1.949
SPXCASH	1.00	0.998	1.014	**0.956**	*0.979*	4.277	1.623	1.244	3.989
SPXFUT	1.00	1.014	1.009	*0.986*	**0.985**	4.178	1.5	1.209	4.164
SSEC	*1.00*	1.061	1.015	**0.99**	1.026	3.346	1.317	1.219	2.409
SSMI	1.00	1.273	1.038	*0.996*	**0.941**	3.257	1.592	1.201	3.137
STOXX50E	1.00	1.045	*0.973*	1.117	**0.968**	2.601	1.361	1.184	2.491
Best Model	2	4	11	4	8	0	2	0	0
Second best	3	4	10	7	7	0	0	0	0

WLS-HAR emerges as the best-performing model for the remaining three indices and as the second-best model for 25 indices. The other HAR models, however, show relatively poor performance. Similarly, the machine learning models are outperformed by the WLS- and RR-models. Yet, the random forest and the LSTM model outperform the standard HAR model for a significant number of markets, consistent with the findings for the daily forecasts. Table 16 in Appendix A.3 summarizes all results.

Table 8: Diebold-Mariano Test Results, Weekly Forecasts - 5% Level
The table displays a summary of all one-sided Diebold-Mariano tests carried out between all nine models and across all 31 indices for weekly realized variance forecasts. The table illustrates for how many indices, out of the total of 31, the model listed in the row was significantly more accurate than the model listed in the column at the 5% level. Appendix A.4 provides statistics on all Diebold-Mariano tests at the index level.

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	4	2	14	0	10	28	20	24
C-HAR	3	-	1	7	3	7	29	18	25
WLS	9	4	-	15	3	11	28	18	24
RR	0	1	1	-	0	7	29	18	24
HAR-J	5	4	2	17	-	10	28	21	25
AR-1	0	0	0	0	0	-	23	9	23
SVM	0	0	0	0	0	0	-	0	0
RF	0	0	0	0	0	0	30	-	19
LSTM	0	0	0	0	0	0	0	0	-

Table 9: Diebold-Mariano Test Results, Weekly Forecasts - 10% Level
The table displays a summary of all one-sided Diebold-Mariano tests carried out between all nine models and across all 31 indices forecasting weekly realized variance. The table illustrates for how many indices, out of the total of 31, the model listed in the row was significantly more accurate than the model listed in the column at the 10% level. Appendix A.4 provides statistics on all Diebold-Mariano tests at the index level.

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	5	3	16	1	12	28	21	25
C-HAR	6	-	5	14	4	12	29	22	27
WLS	13	8	-	16	10	16	28	19	25
RR	2	4	1	-	2	10	29	21	25
HAR-J	10	6	5	21	-	14	28	23	27
AR-1	0	1	0	1	1	-	23	14	23
SVM	0	0	0	0	0	0	-	0	0
RF	0	0	3	1	0	0	30	-	24
LSTM	0	0	0	0	0	0	0	0	-

Tables 8 and 9 show the summary result for the Diebold-Mariano tests across all indices at the 5% and 10% level, respectively. First, we observe that the machine learning models, especially the LSTM and the SVM, are significantly outperformed by most of the HAR variations. Second, at the 5% level, the differences in forecasting accuracy between the HAR models are mostly not significant. Here, the RR-HAR model is the exception. Its forecasts are significantly less accurate than the other HAR models for several indices. Third, at the 10% level, one can contemplate that the WLS model and the HAR-J model seem to perform particularly well, whereas the other HAR extensions are outperformed more frequently.

6.6 Monthly Volatility Forecast

Finally, I apply all models to forecast monthly volatility. Monthly volatility will be defined as in the HAR model. Hence, I forecast the average volatility over the next month, starting from day t:

$$RV_t^m = \frac{1}{22}(RV_{t+1}^d + RV_{t+2}^d \cdots + RV_{t+22}^d). \tag{45}$$

Table 10 presents the results based on the MSE. Here, we observe three important findings: first, there is a lot of heterogeneity within the performance of the HAR models, with some models performing relatively well and others performing relatively poorly. The standard HAR model, for example, fails to be the best-performing model for a single index. Second, one can contemplate that the WLS-HAR, which was one of the best-performing models for forecasting daily and weekly volatility, performs surprisingly poorly when forecasting monthly volatility. Moreover, it fails to outperform the benchmark HAR for most of the indices. Third, we observe that the performance of the machine learning models improves significantly compared to the daily and weekly forecasts. The random forest, in particular, emerges as one of the best-performing models.

The monthly forecast performance, based on the QLIKE, which can be found in Appendix A.3, reveals a similar picture.[8] Yet, while there is again a lot of variation in the performance of the different HAR models, the WLS-HAR shows strongly improved performance and emerges as the best-performing model across markets. The standard HAR, again, performs relatively poorly. Within the machine learning models, we observe that the random forest performs well, while the LSTM and the SVM fail to outperform the benchmark HAR model for a single index.

Finally, when measuring the forecast error using the MAPE, as illustrated in Appendix A.3, we observe a different picture. As for the daily and weekly forecasts, the RR-HAR emerges clearly as the best-performing model, outperforming the benchmark HAR for every single index and being the best model for 25 out of 31 models. However, we also observe a strongly improved forecast accuracy of the LSTM, which can be seen as the second-best overall model based on the MAPE. In addition to these two models, only the WLS-HAR shows good forecasting performance for the monthly volatility.

[8] As for the weekly variance forecasts, only the table for the MSE is presented in the main text. The tables for the QLIKE and the MAPE can be found in Appendix A.3.

Table 10: MSE - Monthly Realized Variance Forecast
The table displays the MSE across nine models for all 31 stock market indices forecasting monthly realized variance. The statistics were divided by the result for the benchmark HAR model for standardization. For each row, the value with the lowest error is printed in bold and the value with the second lowest error is printed in italic. The last two rows count the total number of times that each model was the best or second best model in one of the 31 markets.

	HAR	C-HAR	HAR-J	WLS-HAR	RR-HAR	SVM	RF	AR-1	LSTM
AEX	1.00	*0.954*	0.986	1.129	0.977	1.61	**0.943**	1.11	1.65
AORD	1.00	0.987	0.988	0.964	**0.932**	1.414	*0.934*	0.945	1.164
BFX	1.00	*0.966*	0.974	1.093	0.996	1.631	**0.963**	1.097	1.481
BSESN	1.00	*0.917*	0.95	1.006	0.969	1.185	**0.87**	1.225	1.18
BVLG	1.00	1.019	1.001	0.975	*0.841*	1.177	**0.747**	0.857	1.123
BVSP	1.00	0.96	0.979	**0.91**	*0.933*	1.764	1.19	1.057	1.771
DJI	*1.00*	1.051	1.01	1.022	**0.939**	1.743	1.154	1.111	1.584
FCHI	1.00	*0.984*	1.011	1.147	0.994	1.634	**0.968**	1.157	1.573
FTMIB	1.00	1.023	0.999	0.987	**0.959**	1.265	*0.972*	0.992	1.419
FTSE	1.00	**0.97**	0.994	1.106	*0.987*	1.616	1.013	1.184	1.442
GDAXI	1.00	1.036	*0.999*	1.076	**0.959**	1.603	1.006	1.253	1.644
GSPTSE	1.00	0.499	0.885	0.541	**0.214**	0.251	0.253	1.08	*0.228*
HSI	1.00	1.021	1.003	1.00	**0.888**	1.885	*0.95*	1.198	1.489
IBEX	1.00	*0.973*	0.997	1.148	1.005	1.565	**0.925**	1.157	1.336
IXIC	1.00	*0.954*	1.004	1.053	0.957	1.446	**0.918**	1.16	1.402
KS11	1.00	1.031	1.015	1.00	**0.88**	1.414	*0.974*	1.053	1.391
KSE	1.00	1.027	1.002	2.346	*0.99*	1.669	**0.962**	1.25	1.443
MXX	1.00	**0.858**	0.939	0.958	*0.91*	1.548	1.00	1.13	1.309
N225	1.00	*0.993*	1.00	1.06	**0.967**	1.608	1.114	1.145	1.57
NSEI	1.00	*0.865*	1.672	1.083	0.948	1.193	**0.837**	1.215	1.055
OMXC20	1.00	1.808	1.427	1.117	*0.838*	**0.723**	0.881	1.68	1.288
OMXHPI	1.00	1.304	1.12	1.108	0.805	**0.61**	0.7	1.824	*0.675*
OMXSPI	1.00	0.89	0.874	1.024	**0.855**	*0.872*	0.98	1.591	1.031
OSEAX	1.00	**0.901**	*0.942*	0.98	0.951	1.508	0.958	1.309	1.448
RUT	*1.00*	1.03	1.006	1.129	**0.98**	1.836	1.03	1.062	1.657
SMSI	1.00	1.041	*0.983*	1.162	1.027	1.359	**0.886**	1.158	1.324
SPXCASH	*1.00*	1.037	1.012	1.016	**0.924**	2.012	1.099	1.064	2
SPXFUT	1.00	1.047	*0.997*	0.998	**0.951**	1.971	1.152	1.092	1.907
SSEC	*1.00*	1.002	1.005	1.022	1.019	2.289	**0.99**	1.37	1.529
SSMI	1.00	0.972	1.026	1.093	*0.962*	1.437	**0.942**	1.024	1.368
STOXX50E	1.00	**0.982**	0.989	1.691	0.983	1.467	0.989	1.156	1.521
Best Model	0	4	0	1	12	2	12	0	0
Second best	4	8	4	0	8	1	4	0	2

Tables 11 and 12 display a summary of the one-sided Diebold-Mariano tests between all models forecasting monthly volatility for the 5% and 10% level, respectively. Overall, the results seem to confirm the findings based on the MSE, as discussed above. The random forest seems to be the best-performing model based on the one-sided Diebold-Mariano tests, showing a significantly higher forecasting accuracy than most other models for almost half of the indices. Within the different HAR models, the standard HAR, the

Table 11: Diebold-Mariano Test Results, Monthly Forecasts - 5% Level
The table displays a summary of all one-sided Diebold-Mariano tests carried out between all nine models and across all 31 indices for monthly realized variance forecasts. The table illustrates for how many indices, out of the total of 31, the model listed in the row was significantly more accurate than the model listed in the column at the 5% level. Appendix A.4 provides statistics on all Diebold-Mariano tests at the index level.

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	5	10	12	0	10	25	6	22
C-HAR	7	-	10	12	5	9	26	3	25
WLS	5	5	-	4	3	8	23	5	19
RR	4	6	6	-	2	7	27	7	21
HAR-J	3	6	9	12	-	12	26	7	24
AR-1	2	2	2	1	2	-	19	0	13
SVM	1	1	2	0	1	0	-	0	1
RF	13	14	14	16	12	8	28	-	25
LSTM	1	1	1	0	1	0	1	0	-

Table 12: Diebold-Mariano Test Results, Monthly Forecasts - 10% Level
The table displays a summary of all one-sided Diebold-Mariano tests carried out between all nine models and across all 31 indices forecasting monthly realized variance. The table illustrates for how many indices, out of the total of 31, the model listed in the row was significantly more accurate than the model listed in the column at the 10% level. Appendix A.4 provides statistics on all Diebold-Mariano tests at the index level.

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	7	12	13	2	15	25	6	25
C-HAR	9	-	14	14	7	14	26	6	26
WLS	7	6	-	5	3	9	23	7	21
RR	5	7	12	-	5	10	27	8	25
HAR-J	9	7	12	13	-	15	26	8	26
AR-1	2	2	5	1	2	-	19	0	15
SVM	1	1	4	0	1	0	-	1	2
RF	15	15	18	16	13	12	28	-	27
LSTM	1	1	2	0	1	0	2	0	-

C-HAR and the HAR-J all show comparable forecasting accuracy. The SVM and the LSTM, however, show relatively poor performance based on the Diebold-Mariano test. Interestingly, the WLS-HAR and the RR-HAR do not seem to perform particularly well, based on the Diebold-Mariano tests. This finding, as well as the poor performance of SVM and LSTM, stand somewhat in contrast to the results based on the three loss functions as reported above.

7 Robustness Check - LSTM Architecture

Except for the monthly forecasts, the results for the different loss functions and different forecasting horizons revealed a relatively poor performance of the LSTM model, which stands in contrast to prior research on the model (see for example Bucci, 2020 or Rahimikia and Poon, 2023). This could be due to the different estimation approach taken in this thesis, or due to differences in the architecture of the LSTM. Only for the monthly variance forecast and using MAPE as a loss function, the LSTM emerges as one of the more powerful models. Given this finding of my thesis, I will use this final section to investigate if changes in the LSTM's architecture can improve its performance. Some prior studies on the LSTM have compared many different architectures to derive the best-performing model (see for example Rahimikia and Poon, 2023). Yet, as mentioned before, this approach is not feasible within the scope of this thesis as I investigate the role of the LSTM across 31 markets using monthly re-estimation. Thus, in this section, I will investigate potential improvements of the LSTM for three markets in which its performance was particularly inferior to other forecasting techniques. For this exercise, I will focus on the results for the daily variance forecast measured by the MSE as provided in Table 2. I will investigate changes in the LSTM architecture for the Euronext Amsterdam (AEX), the Bovespa Index (BVSP), and the S&P-500 Cash index (SPXCASH).[9]

The first change in the LSTM architecture will be to evaluate the performance for all three markets using the mean squared loss function instead of the mean absolute loss function as described in Section 5.3. This is an intuitive first step given that I am investigating the effect of changes to the LSTM's architecture measured by the MSE. As a second step, I increase the dropout rate of the model, which can be seen as an effective way to avoid overfitting of the model (Srivastava et al., 2014). The LSTM models estimated in Section 6 were based on a drop-out rate of 20%, meaning that 20% of the data points were randomly excluded from the training of the model. Selecting a drop-out rate between 20% and 50% is a common choice in the literature. In fact, Rahimikia and Poon (2023)

[9]Please note that, even though the LSTM performance for both indices was also relatively poorly, I decided not to investigate potential improvements for the BVLG (Portuguese Stock Index) or the FTMIB (FTSEI MIB Italy) as there was no complete data for both indices. Hence, the underperformance of the LSTM for these indices may be partly due to the smaller sample size which poses a challenge for neural networks.

Table 13: LSTM Hyperparameters - Robustness Check
The table displays the effects of different changes to the hyperparameters of the LSTM model estimated in Section 6 for three stock markets: Euronext Amsterdam (AEX), Bovespa Sao Paolo (BVSP), and the S&P-500 Cash index (SPXCASH). The numbers presented show the MSE for the 1-day ahead realized variance forecast and were divided by the MSE of the standard HAR model for comparison. The first row presents the MSE as presented in Section 6 using the mean absolute error as a loss function, a dropout rate of 20%, and 10 epochs to train the model.

	AEX (Netherlands)	BVSP (Brazil)	SPXCASH (USA)
Initial Loss	2.072	2.364	2.232
MSE loss function	2.069	2.302	2.006
Drop out 35%	1.967	2.488	2.076
Drop out 50%	1.953	2.347	2.141
Epochs 20	1.997	2.553	1.964
Epochs 30	1.985	2.430	2.101
Epochs 50	1.862	2.365	1.993

use a dropout rate of 50%.[10] Hence, given that the drop-out rate used in Section 6 is at the lower bound of the recommended parameters, I increase it first to 30% and then to 50% to investigate its effect on the LSTM's forecasting performance. Finally, I increase the number of epochs for the LSTM model. This hyperparameter defines how often the model sees the entire data set. This parameter is computationally expensive, especially when a model is frequently re-estimated as in the case of this thesis. Therefore, the LSTM models estimated in Section 6 were estimated based on 10 epochs. To investigate whether the LSTM's forecasting performance is robust to the number of epochs I will increase the number gradually to 20, 30, and 50, where the latter is the number of epochs used in Rahimikia and Poon (2023). Table 13 summarizes the results of the different adjustments for the three selected stock markets.

The results of the robustness check reveal several interesting findings. First, as expected, training the LSTM model with the MSE as a loss function improves the loss of the out-of-sample forecasts. The reduction in the loss, however, is relatively small and probably comes at the expense of a higher loss if measured via QLIKE or MAPE. Second,

[10]Bucci (2020), unfortunately, does not report the drop-out rate used for his analysis.

one can observe that the effect of an increase in the dropout rate differed across markets. A higher dropout rate strictly reduces the MSE for the AEX, it increases the loss for the SPXCASH, and the results are mixed for the BVSP. Third, increasing the number of epochs decreases the loss across all markets. For the SPXCASH, however, the impact becomes weaker with a further increase in epochs. Here, it is important to note that the dropout rate was kept at 20% for the last exercise given the mixed results of an increased dropout rate and to distinguish the effect of a higher number of epochs from the effect of the higher dropout rate. Hence, different combinations of the dropout rate and the number of epochs could potentially yield distinct results.

Overall, one can conclude that the LSTM architecture shows limited robustness: on the one hand, the MSE changes when the hyperparameters are modified, especially when increasing the number of epochs. On the other hand, however, the reductions in the MSE are relatively small and the LSTM is still significantly outperformed by the benchmark HAR model. A further improvement of the LSTM may be possible when experimenting with further changes in the LSTM architecture, especially with combinations of different hyperparameters as done, for example, in Rahimikia and Poon (2023). Yet, this procedure, similar to a full grid search, is computationally expensive and weakens the comparison of the LSTM's performance across markets as the hyperparameter tuning would need to be done individually for each market.

Lastly, it is essential to recognize that for this thesis I trained the LSTM model only with the lagged values of daily realized variances. Other studies on the LSTM use additional data to train the model such as the dividend yield, the market excess return or the default spread (Bucci, 2020). Christensen et al. (2023) also find that their neural network performs better when taking into account additional data such as the 1-week momentum factor or the CBOE VIX. Similarly, Rahimikia and Poon (2023) add news and market structure variables, such as the bid-ask spread, to the training data. The good performance of neural networks in these studies, against the backdrop of the relatively poor performance of the LSTM in this thesis, underlines that the LSTM may need larger and more diverse training datasets to use its full potential.

8 Conclusion

Within the scope of this thesis, I compared 8 different models for forecasting realized volatility and investigated their performance across 31 international stock markets. The forecasting models included several HAR as well as machine learning models. The models were re-estimated monthly and their performance was tested via recursive backtesting, hence, applying a quasi out-of-sample approach. I measured the models' forecasting performance for daily, weekly, and monthly forecasting horizons, and using three different loss functions: MSE, QLIKE, and MAPE.

For the daily variance forecasts, I contemplate a solid performance of different HAR extensions such as the C-HAR or the WLS-HAR. In contrast, the performance of the machine learning models is relatively poor but improves for loss functions different from the MSE. With respect to the weekly forecasts, the picture is relatively similar: the HAR extensions show lower forecasting errors across markets. The exact ranking of the models, however, differs compared to the daily forecasts. Again, the machine learning models perform relatively poorly based on the MSE but can improve their performance when looking at other loss functions and can outperform the benchmark HAR for several markets. Finally, for the monthly volatility forecasts, the performance of the machine learning models increases significantly, especially for the random forest and the LSTM. Within the different HAR models, we observe a decreased forecasting accuracy for several models.

Given these results and the overall approach of this thesis, there are a few caveats that shall be pointed out. First, in contrast to other studies on single stock markets, the hyperparameters of the machine learning models were determined ex-ante and not based on model performance or the outcome of a comprehensive grid search. This was not feasible given the monthly re-estimation of the models and the large number of stock market indices. Yet, the robustness check in Section 7 indicates that changes to the LSTM's architecture may lead to improved performance. The ex-ante selection of hyperparameters may, at least partly, explain the relatively poor performance of the LSTM compared to prior studies (for example Bucci, 2020). Similarly, the performance of the SVM may be improved through using different algorithms or selecting different hyperparameters given its particularly poor performance in this thesis. Second, as mentioned in Section

7, I trained the LSTM model solely on lagged values of realized volatility. Prior studies, however, frequently add additional variables when training the LSTM. This indicates that the LSTM may perform better when trained on a larger, more diverse data set. Finally, the monthly re-estimation of parameters may favor the WLS- and RR-HAR models given that prior research showed that their parameters vary less than the parameters of other HAR extensions (Clements and Preve, 2021). This may be one reason for their consistent performance across most markets in this thesis.

Based on these caveats and the results presented, I identify three possible lines of future research: first, more research on hyperparameter tuning is needed, especially in the context of large sample sizes and frequent re-estimation of models. Going forward, there can be two possible solutions for this: conducting grid-search or extensive hyperparameter tuning for each re-estimation, which would probably improve performance but requires vast computational power and risks overfitting, especially in the context of volatility forecasting where adapting to new data is crucial. Another solution that would be worth investigating is determining effective heuristics for hyperparameter selection that work well across markets and forecasting horizons. A second field for future research are the economic and financial reasons for the differences in model performance across markets. Even though there seemed to be a general trend for the different forecasting horizons and loss functions, there were still significant differences across markets. Investigating whether these differences can be explained by idiosyncratic factors such as index composition, market structure, liquidity, etc., is of great interest to improve volatility forecasting and select the right model for traders and risk managers who are active across international markets. A third and final direction for potential research is the interaction between international stock markets and the question of whether volatility in one market can help forecast future volatility in other markets. Recent research has emphasized the role of US stock markets in forecasting realized volatility across international markets (Buncic and Gisler, 2016; Liang et al., 2022). More research on cross-market dependencies may help to better understand model performance across markets and to improve volatility forecasting.

References

Aggarwal, C. C. et al. (2018). Neural networks and deep learning. *Springer*, 10(978):3.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, pages 885–905.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics*, 89(4):701–720.

Audrino, F. and Knaus, S. D. (2016). Lassoing the har model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*, 35(8-10):1485–1521.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, 2(1):1–37.

Bartz, E., Bartz-Beielstein, T., Zaefferer, M., and Mersmann, O. (2023). *Hyperparameter Tuning for Machine and Deep Learning with R: A Practical Guide*. Springer Nature.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18.

Bucci, A. (2020). Realized volatility forecasting with neural networks. *Journal of Financial Econometrics*, 18(3):502–531.

Buncic, D. and Gisler, K. I. (2016). Global equity market volatility spillovers: A broader role for the united states. *International Journal of Forecasting*, 32(4):1317–1339.

Carr, P., Wu, L., and Zhang, Z. (2019). Using machine learning to predict realized variance. *arXiv preprint arXiv:1909.10035*.

Cherkassky, V. and Ma, Y. (2004). Practical selection of svm parameters and noise estimation for svm regression. *Neural networks*, 17(1):113–126.

Christensen, K., Siggaard, M., and Veliyev, B. (2023). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, 21(5):1680–1727.

Clements, A. and Preve, D. P. (2021). A practical guide to harnessing the har volatility model. *Journal of Banking & Finance*, 133:106285.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F. and Reno, R. (2009). Har volatility modelling with heterogeneous leverage and jumps. *Available at SSRN*, 1316953.

Degiannakis, S. (2018). Multiple days ahead realized volatility forecasting: Single, combined and average forecasts. *Global Finance Journal*, 36:41–61.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144.

Engle, R. (2002). New frontiers for arch models. *Journal of applied econometrics*, 17(5):425–446.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007.

Filipović, D. and Khalilzadeh, A. (2021). Machine learning for predicting stock return volatility. *Swiss Finance Institute Research Paper*, (21-95).

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* sage publications, Inc.

Gong, X.-L., Liu, X.-H., Xiong, X., and Zhuang, X.-T. (2019). Forecasting stock volatility process using improved least square support vector machine approach. *Soft Computing*, 23:11867–11881.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Horpestad, J. B., Lyócsa, Š., Molnár, P., and Olsen, T. B. (2019). Asymmetric volatility in equity markets around the world. *The North American Journal of Economics and Finance*, 48:540–554.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101.

Kim, D. and Shin, M. (2023). Volatility models for stylized facts of high-frequency financial data. *Journal of Time Series Analysis*, 44(3):262–279.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kourtis, A., Markellos, R. N., and Symeonidis, L. (2016). An international comparison of implied, realized, and garch volatility forecasts. *Journal of Futures Markets*, 36(12):1164–1193.

Li, S. Z. and Tang, Y. (2022). Automated risk forecasting. *Available at SSRN 3776915*.

Liang, C., Wei, Y., Lei, L., and Ma, F. (2022). Global equity market volatility forecasting: new evidence. *International Journal of Finance & Economics*, 27(1):594–609.

Liu, Y. (2019). Novel volatility forecasting using deep learning–long short term memory recurrent neural networks. *Expert Systems with Applications*, 132:99–109.

Luong, C. and Dokuchaev, N. (2018). Forecasting of realised volatility with the random forests algorithm. *Journal of Risk and Financial Management*, 11(4):61.

Mandelbrot, B. B. and Mandelbrot, B. B. (1963). *The variation of certain speculative prices*. The University of Chicago Press.

McAleer, M. and Medeiros, M. C. (2008). Realized volatility: A review. *Econometric reviews*, 27(1-3):10–45.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, pages 347–370.

Poon, S.-H. and Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of economic literature*, 41(2):478–539.

Pradhan, R. P., Arvin, M. B., and Ghoshray, A. (2015). The dynamics of economic growth, oil prices, stock market depth, and other macroeconomic variables: Evidence from the g-20 countries. *International Review of Financial Analysis*, 39:84–95.

Probst, P., Wright, M. N., and Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, 9(3):e1301.

Rahimikia, E. and Poon, S.-H. (2020). Alternative data for realised volatility forecasting: Limit order book and news stories. *Available at SSRN*, 3684040.

Rahimikia, E. and Poon, S.-H. (2023). Machine learning for realised volatility forecasting. *Available at SSRN*, 3707796.

Schölkopf, B. and Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14:199–222.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Witten, D. and James, G. (2013). *An introduction to statistical learning with applications in R*. springer publication.

A Appendix

A.1 Additional Information on stock indices

Table 14: Stock Indices - Additional Information
The table presents the official abbreviation of the index as used in the tables in the main text of the thesis. Additionally, the table provides the official name of the thesis, the country of origin, the number of stocks included in the index, as well as the start and end date of the sample based on the respective index. * The number of companies had to be estimated by the author as no official number is provided by the exchange.

Index	Name	Country	Stocks	Start	End
AEX	Euronext Amsterdam	Netherlands	25	03/01/2000	25/02/2022
AORD	S&P/ASX All Ordinaries	Australia	500	04/01/2000	25/02/2022
BFX	Euronext Brussels	Belgium	20	03/01/2000	25/02/2022
BSESN	S&P BSE SENSEX	India	30	03/01/2000	25/02/2022
BVLG	Portuguese Stock Index	Portugal	37	15/10/2012	25/02/2022
BVSP	Bovespa Index	Brazil	87	03/01/2000	25/02/2022
DJI	Dow Jones Index	USA	30	03/01/2000	25/02/2022
FCHI	CAC 40	France	40	03/01/2000	25/02/2022
FTMIB	FTSE MIB	Italy	40	01/06/2009	25/02/2022
FTSE	FTSE 100 Index	UK	100	04/01/2000	25/02/2022
GDAXI	German DAX	Germany	40	03/01/2000	25/02/2022
GSPTSE	S&P/TSX Composite	Canada	225	02/05/2002	25/02/2022
HSI	Hang Seng Index	Hongkong	50	03/01/2000	25/02/2022
IBEX	IBEX 35	Spain	35	03/01/2000	25/02/2022
IXIC	NASDAQ Composite	USA	3414	03/01/2000	25/02/2022
KS11	KOSPI Composite Index	South Korea	755	04/01/2000	25/02/2022
KSE	Karachi 30 Index	Pakistan	30	03/01/2000	25/02/2022
MXX	IPC Mexico	Mexico	35	03/01/2000	25/02/2022
N225	Nikkei	Japan	225	02/02/2000	25/02/2022
NSEI	Nifty 50	India	50	03/01/2000	25/02/2022
OMXC20	OMX Copenhagen	Denmark	20	03/10/2005	25/02/2022
OMXHPI	OMX Helsinki	Finland	30	03/10/2005	25/02/2022
OMXSPI	OMX Stockholm	Sweden	30	03/10/2005	25/02/2022
OSEAX	Oslo Bors Index	Norway	69	03/09/2001	25/02/2022
RUT	Russel 2000	USA	2000	03/01/2000	25/02/2022
SMSI	Madrid SE General Index	Spain	122	04/07/2005	25/02/2022
SPXCASH	S&P-500 Cash	USA	500	03/01/2000	27/07/2020
SPXFUT	S&P-500 Futures	USA	500	03/01/2000	31/12/2020
SSEC	SSE Composite Index	China	2000*	04/01/2000	25/02/2022
SSMI	Swiss Market Index	Switzerland	20	04/01/2000	25/02/2022
STOXX50E	Stoxx 50 Europe	Europe	50	03/01/2000	25/02/2022

A.2 Realized Volatility Graphs

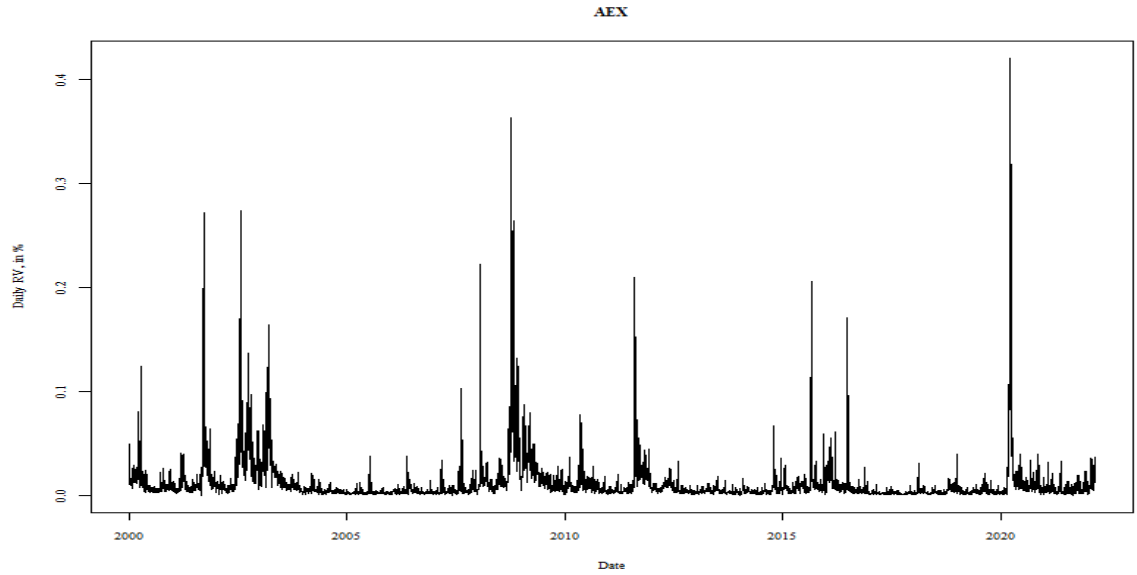

Figure 6: Realized Volatility - AEX (Euronext Amsterdam)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

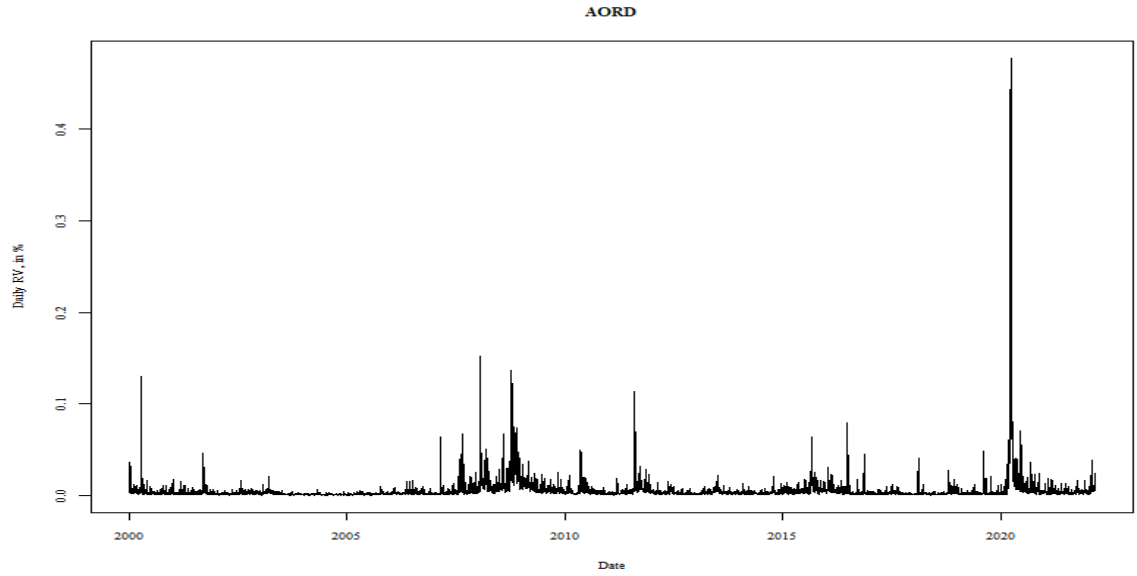

Figure 7: Realized Volatility - AORD (Australia)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

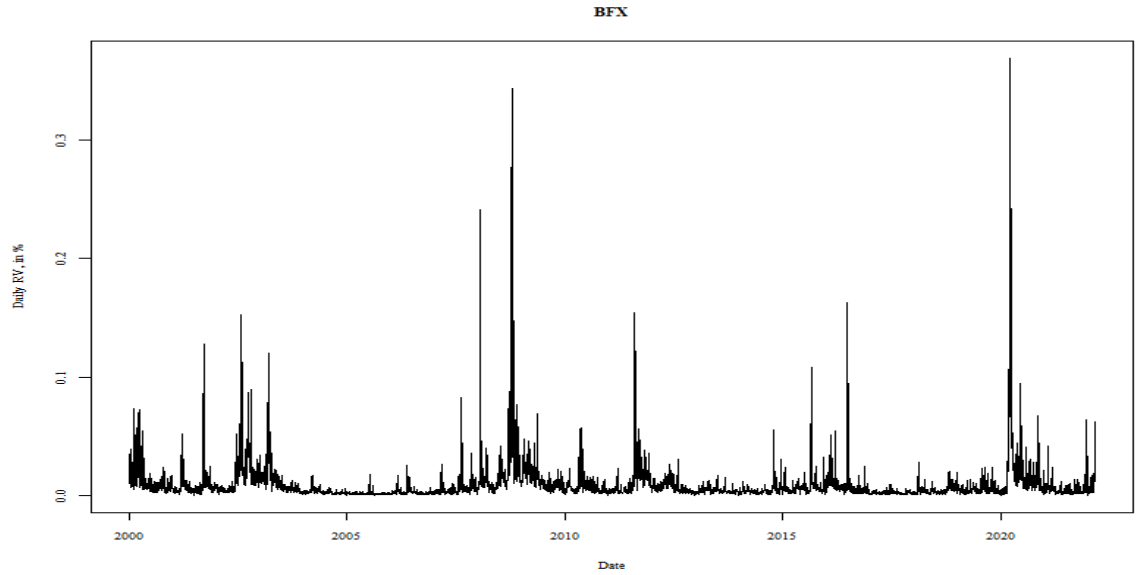

Figure 8: Realized Volatility - BFX (Belgium)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

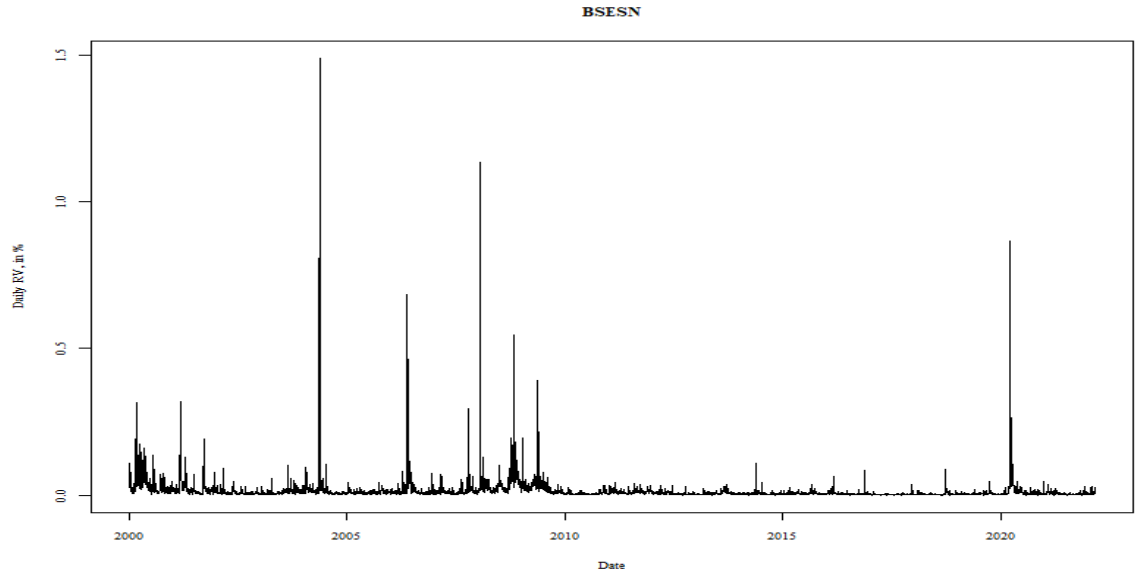

Figure 9: Realized Volatility - BSESN (India)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

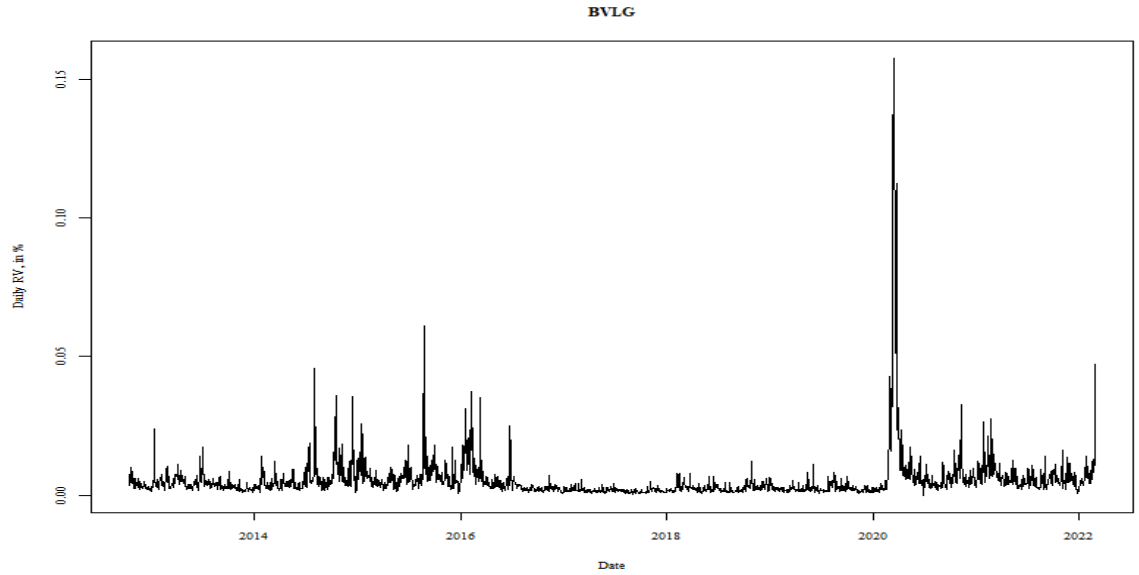

Figure 10: Realized Volatility - BVLG (Portugal)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

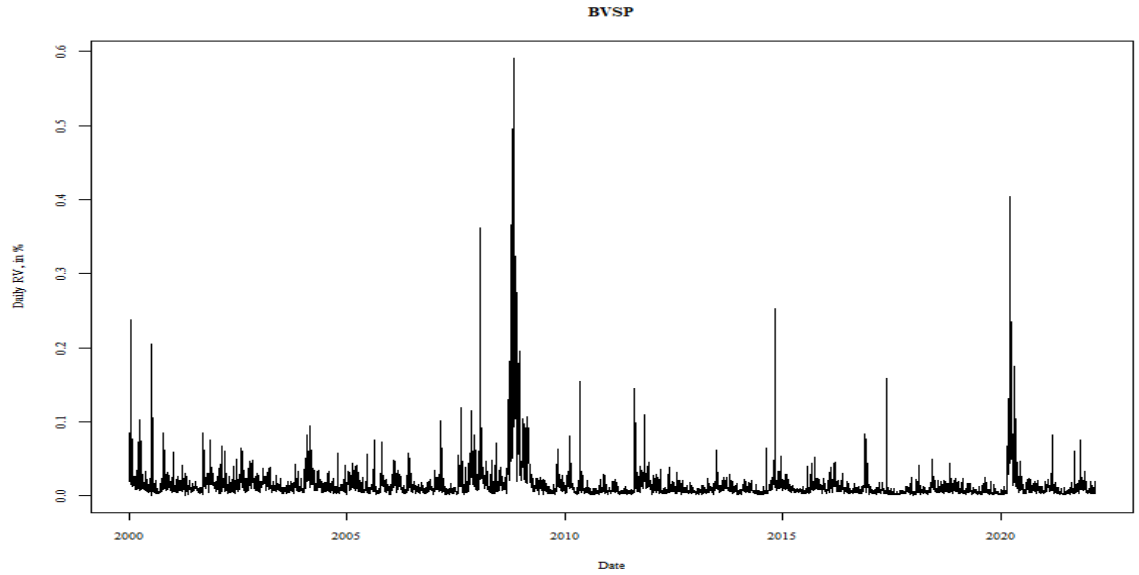

Figure 11: Realized Volatility - BVSP (Brazil)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

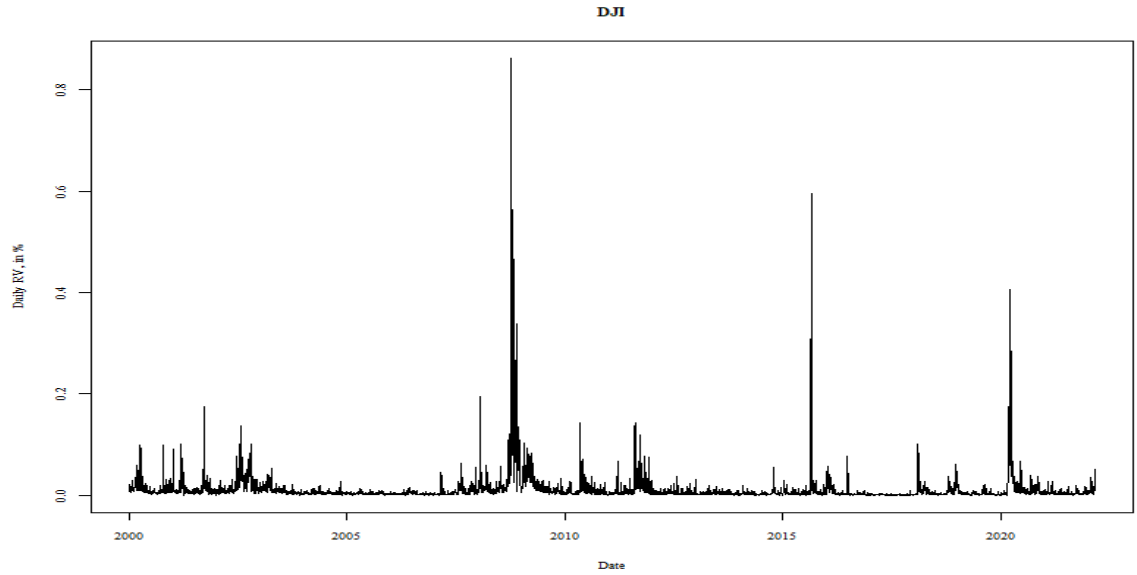

Figure 12: Realized Volatility - DJI (Dow-Jones, USA)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

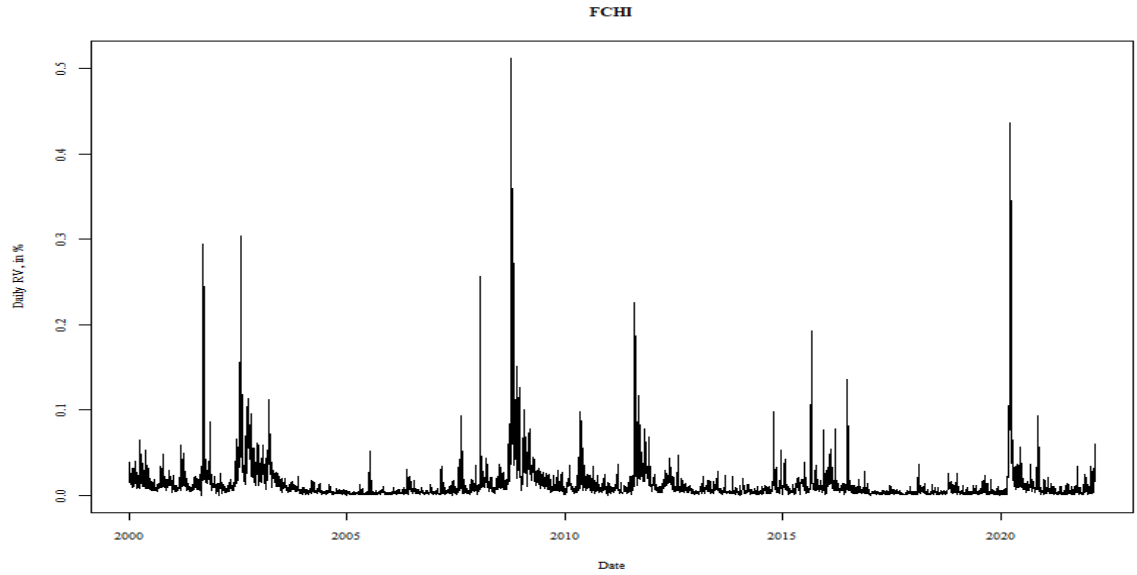

Figure 13: Realized Volatility - FCHI (France)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

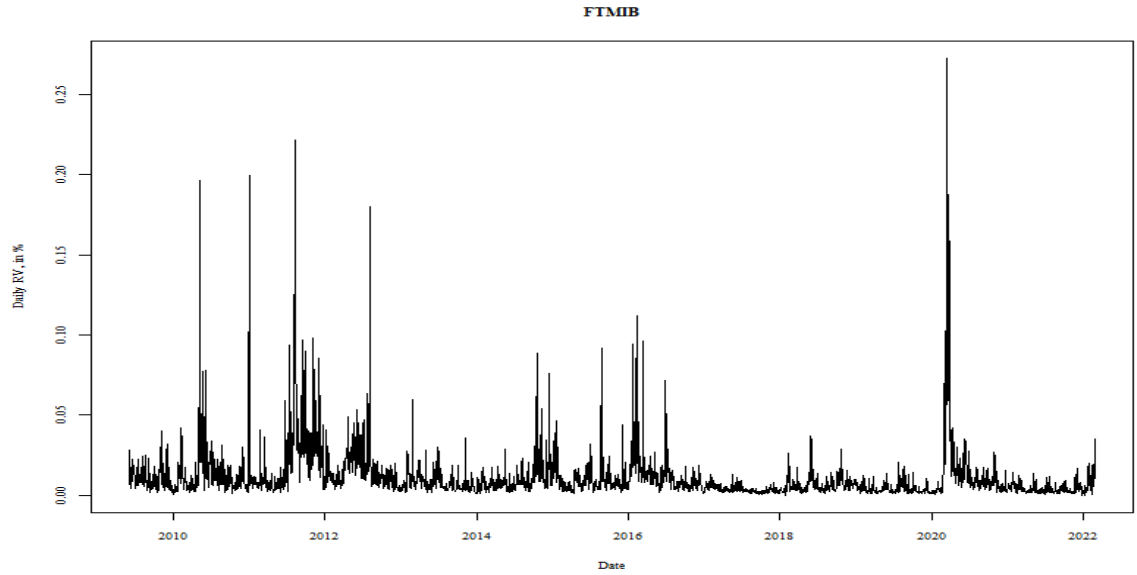

Figure 14: Realized Volatility - FTMIB (Italy)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

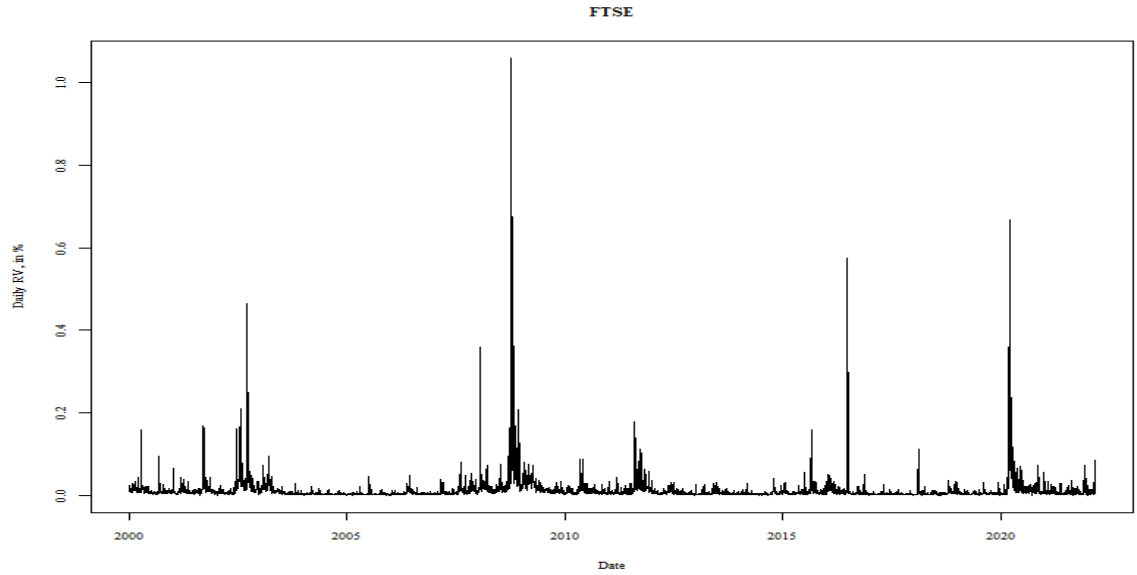

Figure 15: Realized Volatility - FTSE (UK)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

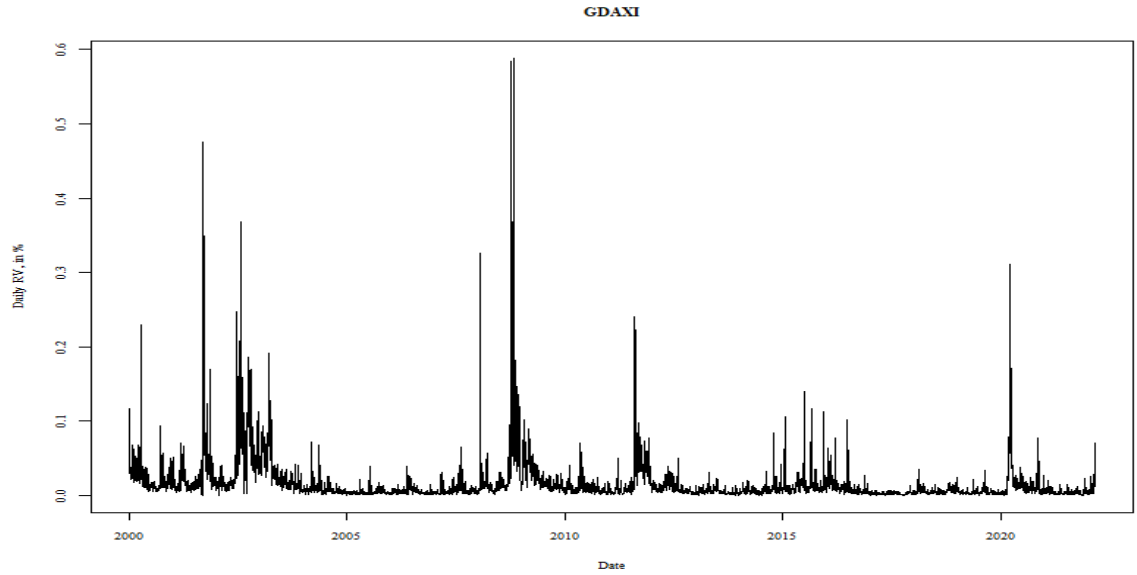

Figure 16: Realized Volatility - GDAXI (Germany)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

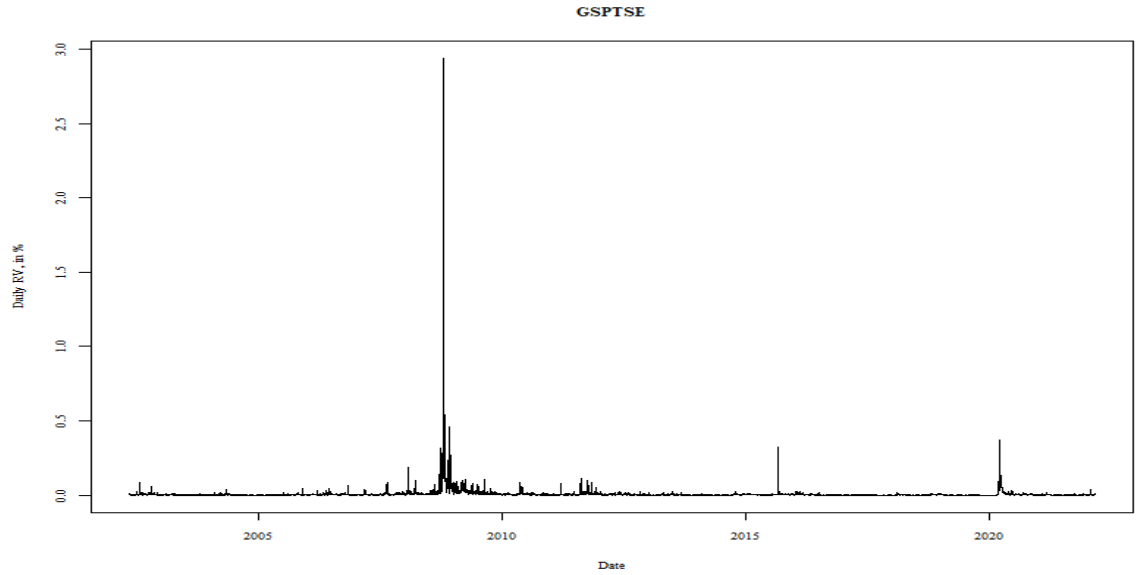

Figure 17: Realized Volatility - GSPTSE (Canada)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

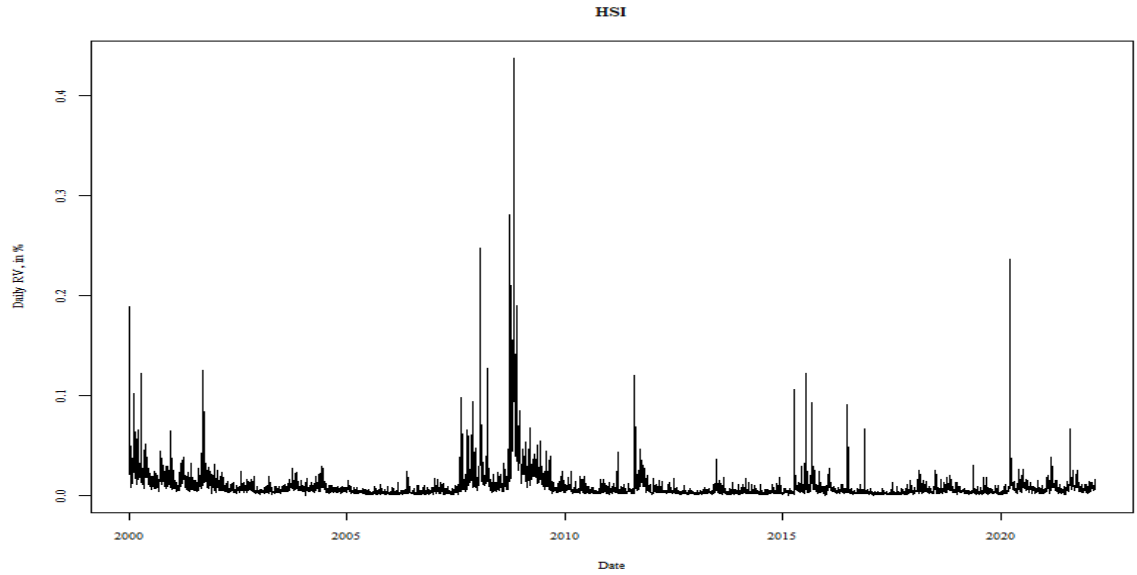

Figure 18: Realized Volatility - HSI (Hong Kong)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

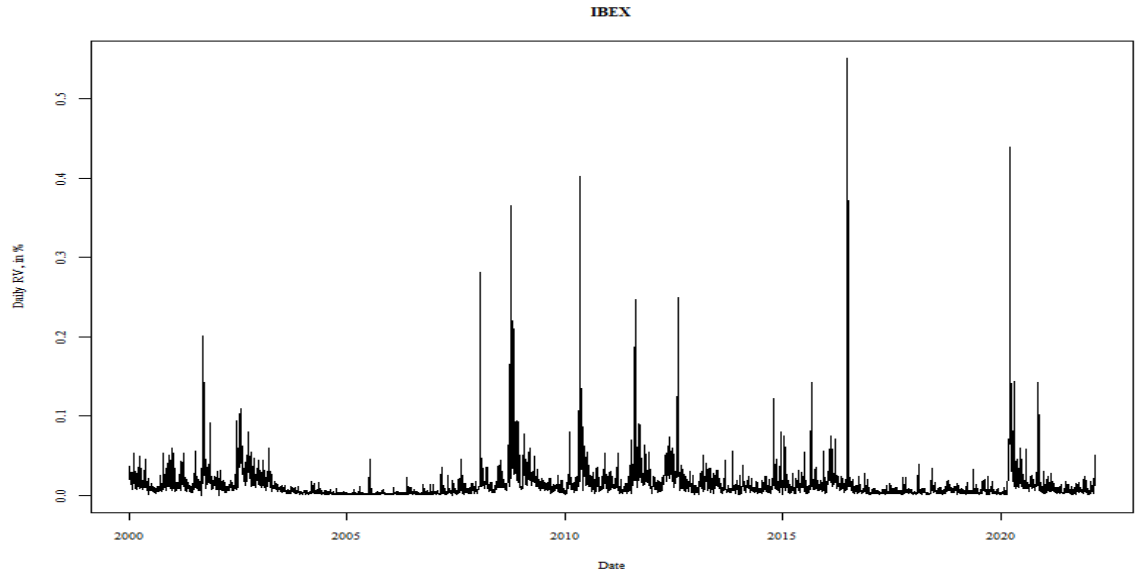

Figure 19: Realized Volatility - IBEX (Spain)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

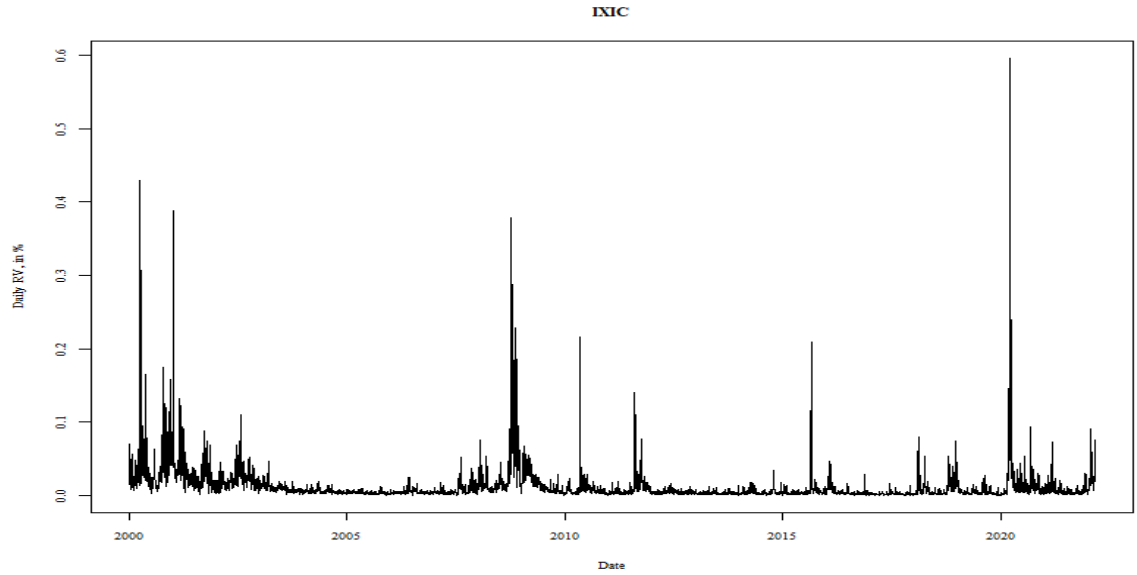

Figure 20: Realized Volatility - IXIC (USA - NASDAQ)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

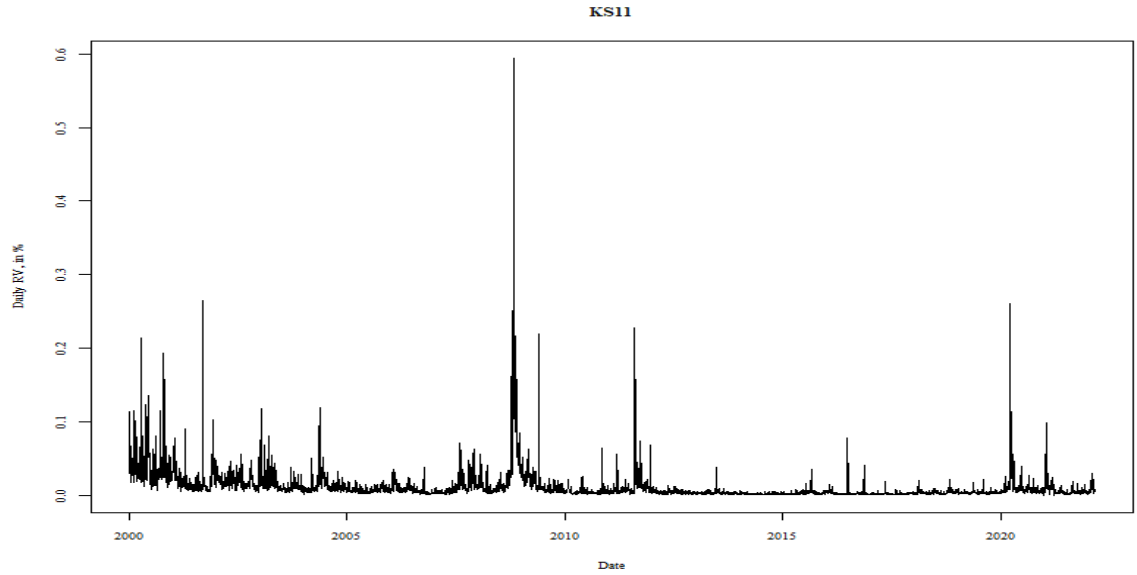

Figure 21: Realized Volatility - KS11 (South Korea)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

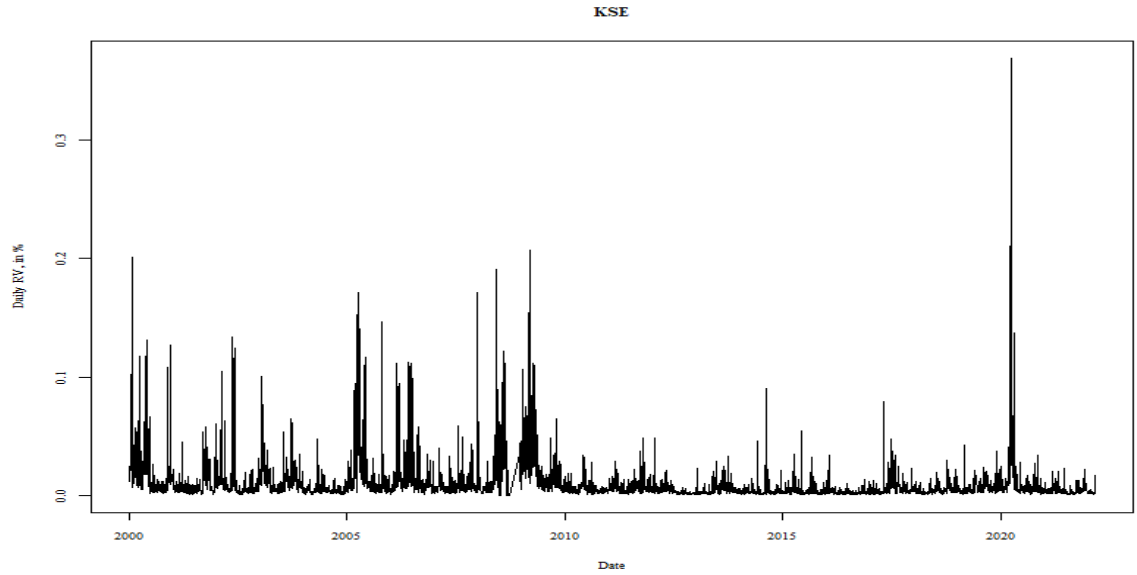

Figure 22: Realized Volatility - KSE (Pakistan)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

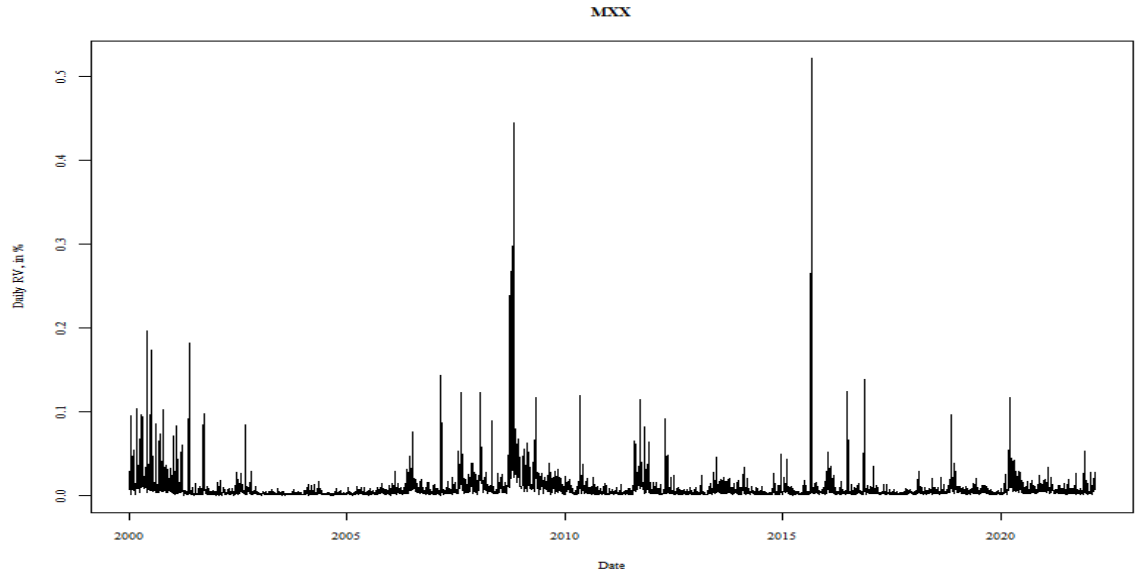

Figure 23: Realized Volatility - MXX (Mexico)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

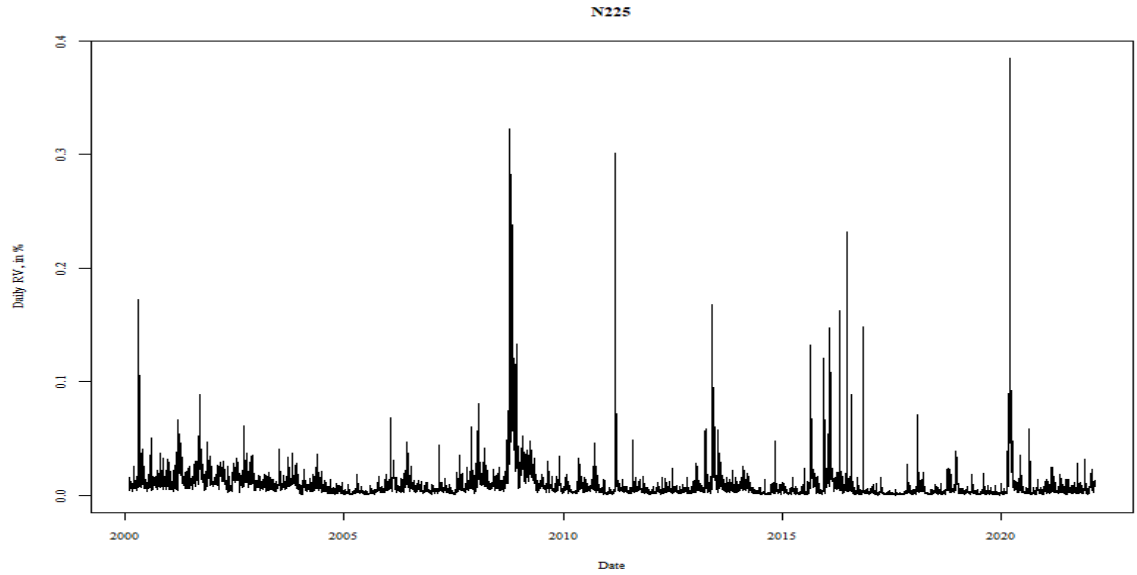

Figure 24: Realized Volatility - N225 (Japan)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

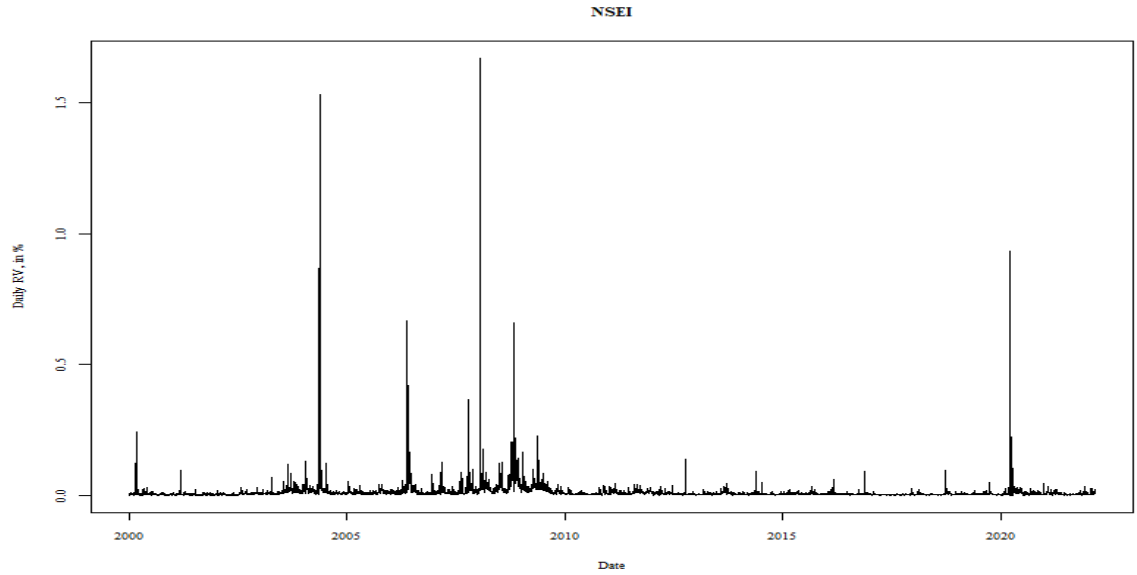

Figure 25: Realized Volatility - NSEI (Nifty 50 - India)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

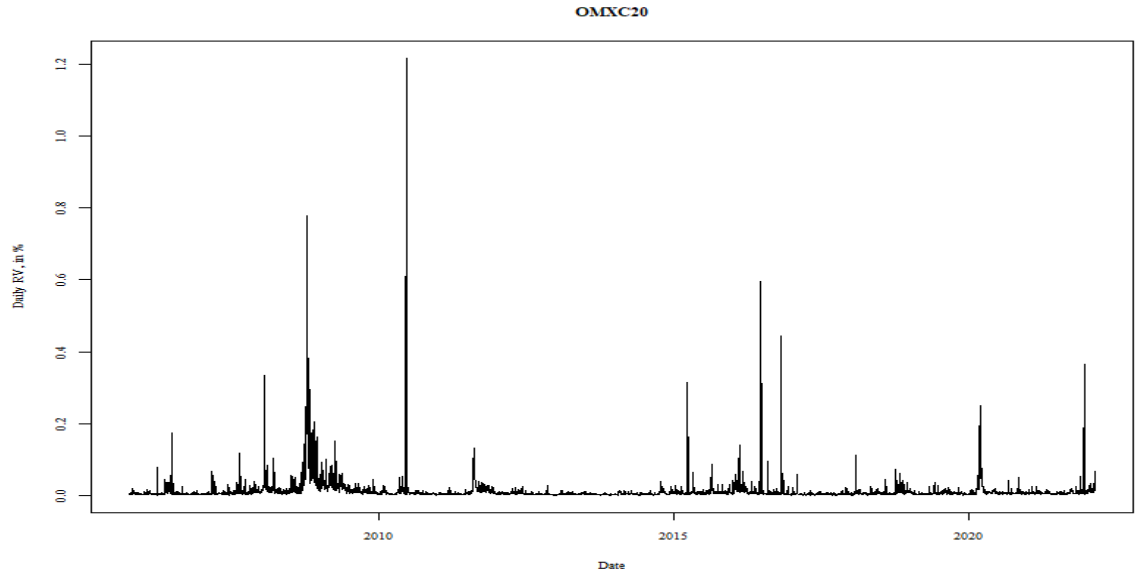

Figure 26: Realized Volatility - OMXC20 (Denmark)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

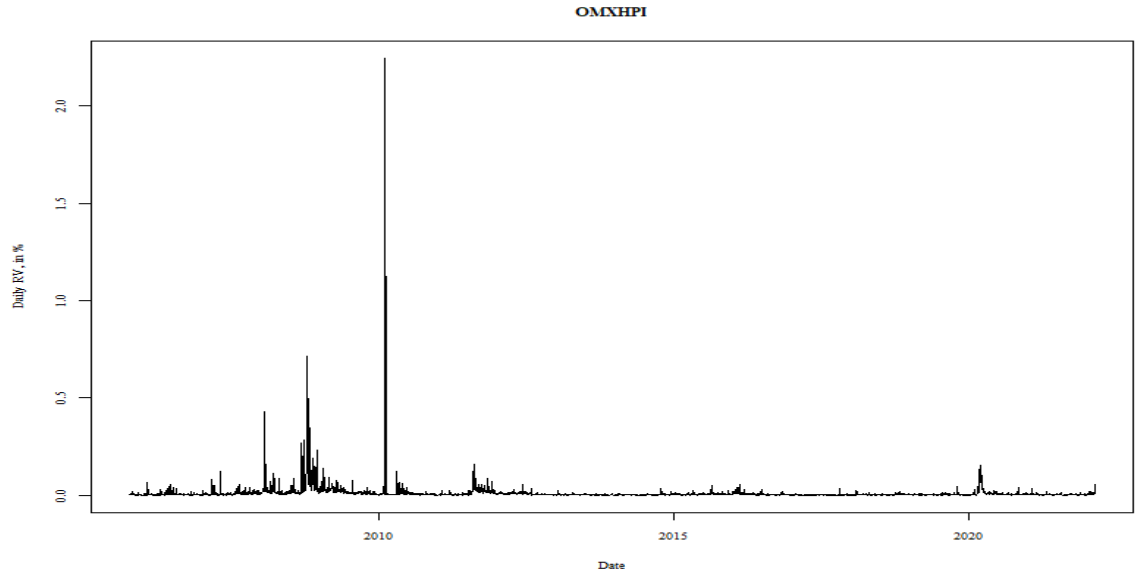

Figure 27: Realized Volatility - OMXHPI (Finland)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

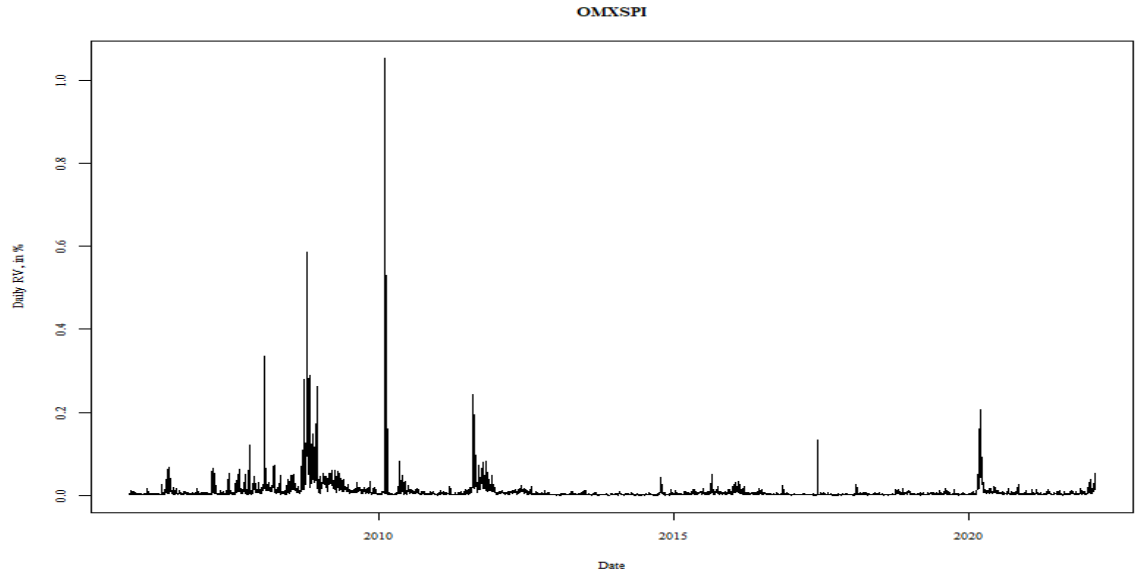

Figure 28: Realized Volatility - OMXSPI (Sweden)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

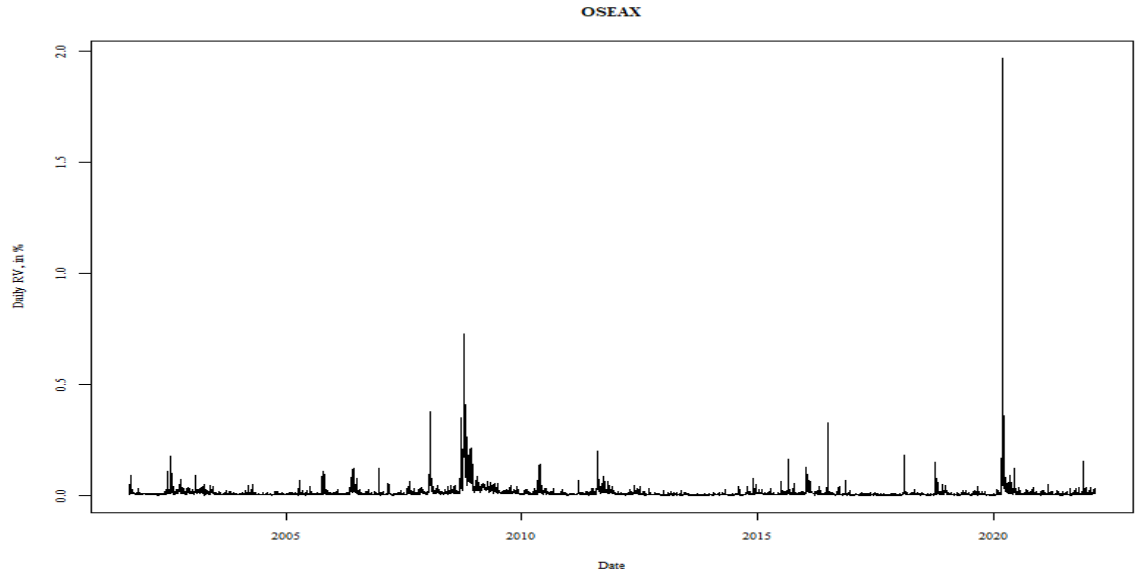

Figure 29: Realized Volatility - OSEAX (Norway)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

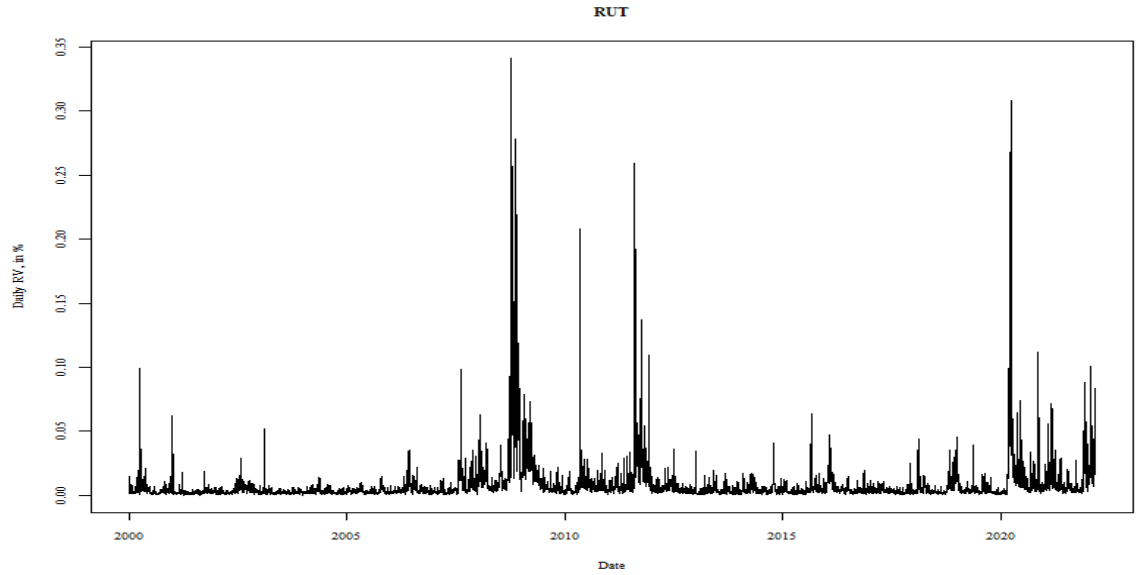

Figure 30: Realized Volatility - RUT (Russel 2000 - USA)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

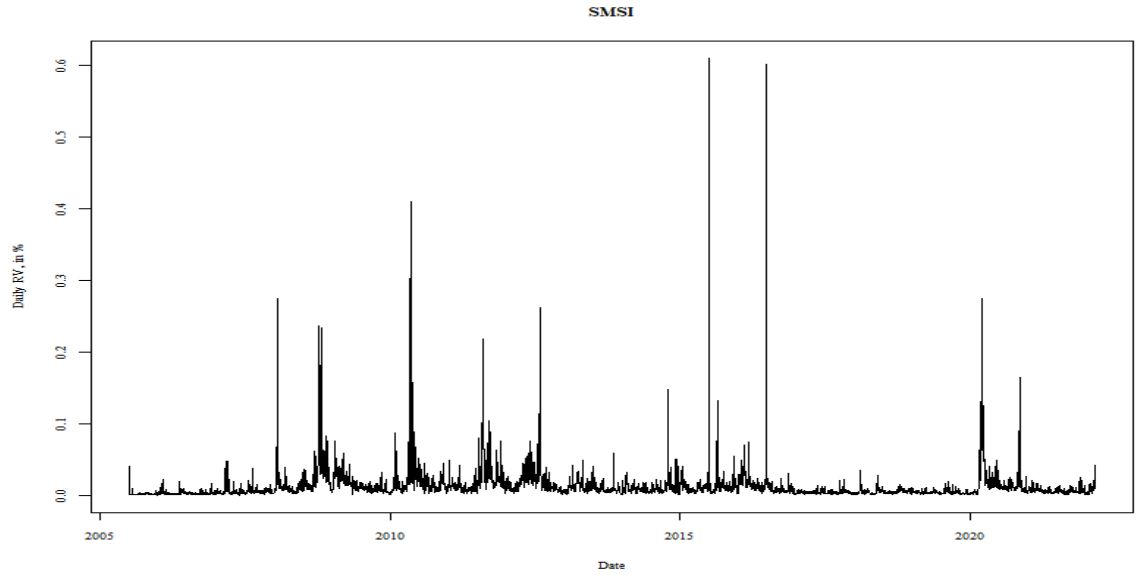

Figure 31: Realized Volatility - SMSI (General Index - Spain)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

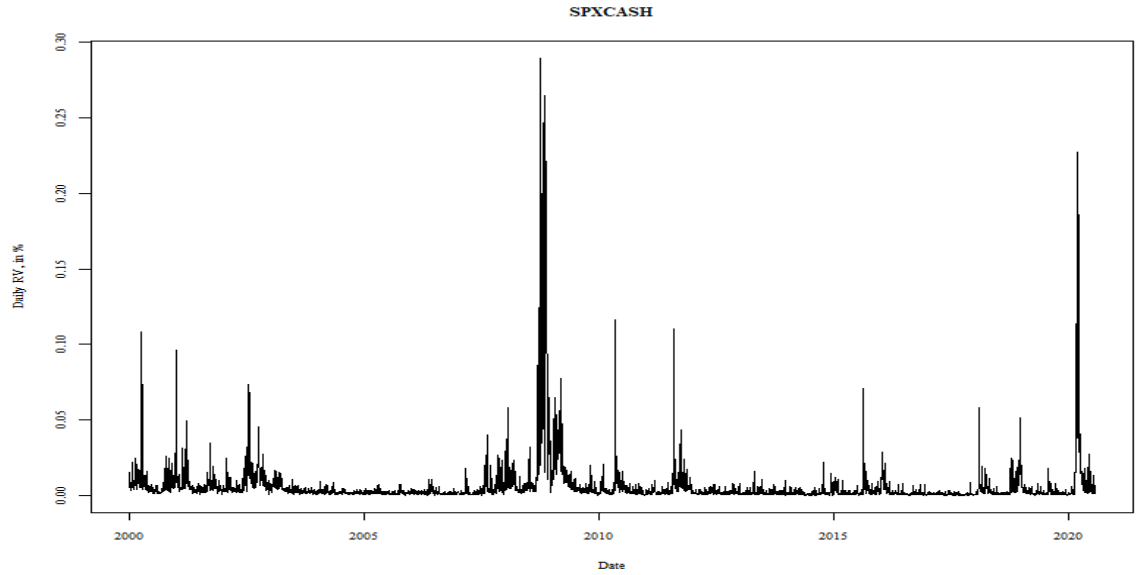

Figure 32: Realized Volatility - SPXCASH (S&P-500 Cash, USA)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

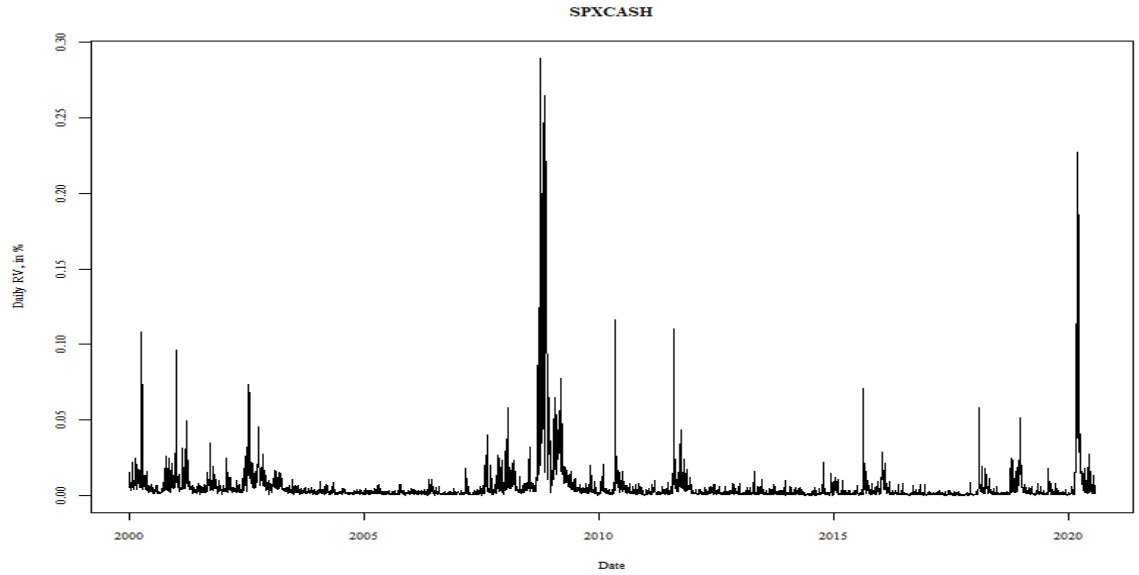

Figure 33: Realized Volatility - SPXFUT (S&P-500 Futures, USA)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

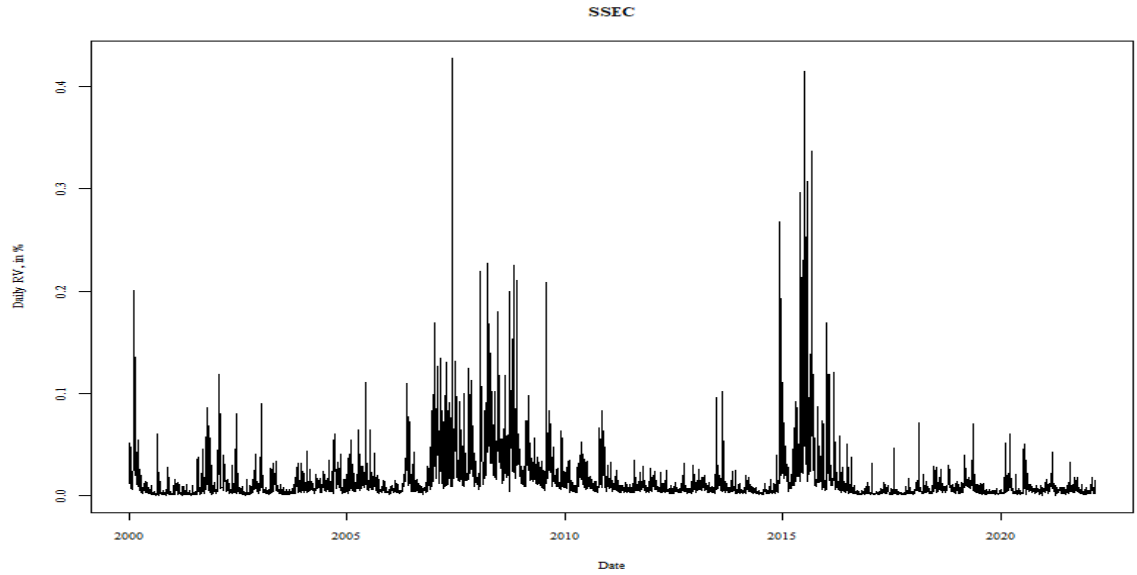

Figure 34: Realized Volatility - SSEC (China)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

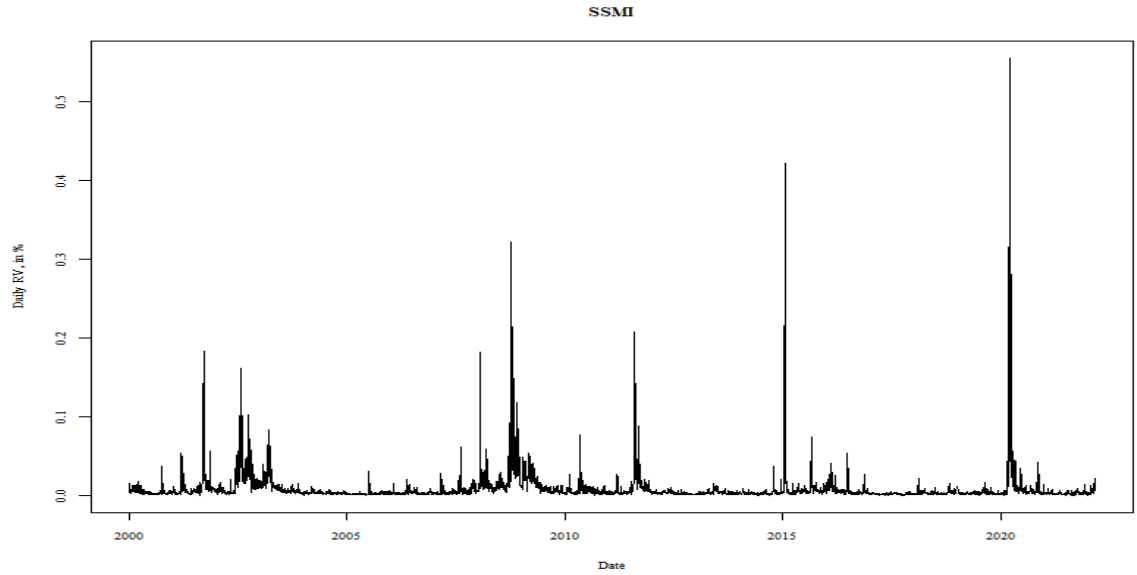

Figure 35: Realized Volatility - SSMI (Switzerland)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

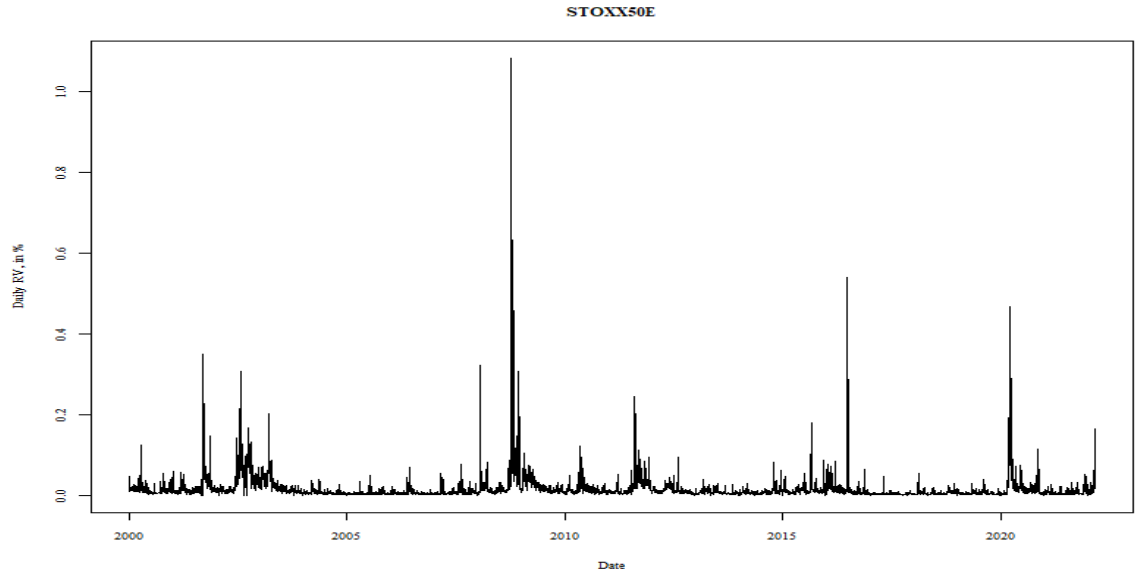

Figure 36: Realized Volatility - STOXX50E (Europe)
The figure illustrates the historical evolution of realized volatility. Further information on the index can be found in Appendix A.1 and Table 1

A.3 QLIKE and MAPE results

Table 15: QLIKE - Weekly Realized Variance Forecast
The table displays the QLIKE across the nine models for all 31 stock market indices forecasting weekly realized variance. For each row, the value with the lowest error is printed in bold and the value with the second lowest error is printed in italic. The last two rows count the total number of times that each model was the best or second best model in one of the 31 markets.

	HAR	C-HAR	HAR-J	WLS-HAR	RR-HAR	SVM	RF	AR-1	LSTM
AEX	1.00	1.004	0.991	**0.851**	*0.882*	3.508	0.967	1.696	3.891
AORD	1.00	*0.982*	0.988	**0.967**	1.036	6.893	1.014	1.42	3.117
BFX	1.00	0.97	0.978	**0.872**	*0.916*	4.508	0.991	1.666	3.48
BSESN	1.00	0.854	0.892	**0.611**	*0.644*	1.596	0.651	1.685	1.894
BVLG	1.00	1.016	0.992	**0.878**	*0.924*	8.274	1.065	1.715	4.954
BVSP	1.00	1.03	**0.982**	0.996	*0.99*	3.735	1.19	1.35	3.636
DJI	1.00	1.073	0.992	**0.868**	*0.926*	3.413	0.998	1.59	4.38
FCHI	1.00	1.04	0.994	**0.83**	*0.858*	3.346	0.958	1.713	3.438
FTMIB	1.00	1.027	0.983	*0.95*	**0.899**	3.197	1.075	1.642	3.369
FTSE	1.00	1.07	0.989	**0.875**	*0.919*	3.329	0.92	1.761	3.085
GDAXI	1.00	1.045	0.992	**0.793**	*0.804*	2.524	0.905	1.962	2.788
GSPTSE	1.00	1.212	0.69	**0.6**	0.686	4.396	*0.678*	1.907	3.03
HSI	1.00	1.02	0.992	**0.975**	*0.991*	4.038	1.075	1.7	2.709
IBEX	1.00	1.019	0.991	**0.896**	*0.944*	3.91	0.967	1.618	3.27
IXIC	1.00	1.172	0.996	**0.806**	*0.84*	2.642	0.914	1.989	3.467
KS11	1.00	1.151	0.996	**0.769**	*0.77*	2.712	0.89	1.868	2.555
KSE	1.00	1.095	*0.992*	1.977	**0.923**	2.521	0.998	1.506	2.677
MXX	1.00	1.017	**0.957**	*0.994*	1.074	3.576	1.036	1.451	2.96
N225	1.00	1.112	1.012	**0.946**	*0.948*	2.67	1.049	1.336	3.051
NSEI	1.00	0.869	0.875	**0.666**	0.727	2.475	*0.693*	1.597	2.364
OMXC20	1.00	1.028	1.037	1.015	**0.931**	1.476	*0.945*	1.465	2.107
OMXHPI	1.00	0.932	1.002	0.741	**0.553**	1.054	*0.652*	1.837	1.551
OMXSPI	1.00	0.823	0.828	*0.799*	**0.693**	1.476	0.823	1.969	2.254
OSEAX	1.00	1.016	**0.941**	*0.992*	1.031	4.336	1.056	1.449	4.469
RUT	**1.00**	1.05	**1**	1.078	1.056	7.518	1.175	1.27	4.887
SMSI	1.00	1.248	1.004	**0.86**	*0.901*	3.037	0.944	1.462	2.574
SPXCASH	1.00	1.097	1.015	**0.895**	*0.911*	4.766	1.051	1.721	3.986
SPXFUT	1.00	1.088	1.016	**0.874**	*0.893*	3.91	0.987	1.807	3.531
SSEC	1.00	1.017	1.00	**0.914**	*0.984*	4.914	0.986	1.722	2.962
SSMI	1.00	1.06	0.999	**0.867**	*0.882*	4.713	1.013	1.619	3.769
STOXX50E	1.00	0.945	0.963	*0.935*	**0.886**	2.723	0.983	1.637	3.662
Best Model	1	0	3	21	6	0	0	0	0
Second best	0	1	2	5	19	0	4	0	0

Table 16: MAPE - Weekly Realized Variance Forecast
The table displays the MAPE across the nine models for all 31 stock market indices forecasting weekly realized variance. For each row, the value with the lowest error is printed in bold and the value with the second lowest error is printed in italic. The last two rows count the total number of times that each model was the best or second best model in one of the 31 markets.

	HAR	C-HAR	HAR-J	WLS-HAR	RR-HAR	SVM	RF	AR-1	LSTM
AEX	1.00	0.999	0.993	*0.765*	**0.728**	1.646	0.886	1.682	1.257
AORD	1.00	1.022	1.005	*0.851*	**0.82**	1.423	1.014	1.362	1.242
BFX	1.00	0.959	0.975	*0.772*	**0.739**	1.77	0.922	1.635	1.291
BSESN	1.00	0.827	0.89	*0.527*	**0.483**	0.834	0.578	1.787	0.665
BVLG	1.00	1.001	0.995	*0.866*	**0.825**	2.439	0.953	1.575	1.152
BVSP	1.00	0.973	*0.971*	0.978	**0.873**	1.578	1.195	1.419	1.537
DJI	1.00	1.08	1.004	**0.743**	*0.754*	1.583	0.911	1.771	1.174
FCHI	1.00	1.03	0.992	*0.736*	**0.717**	1.716	0.85	1.686	1.167
FTMIB	1.00	1.041	0.99	*0.91*	**0.819**	1.937	1.046	1.613	1.307
FTSE	1.00	1.056	0.991	*0.727*	**0.696**	1.343	0.807	1.817	1.06
GDAXI	1.00	1.073	1.004	*0.72*	**0.697**	1.669	0.848	1.963	1.105
GSPTSE	1.00	0.554	0.54	*0.459*	**0.454**	0.691	0.58	2.029	0.67
HSI	1.00	1.07	1.009	*0.879*	**0.847**	1.859	1.001	1.789	1.276
IBEX	1.00	1.012	0.991	*0.782*	**0.766**	1.965	0.895	1.609	1.222
IXIC	1.00	1.195	0.995	**0.696**	*0.715*	1.669	0.849	2.08	1.110
KS11	1.00	1.158	0.996	*0.695*	**0.681**	1.785	0.782	1.836	0.989
KSE	1.00	1.058	*0.986*	1.888	**0.705**	1.345	1.152	1.491	1.533
MXX	1.00	0.885	0.92	*0.839*	**0.736**	1.061	1.003	1.544	1.258
N225	1.00	1.204	1.021	*0.849*	**0.793**	1.937	1.035	1.503	1.135
NSEI	1.00	0.813	0.881	**0.477**	*0.49*	0.744	0.615	1.647	0.736
OMXC20	1.00	1.002	1.03	*0.829*	**0.592**	0.859	0.902	1.74	0.881
OMXHPI	1.00	0.981	1.011	0.625	**0.4**	0.675	0.551	1.987	*0.514*
OMXSPI	1.00	0.782	0.823	*0.672*	**0.508**	1.023	0.728	2.202	0.732
OSEAX	1.00	0.96	0.943	*0.883*	**0.784**	1.255	1.035	1.505	1.29
RUT	1.00	1.024	1.003	*0.869*	**0.855**	1.274	1.089	1.204	1.576
SMSI	1.00	1.253	1.023	*0.734*	**0.701**	1.201	0.938	1.466	1.100
SPXCASH	1.00	1.141	1.024	*0.787*	**0.766**	1.936	1.004	1.901	1.293
SPXFUT	1.00	1.13	1.03	*0.777*	**0.755**	1.876	0.908	1.908	1.259
SSEC	1.00	1.006	0.998	*0.816*	**0.747**	1.377	0.931	1.632	1.407
SSMI	1.00	0.996	0.999	*0.754*	**0.682**	1.487	0.894	1.637	1.255
STOXX50E	1.00	0.924	0.962	*0.838*	**0.717**	1.41	0.901	1.683	1.162
Best Model	0	0	0	3	28	0	0	0	0
Second best	0	0	2	25	3	0	0	0	1

Table 17: QLIKE - Monthly Realized Variance Forecast
The table displays the QLIKE across the nine models for all 31 stock market indices forecasting monthly realized variance. For each row, the value with the lowest error is printed in bold and the value with the second lowest error is printed in italic. The last two rows count the total number of times that each model was the best or second best model in one of the 31 markets.

	HAR	C-HAR	HAR-J	WLS-HAR	RR-HAR	SVM	RF	AR-1	LSTM
AEX	1.00	0.983	0.992	0.926	*0.899*	1.688	**0.872**	1.27	3.211
AORD	1.00	**0.985**	0.988	1.007	1.121	3.582	*0.986*	1.276	2.884
BFX	1.00	0.98	0.987	**0.892**	0.948	2.197	*0.896*	1.316	1.789
BSESN	1.00	0.928	0.96	*0.767*	0.773	1.131	**0.697**	1.357	1.404
BVLG	1.00	0.996	0.997	*0.931*	1.029	2.423	**0.891**	1.15	2.168
BVSP	1.00	**0.982**	*0.984*	1.013	0.997	2.032	1.12	1.174	2.134
DJI	1.00	1.009	*0.99*	**0.956**	1.005	1.917	0.993	1.235	3.018
FCHI	1.00	1.012	0.998	*0.873*	0.885	1.694	**0.833**	1.327	2.153
FTMIB	1.00	0.996	0.992	*0.949*	**0.897**	1.432	1.096	1.24	1.744
FTSE	1.00	1.027	0.995	*0.874*	0.968	2.069	**0.849**	1.451	2.101
GDAXI	1.00	1.018	0.997	**0.82**	*0.83*	1.389	0.838	1.442	1.828
GSPTSE	1.00	*0.837*	0.864	**0.782**	0.856	2.795	0.921	1.568	2.034
HSI	1.00	1.025	0.998	**0.913**	*0.93*	2.625	0.959	1.719	1.785
IBEX	1.00	0.995	0.993	*0.918*	0.984	2.07	**0.914**	1.303	1.524
IXIC	1.00	1.059	0.997	**0.858**	0.906	1.487	*0.882*	1.469	1.889
KS11	1.00	1.042	0.999	*0.655*	0.695	1.36	**0.632**	1.428	1.185
KSE	1.00	1.029	0.999	10.415	*0.873*	1.546	**0.87**	1.35	1.6
MXX	1.00	**0.974**	*0.977*	1.017	1.056	2.538	1.01	1.357	1.921
N225	1.00	1.036	1.007	**0.933**	*0.94*	1.553	0.978	1.175	2.279
NSEI	1.00	0.932	0.954	*0.773*	0.866	1.659	**0.7**	1.348	1.293
OMXC20	1.00	0.987	1.005	0.881	*0.851*	1.088	**0.821**	1.336	1.533
OMXHPI	1.00	1.015	1.016	0.767	**0.64**	0.804	*0.739*	1.492	1.031
OMXSPI	1.00	0.899	0.937	*0.796*	**0.72**	0.972	0.818	1.484	1.196
OSEAX	1.00	0.976	*0.971*	0.984	1.063	2.314	**0.957**	1.192	2.36
RUT	1.00	**0.995**	*0.997*	2.37	1.182	3.908	1.036	1.159	2.368
SMSI	1.00	1.081	1.006	**0.944**	1.01	1.725	*0.947*	1.191	1.666
SPXCASH	*1.00*	1.024	1.004	**0.964**	1.028	2.395	1.013	1.269	2.957
SPXFUT	1.00	1.028	1.003	**0.914**	0.986	2.001	*0.975*	1.294	2.547
SSEC	1.00	0.996	1.001	**0.898**	0.968	2.849	*0.921*	1.654	2.079
SSMI	1.00	1.041	0.993	**0.892**	0.958	2.055	*0.92*	1.217	1.853
STOXX50E	1.00	0.968	0.988	0.975	*0.931*	1.568	**0.905**	1.29	2.104
Best Model	0	4	0	12	3	0	12	0	0
Second best	1	1	5	9	7	0	8	0	0

Table 18: MAPE - Monthly Realized Variance Forecast
The table displays the MAPE across the nine models for all 31 stock market indices forecasting monthly realized variance. For each row, the value with the lowest error is printed in bold and the value with the second lowest error is printed in italic. The last two rows count the total number of times that each model was the best or second best model in one of the 31 markets.

	HAR	C-HAR	HAR-J	WLS-HAR	RR-HAR	SVM	RF	AR-1	LSTM
AEX	1.00	0.985	0.994	0.839	**0.585**	0.938	0.722	1.418	*0.677*
AORD	1.00	1.012	1.001	*0.853*	**0.808**	1.133	0.954	1.322	1.00
BFX	1.00	0.976	0.988	*0.74*	**0.627**	1.106	0.837	1.455	0.754
BSESN	1.00	0.907	0.955	0.648	*0.528*	0.72	0.572	1.541	**0.49**
BVLG	1.00	0.986	0.998	0.816	0.82	1.593	*0.731*	1.353	**0.718**
BVSP	1.00	*0.954*	0.972	1.006	**0.865**	1.117	1.164	1.24	1.075
DJI	1.00	1.038	0.998	0.731	**0.643**	0.982	0.776	1.497	*0.684*
FCHI	1.00	1.008	0.997	0.764	**0.614**	1.04	0.7	1.456	*0.685*
FTMIB	1.00	1.018	1.00	0.9	**0.728**	1.239	1.018	1.446	*0.744*
FTSE	1.00	1.005	0.992	0.702	**0.62**	0.898	0.729	1.562	*0.687*
GDAXI	1.00	1.046	1.003	0.653	**0.608**	1.015	0.696	1.605	*0.615*
GSPTSE	1.00	0.776	0.812	*0.461*	**0.446**	0.561	0.536	1.696	0.479
HSI	1.00	1.052	1.004	*0.787*	**0.747**	1.298	0.812	1.681	0.894
IBEX	1.00	0.993	0.995	*0.74*	**0.712**	1.338	0.789	1.441	0.8
IXIC	1.00	1.1	0.994	*0.65*	**0.569**	1.019	0.716	1.739	0.691
KS11	1.00	1.049	0.999	0.578	0.608	1.05	*0.546*	1.495	**0.473**
KSE	1.00	1.024	0.996	2.697	**0.738**	0.934	0.861	1.354	*0.739*
MXX	1.00	0.879	0.954	*0.83*	**0.746**	0.849	0.942	1.45	0.993
N225	1.00	1.068	1.009	*0.799*	0.805	1.227	0.857	1.254	**0.74**
NSEI	1.00	0.907	0.958	*0.544*	**0.522**	0.659	0.615	1.436	0.55
OMXC20	1.00	0.999	1.009	0.669	**0.547**	*0.607*	0.721	1.444	0.612
OMXHPI	1.00	1.039	1.015	0.621	*0.43*	0.551	0.609	1.643	**0.329**
OMXSPI	1.00	0.879	0.932	0.654	*0.486*	0.709	0.689	1.676	**0.417**
OSEAX	1.00	0.941	0.966	*0.825*	**0.704**	0.891	0.873	1.368	0.884
RUT	1.00	0.992	0.999	*0.897*	**0.826**	0.97	1.09	1.206	1.17
SMSI	1.00	1.126	1.017	*0.75*	**0.68**	0.913	0.9	1.319	0.781
SPXCASH	1.00	1.08	0.985	*0.749*	**0.637**	1.084	0.87	1.564	0.778
SPXFUT	1.00	1.08	0.992	*0.697*	**0.616**	1.065	0.764	1.547	0.727
SSEC	1.00	0.989	1.00	*0.794*	**0.729**	1.089	0.892	1.582	0.988
SSMI	1.00	0.983	0.999	0.701	**0.549**	0.824	0.792	1.385	*0.648*
STOXX50E	1.00	0.948	0.984	0.867	**0.627**	0.907	0.759	1.438	*0.708*
Best Model	0	0	0	0	25	0	0	0	6
Second best	0	1	0	15	3	1	2	0	9

A.4 Individual Diebold-Mariano tests

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.83	-1.52	1.59*	-1.55	0.51	2.54**	1.66**	1.78**
C-HAR	0.83	-	0.22	1.81**	-1.2	0.89	2.57**	1.81**	1.81**
WLS	1.52*	-0.22	-	1.6*	-0.92	0.84	2.53**	1.68**	1.8**
RR	-1.59	-1.81	-1.6	-	-1.81	-0.69	2.6**	1.52*	1.64*
HAR-J	1.55*	1.2	0.92	1.81**	-	1.33*	2.55**	1.79**	1.83**
AR-1	-0.51	-0.89	-0.84	0.69	-1.33	-	2.44**	1.26	1.76**
SVM	-2.54	-2.57	-2.53	-2.6	-2.55	-2.44	-	-2.73	-0.28
RF	-1.66	-1.81	-1.68	-1.52	-1.79	-1.26	2.73**	-	1.41*
LSTM	-1.78	-1.81	-1.8	-1.64	-1.83	-1.76	0.28	-1.41	-

Table 19: AEX Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.09	-1.58	1.93**	-0.27	0.91	3.36**	1.97**	2.73**
C-HAR	-1.09	-	-1.39	0.8	-1.36	-0.08	3.37**	1.77**	2.67**
WLS	1.58*	1.39*	-	1.81**	1.31*	1.49*	3.35**	2.04**	2.76**
RR	-1.93	-0.8	-1.81	-	-1.98	-0.53	3.37**	1.72**	2.65**
HAR-J	0.27	1.36*	-1.31	1.98**	-	1.09	3.36**	1.97**	2.73**
AR-1	-0.91	0.08	-1.49	0.53	-1.09	-	3.31**	1.62*	2.68**
SVM	-3.36	-3.37	-3.35	-3.37	-3.36	-3.31	-	-3.6	-0.17
RF	-1.97	-1.77	-2.04	-1.72	-1.97	-1.62	3.6**	-	2.35**
LSTM	-2.73	-2.67	-2.76	-2.65	-2.73	-2.68	0.17	-2.35	-

Table 20: AEX Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-2.32	1.66**	1.23	-1.53	1.3*	5.67**	-3.58	3.35**
C-HAR	2.32**	-	1.98**	4.3**	3.2**	1.92**	6.2**	-3.41	3.62**
WLS	-1.66	-1.98	-	-0.93	-1.72	-0.23	4.15**	-2.88	2.94**
RR	-1.23	-4.3	0.93	-	-3.31	0.77	5.9**	-4.35	3.23**
HAR-J	1.53*	-3.2	1.72**	3.31**	-	1.59*	5.96**	-3.83	3.48**
AR-1	-1.3	-1.92	0.23	-0.77	-1.59	-	4.77**	-2.8	3.03**
SVM	-5.67	-6.2	-4.15	-5.9	-5.96	-4.77	-	-6.36	0.38
RF	3.58**	3.41**	2.88**	4.35**	3.83**	2.8**	6.36**	-	4.06**
LSTM	-3.35	-3.62	-2.94	-3.23	-3.48	-3.03	-0.38	-4.06	-

Table 21: AEX Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.49	0.39	1.06	-1.48	0.94	1.54*	1.14	1.05
C-HAR	1.49*	-	0.7	1.17	-1.32	1.18	1.57*	1.23	1.09
WLS	-0.39	-0.7	-	1.48*	-1.05	0.8	1.67**	1.29*	0.99
RR	-1.06	-1.17	-1.48	-	-1.32	-0.54	1.72**	-0.19	0.77
HAR-J	1.48*	1.32*	1.05	1.32*	-	1.41*	1.61*	1.35*	1.12
AR-1	-0.94	-1.18	-0.8	0.54	-1.41	-	1.54*	0.3	0.89
SVM	-1.54	-1.57	-1.67	-1.72	-1.61	-1.54	-	-1.54	-0.05
RF	-1.14	-1.23	-1.29	0.19	-1.35	-0.3	1.54*	-	0.81
LSTM	-1.05	-1.09	-0.99	-0.77	-1.12	-0.89	0.05	-0.81	-

Table 22: AORD Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-2.07	-1.37	1.01	-2.13	1.16	3.2**	2.1**	2.44**
C-HAR	2.07**	-	0.29	1.61*	-0.95	1.48*	3.21**	2.16**	2.5**
WLS	1.37*	-0.29	-	2.2**	-0.49	1.31*	3.23**	2.24**	2.49**
RR	-1.01	-1.61	-2.2	-	-1.67	0.69	3.24**	2.11**	2.37**
HAR-J	2.13**	0.95	0.49	1.67**	-	1.58*	3.21**	2.16**	2.51**
AR-1	-1.16	-1.48	-1.31	-0.69	-1.58	-	3.1**	1.41*	2.24**
SVM	-3.2	-3.21	-3.23	-3.24	-3.21	-3.1	-	-3.35	-0.58
RF	-2.1	-2.16	-2.24	-2.11	-2.16	-1.41	3.35**	-	1.58*
LSTM	-2.44	-2.5	-2.49	-2.37	-2.51	-2.24	0.58	-1.58	-

Table 23: AORD Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.07	-2.92	-1.08	-1.42	-0.15	2.89**	-0.08	1.27
C-HAR	1.07	-	-3.26	-1.03	-0.2	0	3.04**	0.12	1.32*
WLS	2.92**	3.26**	-	0.39	2.8**	0.53	3.28**	0.96	1.5*
RR	1.08	1.03	-0.39	-	0.89	0.33	3.6**	1.01	1.43*
HAR-J	1.42*	0.2	-2.8	-0.89	-	0.02	2.97**	0.13	1.32*
AR-1	0.15	0	-0.53	-0.33	-0.02	-	2.31**	0.06	1.25
SVM	-2.89	-3.04	-3.28	-3.6	-2.97	-2.31	-	-4.27	-0.02
RF	0.08	-0.12	-0.96	-1.01	-0.13	-0.06	4.27**	-	1.28
LSTM	-1.27	-1.32	-1.5	-1.43	-1.32	-1.25	0.02	-1.28	-

Table 24: AORD Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.95	-1.26	1.8**	-1.81	1.1	2.13**	1.72**	1.4*
C-HAR	1.95**	-	0.63	2.1**	1.11	1.61*	2.14**	1.89**	1.5*
WLS	1.26	-0.63	-	1.72**	-0.43	1.18	2.12**	1.7**	1.48*
RR	-1.8	-2.1	-1.72	-	-2.06	0.61	2.13**	1.57*	1.3*
HAR-J	1.81**	-1.11	0.43	2.06**	-	1.63*	2.14**	1.88**	1.49*
AR-1	-1.1	-1.61	-1.18	-0.61	-1.63	-	1.96**	0.33	1.15
SVM	-2.13	-2.14	-2.12	-2.13	-2.14	-1.96	-	-2.2	-0.23
RF	-1.72	-1.89	-1.7	-1.57	-1.88	-0.33	2.2**	-	1.11
LSTM	-1.4	-1.5	-1.48	-1.3	-1.49	-1.15	0.23	-1.11	-

Table 25: BFX Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-2.6	-1.71	2.05**	-2.41	1.69**	3.53**	2.74**	2.91**
C-HAR	2.6**	-	0.85	2.41**	2.2**	2.52**	3.53**	2.78**	2.97**
WLS	1.71**	-0.85	-	1.95**	0.03	2.03**	3.52**	2.73**	2.96**
RR	-2.05	-2.41	-1.95	-	-2.28	0.56	3.55**	2.72**	2.85**
HAR-J	2.41**	-2.2	-0.03	2.28**	-	2.4**	3.53**	2.75**	2.95**
AR-1	-1.69	-2.52	-2.03	-0.56	-2.4	-	3.46**	2.28**	2.82**
SVM	-3.53	-3.53	-3.52	-3.55	-3.53	-3.46	-	-3.65	-0.38
RF	-2.74	-2.78	-2.73	-2.72	-2.75	-2.28	3.65**	-	2.27**
LSTM	-2.91	-2.97	-2.96	-2.85	-2.95	-2.82	0.38	-2.27	-

Table 26: BFX Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-3.09	1.19	2.36**	-2.37	1.5*	5.58**	-2.01	2.42**
C-HAR	3.09**	-	1.71**	3.97**	2.57**	1.9**	5.72**	0.06	2.65**
WLS	-1.19	-1.71	-	-0.63	-1.51	0.73	4.36**	-1.56	2.11**
RR	-2.36	-3.97	0.63	-	-3.72	1.09	5.62**	-3.05	2.27**
HAR-J	2.37**	-2.57	1.51*	3.72**	-	1.77**	5.71**	-0.58	2.59**
AR-1	-1.5	-1.9	-0.73	-1.09	-1.77	-	4.32**	-1.83	1.91**
SVM	-5.58	-5.72	-4.36	-5.62	-5.71	-4.32	-	-5.97	-0.69
RF	2.01**	-0.06	1.56*	3.05**	0.58	1.83**	5.97**	-	2.65**
LSTM	-2.42	-2.65	-2.11	-2.27	-2.59	-1.91	0.69	-2.65	-

Table 27: BFX Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.97	-1.18	1.28	-1.06	-0.67	1.6*	0.31	0.2
C-HAR	0.97	-	0.68	1.23	-1.04	0.52	1.58*	1	0.26
WLS	1.18	-0.68	-	1.34*	-0.93	0.19	1.56*	1.09	0.23
RR	-1.28	-1.23	-1.34	-	-1.23	-0.88	1.63*	-1.74	0.18
HAR-J	1.06	1.04	0.93	1.23	-	0.75	1.54*	1.07	0.28
AR-1	0.67	-0.52	-0.19	0.88	-0.75	-	1.45*	0.63	0.23
SVM	-1.6	-1.58	-1.56	-1.63	-1.54	-1.45	-	-1.66	0
RF	-0.31	-1	-1.09	1.74**	-1.07	-0.63	1.66**	-	0.2
LSTM	-0.2	-0.26	-0.23	-0.18	-0.28	-0.23	0	-0.2	-

Table 28: BSESN Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.39	0.88	0.81	-1.92	0.88	3.46**	0.66	1.27
C-HAR	1.39*	-	1.11	1.58*	-0.38	0.95	3.34**	1.11	1.38*
WLS	-0.88	-1.11	-	-0.83	-1.23	0.85	3.58**	-0.75	1.07
RR	-0.81	-1.58	0.83	-	-2	0.87	3.49**	0.46	1.24
HAR-J	1.92**	0.38	1.23	2**	-	0.97	3.35**	1.25	1.41*
AR-1	-0.88	-0.95	-0.85	-0.87	-0.97	-	0.73	-0.89	0.35
SVM	-3.46	-3.34	-3.58	-3.49	-3.35	-0.73	-	-4.03	-0.05
RF	-0.66	-1.11	0.75	-0.46	-1.25	0.89	4.03**	-	1.19
LSTM	-1.27	-1.38	-1.07	-1.24	-1.41	-0.35	0.05	-1.19	-

Table 29: BSESN Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.5	1.26	4.01**	-1.29	0.99	2.83**	-2.14	1.46*
C-HAR	1.5*	-	1.34*	2.09**	1.22	1.01	7.23**	-4.36	2.41**
WLS	-1.26	-1.34	-	-0.94	-1.29	0.95	0	-1.53	0.02
RR	-4.01	-2.09	0.94	-	-1.92	0.95	1.66**	-2.61	0.95
HAR-J	1.29*	-1.22	1.29*	1.92**	-	1	7.07**	-4.38	2.32**
AR-1	-0.99	-1.01	-0.95	-0.95	-1	-	-0.81	-1.04	-0.8
SVM	-2.83	-7.23	0	-1.66	-7.07	0.81	-	-8.17	0.05
RF	2.14**	4.36**	1.53*	2.61**	4.38**	1.04	8.17**	-	2.76**
LSTM	-1.46	-2.41	-0.02	-0.95	-2.32	0.8	-0.05	-2.76	-

Table 30: BSESN Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.58	-1.16	1.38*	-0.83	0.27	1.88**	1.56*	1.59*
C-HAR	0.58	-	0.2	1.22	0.41	0.66	1.86**	1.52*	1.61*
WLS	1.16	-0.2	-	1.35*	0.75	0.47	1.88**	1.55*	1.6*
RR	-1.38	-1.22	-1.35	-	-1.33	-0.7	1.92**	1.57*	1.45*
HAR-J	0.83	-0.41	-0.75	1.33*	-	0.44	1.87**	1.54*	1.6*
AR-1	-0.27	-0.66	-0.47	0.7	-0.44	-	1.8**	1.4*	1.57*
SVM	-1.88	-1.86	-1.88	-1.92	-1.87	-1.8	-	-2.17	-0.14
RF	-1.56	-1.52	-1.55	-1.57	-1.54	-1.4	2.17**	-	0.68
LSTM	-1.59	-1.61	-1.6	-1.45	-1.6	-1.57	0.14	-0.68	-

Table 31: BVLG Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.41	-0.69	1.77**	-1.41	0.62	2.6**	2**	2.23**
C-HAR	0.41	-	0.24	1.48*	0.2	0.64	2.59**	1.99**	2.24**
WLS	0.69	-0.24	-	1.75**	-0.25	0.67	2.6**	1.99**	2.23**
RR	-1.77	-1.48	-1.75	-	-1.75	-0.3	2.6**	1.88**	2.17**
HAR-J	1.41*	-0.2	0.25	1.75**	-	0.66	2.6**	2**	2.24**
AR-1	-0.62	-0.64	-0.67	0.3	-0.66	-	2.59**	1.77**	2.17**
SVM	-2.6	-2.59	-2.6	-2.6	-2.6	-2.59	-	-2.68	-0.13
RF	-2	-1.99	-1.99	-1.88	-2	-1.77	2.68**	-	1.68**
LSTM	-2.23	-2.24	-2.23	-2.17	-2.24	-2.17	0.13	-1.68	-

Table 32: BVLG Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.4*	-1.85	-1.89	0.68	-1.91	1.19	-1.95	0.78
C-HAR	-1.4	-	-2.14	-1.99	-1.57	-2.31	0.88	-2.05	0.61
WLS	1.85**	2.14**	-	-1.63	1.93**	-1.3	1.97**	-1.69	1.17
RR	1.89**	1.99**	1.63*	-	1.9**	0.4	3.54**	-1	1.65*
HAR-J	-0.68	1.57*	-1.93	-1.9	-	-2	1.15	-1.97	0.76
AR-1	1.91**	2.31**	1.3*	-0.4	2**	-	2.66**	-0.64	1.51*
SVM	-1.19	-0.88	-1.97	-3.54	-1.15	-2.66	-	-3.87	0.09
RF	1.95**	2.05**	1.69**	1	1.97**	0.64	3.87**	-	1.74**
LSTM	-0.78	-0.61	-1.17	-1.65	-0.76	-1.51	-0.09	-1.74	-

Table 33: BVLG Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.26	1.5*	2.09**	0.53	2**	2.58**	1.7**	2.1**
C-HAR	0.26	-	1.51*	2.09**	1.2	1.82**	2.59**	1.71**	2.11**
WLS	-1.5	-1.51	-	2.89**	-1.27	0.98	2.65**	1.7**	1.99**
RR	-2.09	-2.09	-2.89	-	-1.94	-0.08	2.63**	0.1	1.89**
HAR-J	-0.53	-1.2	1.27	1.94**	-	1.74**	2.57**	1.57*	2.09**
AR-1	-2	-1.82	-0.98	0.08	-1.74	-	2.52**	0.1	1.91**
SVM	-2.58	-2.59	-2.65	-2.63	-2.57	-2.52	-	-2.66	-0.3
RF	-1.7	-1.71	-1.7	-0.1	-1.57	-0.1	2.66**	-	1.88**
LSTM	-2.1	-2.11	-1.99	-1.89	-2.09	-1.91	0.3	-1.88	-

Table 34: BVSP Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.85	1.26	2.16**	-1.13	1.93**	4.09**	3.69**	3.62**
C-HAR	0.85	-	1.32*	2.04**	0.33	1.95**	4.1**	3.85**	3.65**
WLS	-1.26	-1.32	-	3.11**	-1.68	2**	4.13**	3.59**	3.59**
RR	-2.16	-2.04	-3.11	-	-2.37	1.41*	4.13**	3.06**	3.52**
HAR-J	1.13	-0.33	1.68**	2.37**	-	2.15**	4.1**	3.74**	3.64**
AR-1	-1.93	-1.95	-2	-1.41	-2.15	-	4.09**	1.27	3.41**
SVM	-4.09	-4.1	-4.13	-4.13	-4.1	-4.09	-	-4.04	-0.64
RF	-3.69	-3.85	-3.59	-3.06	-3.74	-1.27	4.04**	-	3.24**
LSTM	-3.62	-3.65	-3.59	-3.52	-3.64	-3.41	0.64	-3.24	-

Table 35: BVSP Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.84	-1.95	0.07	-1.67	1.73**	5.27**	2.26**	3.88**
C-HAR	1.84**	-	-0.16	1.26	1.56*	2.28**	5.17**	2.65**	4.16**
WLS	1.95**	0.16	-	4.26**	1.01	3.64**	5.57**	3.71**	4.18**
RR	-0.07	-1.26	-4.26	-	-0.87	2.47**	5.6**	3.14**	3.86**
HAR-J	1.67**	-1.56	-1.01	0.87	-	2.25**	5.23**	2.56**	4.04**
AR-1	-1.73	-2.28	-3.64	-2.47	-2.25	-	5.39**	1.23	3.42**
SVM	-5.27	-5.17	-5.57	-5.6	-5.23	-5.39	-	-5.52	-0.3
RF	-2.26	-2.65	-3.71	-3.14	-2.56	-1.23	5.52**	-	3.1**
LSTM	-3.88	-4.16	-4.18	-3.86	-4.04	-3.42	0.3	-3.1	-

Table 36: BVSP Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.09	-0.92	1.35*	-1.19	1.75**	1.62*	1.09	0.58
C-HAR	1.09	-	1.07	1.22	0.87	1.37*	1.54*	1.09	0.69
WLS	0.92	-1.07	-	1.37*	-1.17	1.74**	1.62*	1.09	0.59
RR	-1.35	-1.22	-1.37	-	-1.29	-0.43	1.68**	0.89	0.47
HAR-J	1.19	-0.87	1.17	1.29*	-	1.5*	1.57*	1.11	0.66
AR-1	-1.75	-1.37	-1.74	0.43	-1.5	-	1.58*	0.82	0.49
SVM	-1.62	-1.54	-1.62	-1.68	-1.57	-1.58	-	-1.98	-0.07
RF	-1.09	-1.09	-1.09	-0.89	-1.11	-0.82	1.98**	-	0.35
LSTM	-0.58	-0.69	-0.59	-0.47	-0.66	-0.49	0.07	-0.35	-

Table 37: DJI Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.54	-1.96	1.91**	-1.62	2.05**	3.83**	2.29**	3.24**
C-HAR	1.54*	-	0.64	1.78**	1.12	2.3**	3.78**	2.32**	3.35**
WLS	1.96**	-0.64	-	2.11**	0.46	2.46**	3.82**	2.36**	3.32**
RR	-1.91	-1.78	-2.11	-	-1.86	1.67**	3.84**	2.21**	3.17**
HAR-J	1.62*	-1.12	-0.46	1.86**	-	2.3**	3.8**	2.32**	3.31**
AR-1	-2.05	-2.3	-2.46	-1.67	-2.3	-	3.66**	1.7**	2.94**
SVM	-3.83	-3.78	-3.82	-3.84	-3.8	-3.66	-	-4.4	-0.3
RF	-2.29	-2.32	-2.36	-2.21	-2.32	-1.7	4.4**	-	2.03**
LSTM	-3.24	-3.35	-3.32	-3.17	-3.31	-2.94	0.3	-2.03	-

Table 38: DJI Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.98**	-0.91	1.82**	0.38	1.81**	5.73**	3.61**	3.34**
C-HAR	-1.98	-	-2.51	-0.57	-2.32	1.46*	5.32**	1.53*	3.06**
WLS	0.91	2.51**	-	1.68**	1.09	1.85**	5.58**	3.4**	3.42**
RR	-1.82	0.57	-1.68	-	-1.21	1.56*	5.88**	3.2**	3.12**
HAR-J	-0.38	2.32**	-1.09	1.21	-	1.79**	5.63**	3.06**	3.32**
AR-1	-1.81	-1.46	-1.85	-1.56	-1.79	-	3.17**	-1.07	1.48*
SVM	-5.73	-5.32	-5.58	-5.88	-5.63	-3.17	-	-5.85	-0.8
RF	-3.61	-1.53	-3.4	-3.2	-3.06	1.07	5.85**	-	2.7**
LSTM	-3.34	-3.06	-3.42	-3.12	-3.32	-1.48	0.8	-2.7	-

Table 39: DJI Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0.49	-1.01	0.89	-0.43	1.68**	2.47**	1.43*	1.11
C-HAR	-0.49	-	-0.73	0.33	-1.05	1.31*	2.57**	1.36*	1.05
WLS	1.01	0.73	-	1.04	0.45	1.7**	2.39**	1.36*	1.17
RR	-0.89	-0.33	-1.04	-	-1.68	0.9	2.54**	1.29*	1.04
HAR-J	0.43	1.05	-0.45	1.68**	-	1.65**	2.52**	1.52*	1.11
AR-1	-1.68	-1.31	-1.7	-0.9	-1.65	-	2.33**	0.24	0.94
SVM	-2.47	-2.57	-2.39	-2.54	-2.52	-2.33	-	-2.67	-0.28
RF	-1.43	-1.36	-1.36	-1.29	-1.52	-0.24	2.67**	-	0.89
LSTM	-1.11	-1.05	-1.17	-1.04	-1.11	-0.94	0.28	-0.89	-

Table 40: FCHI Diebold-Mariano Test statistics, k = 1 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0.85	-1.36	2.08**	-0.12	1.04	4.02**	2.3**	3.14**
C-HAR	-0.85	-	-1.37	0.35	-0.92	0.4	4.01**	2.06**	3.1**
WLS	1.36*	1.37*	-	1.67**	1.31*	1.94**	4.01**	2.43**	3.19**
RR	-2.08	-0.35	-1.67	-	-2.16	0.17	4.03**	2.09**	3.08**
HAR-J	0.12	0.92	-1.31	2.16**	-	1.03	4.02**	2.3**	3.14**
AR-1	-1.04	-0.4	-1.94	-0.17	-1.03	-	3.95**	1.84**	3.09**
SVM	-4.02	-4.01	-4.01	-4.03	-4.02	-3.95	-	-4.22	-0.27
RF	-2.3	-2.06	-2.43	-2.09	-2.3	-1.84	4.22**	-	2.75**
LSTM	-3.14	-3.1	-3.19	-3.08	-3.14	-3.09	0.27	-2.75	-

Table 41: FCHI Diebold-Mariano Test statistics, k = 5 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0.12	0.97	4.4**	1.26	1.78**	5.77**	-3.77	3.07**
C-HAR	-0.12	-	0.78	3.94**	0.27	1.62*	5.94**	-3.69	3.05**
WLS	-0.97	-0.78	-	-0.01	-0.9	0.49	4.76**	-3.54	2.89**
RR	-4.4	-3.94	0.01	-	-3.43	0.61	5.7**	-4.22	2.84**
HAR-J	-1.26	-0.27	0.9	3.43**	-	1.7**	5.76**	-3.93	3.05**
AR-1	-1.78	-1.62	-0.49	-0.61	-1.7	-	5.5**	-3.57	2.72**
SVM	-5.77	-5.94	-4.76	-5.7	-5.76	-5.5	-	-6.07	0.02
RF	3.77**	3.69**	3.54**	4.22**	3.93**	3.57**	6.07**	-	3.83**
LSTM	-3.07	-3.05	-2.89	-2.84	-3.05	-2.72	-0.02	-3.83	-

Table 42: FCHI Diebold-Mariano Test statistics, k = 22 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0.94	1.18	1.56*	0.91	-0.26	1.81**	0.88	1.45*
C-HAR	-0.94	-	1.1	0.41	-1	-0.64	1.88**	0.73	1.35*
WLS	-1.18	-1.1	-	-0.25	-1.37	-0.79	1.87**	0.4	1.31*
RR	-1.56	-0.41	0.25	-	-1.44	-0.93	1.82**	0.37	1.33*
HAR-J	-0.91	1	1.37*	1.44*	-	-0.56	1.86**	0.83	1.38*
AR-1	0.26	0.64	0.79	0.93	0.56	-	1.72**	0.69	1.48*
SVM	-1.81	-1.88	-1.87	-1.82	-1.86	-1.72	-	-1.96	-0.08
RF	-0.88	-0.73	-0.4	-0.37	-0.83	-0.69	1.96**	-	1.26
LSTM	-1.45	-1.35	-1.31	-1.33	-1.38	-1.48	0.08	-1.26	-

Table 43: FTMIB Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0.6	1.77**	1.92**	-0.18	-0.49	2.46**	1.16	2.04**
C-HAR	-0.6	-	0.14	0.27	-1.07	-0.6	2.49**	1.32*	2.01**
WLS	-1.77	-0.14	-	0.54	-1.52	-1.06	2.46**	1.01	1.99**
RR	-1.92	-0.27	-0.54	-	-1.33	-1.5	2.45**	0.87	1.97**
HAR-J	0.18	1.07	1.52*	1.33*	-	-0.26	2.47**	1.31*	2.05**
AR-1	0.49	0.6	1.06	1.5*	0.26	-	2.44**	1.09	2.05**
SVM	-2.46	-2.49	-2.46	-2.45	-2.47	-2.44	-	-2.57	-0.17
RF	-1.16	-1.32	-1.01	-0.87	-1.31	-1.09	2.57**	-	1.81**
LSTM	-2.04	-2.01	-1.99	-1.97	-2.05	-2.05	0.17	-1.81	-

Table 44: FTMIB Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.11	-0.31	3.43**	-1.59	-2.66	3.43**	-2.92	1.82**
C-HAR	1.11	-	0.82	2.37**	0.5	-2.52	3.42**	-2.83	1.89**
WLS	0.31	-0.82	-	2.23**	-0.69	-2.34	3.33**	-3.27	1.84**
RR	-3.43	-2.37	-2.23	-	-3.12	-3.22	3.27**	-3.23	1.69**
HAR-J	1.59*	-0.5	0.69	3.12**	-	-2.65	3.46**	-2.79	1.87**
AR-1	2.66**	2.52**	2.34**	3.22**	2.65**	-	3.81**	0.17	2.24**
SVM	-3.43	-3.42	-3.33	-3.27	-3.46	-3.81	-	-3.63	0.49
RF	2.92**	2.83**	3.27**	3.23**	2.79**	-0.17	3.63**	-	2.21**
LSTM	-1.82	-1.89	-1.84	-1.69	-1.87	-2.24	-0.49	-2.21	-

Table 45: FTMIB Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.46*	-1	1.26	0.17	1.19	1.31*	1.02	0.3
C-HAR	-1.46	-	-1.09	1.2	-1.28	1.17	1.3*	0.95	0.29
WLS	1	1.09	-	1.16	1.03	1.15	1.28	1.02	0.34
RR	-1.26	-1.2	-1.16	-	-1.23	1.15	1.32*	0.09	0.25
HAR-J	-0.17	1.28*	-1.03	1.23	-	1.18	1.3*	1	0.3
AR-1	-1.19	-1.17	-1.15	-1.15	-1.18	-	1.42*	-1.28	0.15
SVM	-1.31	-1.3	-1.28	-1.32	-1.3	-1.42	-	-1.37	-0.02
RF	-1.02	-0.95	-1.02	-0.09	-1	1.28	1.37*	-	0.24
LSTM	-0.3	-0.29	-0.34	-0.25	-0.3	-0.15	0.02	-0.24	-

Table 46: FTSE Diebold-Mariano Test statistics, k = 1 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.44*	-2.04	2.16**	0.52	1.92**	2.97**	2.14**	2.31**
C-HAR	-1.44	-	-1.92	-0.17	-1.41	1.13	2.98**	2.1**	2.24**
WLS	2.04**	1.92**	-	2.1**	2.08**	2.24**	2.97**	2.21**	2.39**
RR	-2.16	0.17	-2.1	-	-2.07	1.42*	2.96**	2.06**	2.24**
HAR-J	-0.52	1.41*	-2.08	2.07**	-	1.92**	2.97**	2.13**	2.31**
AR-1	-1.92	-1.13	-2.24	-1.42	-1.92	-	2.81**	1.65*	2.11**
SVM	-2.97	-2.98	-2.97	-2.96	-2.97	-2.81	-	-3.42	-0.21
RF	-2.14	-2.1	-2.21	-2.06	-2.13	-1.65	3.42**	-	1.38*
LSTM	-2.31	-2.24	-2.39	-2.24	-2.31	-2.11	0.21	-1.38	-

Table 47: FTSE Diebold-Mariano Test statistics, k = 5 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.1	0.81	2.87**	-0.62	1.82**	5.29**	0.49	2.51**
C-HAR	1.1	-	0.96	3.08**	1.12	1.8**	5.2**	1.04	2.6**
WLS	-0.81	-0.96	-	0.48	-0.84	1.83**	4.7**	-0.42	2.35**
RR	-2.87	-3.08	-0.48	-	-3.07	1.44*	5.31**	-2.02	2.14**
HAR-J	0.62	-1.12	0.84	3.07**	-	1.78**	5.24**	0.67	2.53**
AR-1	-1.82	-1.8	-1.83	-1.44	-1.78	-	2.09**	-1.72	0.75
SVM	-5.29	-5.2	-4.7	-5.31	-5.24	-2.09	-	-5.74	-0.56
RF	-0.49	-1.04	0.42	2.02**	-0.67	1.72**	5.74**	-	2.44**
LSTM	-2.51	-2.6	-2.35	-2.14	-2.53	-0.75	0.56	-2.44	-

Table 48: FTSE Diebold-Mariano Test statistics, k = 22 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.63	-1.22	1.48*	-1.79	0.98	1.9**	1.07	1.02
C-HAR	1.63*	-	1.01	1.6*	1.36*	1.59*	1.88**	1.32*	1.14
WLS	1.22	-1.01	-	1.35*	0.01	1.39*	1.86**	1.24	1.11
RR	-1.48	-1.6	-1.35	-	-1.67	-0.23	1.92**	0.79	0.97
HAR-J	1.79**	-1.36	-0.01	1.67**	-	1.59*	1.9**	1.3*	1.1
AR-1	-0.98	-1.59	-1.39	0.23	-1.59	-	1.9**	0.72	0.98
SVM	-1.9	-1.88	-1.86	-1.92	-1.9	-1.9	-	-1.97	0
RF	-1.07	-1.32	-1.24	-0.79	-1.3	-0.72	1.97**	-	0.89
LSTM	-1.02	-1.14	-1.11	-0.97	-1.1	-0.98	0	-0.89	-

Table 49: GDAXI Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.56	0.07	0.25	-1.32	0.86	3.46**	1.27	2.76**
C-HAR	1.56*	-	1.31*	1.69**	1.67**	1.45*	3.49**	1.79**	2.84**
WLS	-0.07	-1.31	-	0.12	-1.01	0.79	3.43**	1.18	2.76**
RR	-0.25	-1.69	-0.12	-	-1.57	0.81	3.5**	1.29*	2.74**
HAR-J	1.32*	-1.67	1.01	1.57*	-	1.34*	3.49**	1.66**	2.81**
AR-1	-0.86	-1.45	-0.79	-0.81	-1.34	-	3.45**	0.39	2.62**
SVM	-3.46	-3.49	-3.43	-3.5	-3.49	-3.45	-	-3.69	-0.27
RF	-1.27	-1.79	-1.18	-1.29	-1.66	-0.39	3.69**	-	2.55**
LSTM	-2.76	-2.84	-2.76	-2.74	-2.81	-2.62	0.27	-2.55	-

Table 50: GDAXI Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.43	2**	2.98**	-0.97	1.33*	4.62**	-2.55	2.67**
C-HAR	0.43	-	1.84**	1.94**	-0.41	1.46*	4.59**	-2.55	2.73**
WLS	-2	-1.84	-	-1.2	-1.96	0.16	3.56**	-2.89	2.21**
RR	-2.98	-1.94	1.2	-	-2.98	0.92	4.51**	-2.86	2.49**
HAR-J	0.97	0.41	1.96**	2.98**	-	1.49*	4.67**	-2.4	2.72**
AR-1	-1.33	-1.46	-0.16	-0.92	-1.49	-	3.29**	-2.19	2.05**
SVM	-4.62	-4.59	-3.56	-4.51	-4.67	-3.29	-	-5.09	0.09
RF	2.55**	2.55**	2.89**	2.86**	2.4**	2.19**	5.09**	-	3.31**
LSTM	-2.67	-2.73	-2.21	-2.49	-2.72	-2.05	-0.09	-3.31	-

Table 51: GDAXI Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.07	0.77	0.88	-1.06	0.31	0.97	0.93	0.28
C-HAR	1.07	-	1.03	1.01	1.15	1.25	1	1	0.5
WLS	-0.77	-1.03	-	0.92	-1.03	-0.09	0.98	0.95	0.26
RR	-0.88	-1.01	-0.92	-	-1	-0.53	1	0.97	0.21
HAR-J	1.06	-1.15	1.03	1	-	1.25	1	0.99	0.48
AR-1	-0.31	-1.25	0.09	0.53	-1.25	-	0.86	0.77	0.28
SVM	-0.97	-1	-0.98	-1	-1	-0.86	-	-1.02	-0.01
RF	-0.93	-1	-0.95	-0.97	-0.99	-0.77	1.02	-	0.1
LSTM	-0.28	-0.5	-0.26	-0.21	-0.48	-0.28	0.01	-0.1	-

Table 52: GSPTSE Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-2	0.91	1.48*	-1.63	1.14	2.11**	1.46*	1.14
C-HAR	2**	-	1.73**	2.06**	2.06**	1.37*	2.24**	2.15**	1.72**
WLS	-0.91	-1.73	-	-0.19	-1.37	1.18	1.88**	0.73	0.98
RR	-1.48	-2.06	0.19	-	-1.89	1.03	2.13**	1.22	1.02
HAR-J	1.63*	-2.06	1.37*	1.89**	-	1.24	2.2**	1.98**	1.5*
AR-1	-1.14	-1.37	-1.18	-1.03	-1.24	-	0.22	-0.72	0.07
SVM	-2.11	-2.24	-1.88	-2.13	-2.2	-0.22	-	-2.27	-0.08
RF	-1.46	-2.15	-0.73	-1.22	-1.98	0.72	2.27**	-	0.75
LSTM	-1.14	-1.72	-0.98	-1.02	-1.5	-0.07	0.08	-0.75	-

Table 53: GSPTSE Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.63	-1.69	-1.74	-1.15	0.93	-1.71	-1.72	-1.71
C-HAR	1.63*	-	1.39*	-2.81	2.12**	0.99	-2.53	-2.63	-2.57
WLS	1.69**	-1.39	-	-1.89	2.48**	0.98	-1.76	-1.78	-1.77
RR	1.74**	2.81**	1.89**	-	2.27**	1	-0.72	-0.92	-0.82
HAR-J	1.15	-2.12	-2.48	-2.27	-	0.94	-2.19	-2.21	-2.2
AR-1	-0.93	-0.99	-0.98	-1	-0.94	-	-1	-1	-1
SVM	1.71**	2.53**	1.76**	0.72	2.19**	1	-	-2.67	-0.66
RF	1.72**	2.63**	1.78**	0.92	2.21**	1	2.67**	-	0.74
LSTM	1.71**	2.57**	1.77**	0.82	2.2**	1	0.66	-0.74	-

Table 54: GSPTSE Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.16	-0.66	1.58*	-0.15	0.92	1.83**	1.45*	0.95
C-HAR	1.16	-	0.98	1.64*	0.67	1.18	1.83**	1.44*	0.97
WLS	0.66	-0.98	-	1.55*	-0.04	0.92	1.82**	1.45*	0.95
RR	-1.58	-1.64	-1.55	-	-1.7	-0.74	1.85**	1.09	0.77
HAR-J	0.15	-0.67	0.04	1.7**	-	1.33*	1.83**	1.38*	0.95
AR-1	-0.92	-1.18	-0.92	0.74	-1.33	-	1.8**	1	0.82
SVM	-1.83	-1.83	-1.82	-1.85	-1.83	-1.8	-	-1.92	-0.16
RF	-1.45	-1.44	-1.45	-1.09	-1.38	-1	1.92**	-	0.58
LSTM	-0.95	-0.97	-0.95	-0.77	-0.95	-0.82	0.16	-0.58	-

Table 55: HSI Diebold-Mariano Test statistics, k = 1 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.07	-1.77	1.67**	-0.19	1.92**	2.53**	1.87**	2.04**
C-HAR	0.07	-	-0.75	1.29*	-0.46	1.97**	2.52**	1.86**	2.04**
WLS	1.77**	0.75	-	1.72**	0.6	1.92**	2.53**	1.91**	2.07**
RR	-1.67	-1.29	-1.72	-	-1.43	1.95**	2.53**	1.79**	1.99**
HAR-J	0.19	0.46	-0.6	1.43*	-	2.07**	2.53**	1.87**	2.04**
AR-1	-1.92	-1.97	-1.92	-1.95	-2.07	-	2.47**	1.36*	1.84**
SVM	-2.53	-2.52	-2.53	-2.53	-2.53	-2.47	-	-2.72	-0.36
RF	-1.87	-1.86	-1.91	-1.79	-1.87	-1.36	2.72**	-	1.38*
LSTM	-2.04	-2.04	-2.07	-1.99	-2.04	-1.84	0.36	-1.38	-

Table 56: HSI Diebold-Mariano Test statistics, k = 5 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	2.57**	-1.31	-1.64	-0.16	3.1**	4.67**	1.87**	3.38**
C-HAR	-2.57	-	-2.18	-2.66	-2.97	2.87**	4.67**	1.57*	3.24**
WLS	1.31*	2.18**	-	-0.52	1.04	2.97**	4.61**	1.85**	3.47**
RR	1.64*	2.66**	0.52	-	1.48*	3.8**	4.9**	2.43**	3.52**
HAR-J	0.16	2.97**	-1.04	-1.48	-	3.1**	4.68**	1.88**	3.39**
AR-1	-3.1	-2.87	-2.97	-3.8	-3.1	-	4.53**	-1.42	2.2**
SVM	-4.67	-4.67	-4.61	-4.9	-4.68	-4.53	-	-4.89	-0.94
RF	-1.87	-1.57	-1.85	-2.43	-1.88	1.42*	4.89**	-	2.62**
LSTM	-3.38	-3.24	-3.47	-3.52	-3.39	-2.2	0.94	-2.62	-

Table 57: HSI Diebold-Mariano Test statistics, k = 22 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.94	-1.63	1.71**	-0.44	-0.46	2.5**	1.39*	0.49
C-HAR	0.94	-	-0.65	1.85**	0.46	-0.23	2.49**	1.44*	0.5
WLS	1.63*	0.65	-	1.91**	1.11	-0.06	2.49**	1.48*	0.5
RR	-1.71	-1.85	-1.91	-	-1.61	-1.02	2.55**	0.79	0.43
HAR-J	0.44	-0.46	-1.11	1.61*	-	-0.35	2.47**	1.4*	0.49
AR-1	0.46	0.23	0.06	1.02	0.35	-	2.31**	1.08	0.51
SVM	-2.5	-2.49	-2.49	-2.55	-2.47	-2.31	-	-2.72	-0.15
RF	-1.39	-1.44	-1.48	-0.79	-1.4	-1.08	2.72**	-	0.41
LSTM	-0.49	-0.5	-0.5	-0.43	-0.49	-0.51	0.15	-0.41	-

Table 58: IBEX Diebold-Mariano Test statistics, k = 1 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.12	-0.73	1.93**	-1.89	1.08	3.67**	0.88	2.72**
C-HAR	1.12	-	0.19	2.22**	-0.29	1.17	3.67**	1.11	2.76**
WLS	0.73	-0.19	-	1.27	-0.52	1.25	3.62**	1.01	2.76**
RR	-1.93	-2.22	-1.27	-	-2.27	0.83	3.67**	0.59	2.66**
HAR-J	1.89**	0.29	0.52	2.27**	-	1.23	3.66**	1.1	2.76**
AR-1	-1.08	-1.17	-1.25	-0.83	-1.23	-	3.26**	-0.37	2.43**
SVM	-3.67	-3.67	-3.62	-3.67	-3.66	-3.26	-	-3.89	-0.21
RF	-0.88	-1.11	-1.01	-0.59	-1.1	0.37	3.89**	-	2.58**
LSTM	-2.72	-2.76	-2.76	-2.66	-2.76	-2.43	0.21	-2.58	-

Table 59: IBEX Diebold-Mariano Test statistics, k = 5 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-2.09	2.61**	5.17**	-0.36	1.41*	6.94**	-3.54	2.44**
C-HAR	2.09**	-	2.66**	4.23**	1.41*	1.45*	7.23**	-3.25	2.63**
WLS	-2.61	-2.66	-	-1.69	-2.99	1.15	4.43**	-3.6	1.73**
RR	-5.17	-4.23	1.69**	-	-3.25	1.29*	6.68**	-4.48	2.2**
HAR-J	0.36	-1.41	2.99**	3.25**	-	1.45*	6.76**	-3.24	2.46**
AR-1	-1.41	-1.45	-1.15	-1.29	-1.45	-	0.41	-1.67	0.02
SVM	-6.94	-7.23	-4.43	-6.68	-6.76	-0.41	-	-7.61	-0.6
RF	3.54**	3.25**	3.6**	4.48**	3.24**	1.67**	7.61**	-	3.08**
LSTM	-2.44	-2.63	-1.73	-2.2	-2.46	-0.02	0.6	-3.08	-

Table 60: IBEX Diebold-Mariano Test statistics, k = 22 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.57	-1.12	1.81**	-1.1	-0.22	1.48*	1	0.78
C-HAR	1.57*	-	-1	1.78**	-0.6	-0.03	1.48*	1.17	0.8
WLS	1.12	1	-	1.21	1.11	1.75**	1.45*	1.09	0.87
RR	-1.81	-1.78	-1.21	-	-1.32	-0.37	1.47*	0.75	0.76
HAR-J	1.1	0.6	-1.11	1.32*	-	0.12	1.46*	1.06	0.81
AR-1	0.22	0.03	-1.75	0.37	-0.12	-	1.37*	0.49	0.82
SVM	-1.48	-1.48	-1.45	-1.47	-1.46	-1.37	-	-1.5	-0.21
RF	-1	-1.17	-1.09	-0.75	-1.06	-0.49	1.5*	-	0.73
LSTM	-0.78	-0.8	-0.87	-0.76	-0.81	-0.82	0.21	-0.73	-

Table 61: IXIC Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.35	-1.78	-0.23	-0.51	0.42	2.9**	1.01	2.25**
C-HAR	0.35	-	-1.33	0.17	0.34	0.48	2.93**	1.13	2.27**
WLS	1.78**	1.33*	-	1.72**	1.8**	1.32*	2.89**	1.2	2.31**
RR	0.23	-0.17	-1.72	-	0.2	0.56	2.89**	0.97	2.25**
HAR-J	0.51	-0.34	-1.8	-0.2	-	0.43	2.9**	1.01	2.25**
AR-1	-0.42	-0.48	-1.32	-0.56	-0.43	-	2.81**	0.63	2.2**
SVM	-2.9	-2.93	-2.89	-2.89	-2.9	-2.81	-	-3.08	-0.19
RF	-1.01	-1.13	-1.2	-0.97	-1.01	-0.63	3.08**	-	2.06**
LSTM	-2.25	-2.27	-2.31	-2.25	-2.25	-2.2	0.19	-2.06	-

Table 62: IXIC Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-2.01	1.52*	-0.91	0.88	0.78	5.02**	-2.98	2.26**
C-HAR	2.01**	-	1.69**	2.52**	1.88**	0.91	5.89**	-2.72	2.56**
WLS	-1.52	-1.69	-	-1.46	-1.57	0.47	2.51**	-2.2	1.44*
RR	0.91	-2.52	1.46*	-	1	0.8	5.4**	-3.38	2.31**
HAR-J	-0.88	-1.88	1.57*	-1	-	0.78	4.89**	-2.91	2.21**
AR-1	-0.78	-0.91	-0.47	-0.8	-0.78	-	0.79	-1.15	0.57
SVM	-5.02	-5.89	-2.51	-5.4	-4.89	-0.79	-	-6.73	-0.19
RF	2.98**	2.72**	2.2**	3.38**	2.91**	1.15	6.73**	-	3**
LSTM	-2.26	-2.56	-1.44	-2.31	-2.21	-0.57	0.19	-3	-

Table 63: IXIC Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.5*	-1.82	-0.22	1.29*	0.72	1.38*	1.13	0.97
C-HAR	-1.5	-	-1.73	-0.94	-1.02	-0.02	1.34*	1.05	0.89
WLS	1.82**	1.73**	-	1.06	1.66**	1.4*	1.46*	1.3*	1.07
RR	0.22	0.94	-1.06	-	0.6	1.22	1.5*	1.35*	0.97
HAR-J	-1.29	1.02	-1.66	-0.6	-	0.29	1.34*	1.05	0.92
AR-1	-0.72	0.02	-1.4	-1.22	-0.29	-	1.43*	1.2	0.88
SVM	-1.38	-1.34	-1.46	-1.5	-1.34	-1.43	-	-1.65	-0.09
RF	-1.13	-1.05	-1.3	-1.35	-1.05	-1.2	1.65**	-	0.41
LSTM	-0.97	-0.89	-1.07	-0.97	-0.92	-0.88	0.09	-0.41	-

Table 64: KS11 Diebold-Mariano Test statistics, k = 1 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.65*	-1.66	-1.52	1.18	-0.92	2.37**	1.88**	2.48**
C-HAR	-1.65	-	-1.68	-1.61	-1.73	-1.53	2.28**	1.59*	2.34**
WLS	1.66**	1.68**	-	-0.35	1.51*	0.7	2.47**	2.15**	2.62**
RR	1.52*	1.61*	0.35	-	1.44*	0.75	2.49**	2.17**	2.62**
HAR-J	-1.18	1.73**	-1.51	-1.44	-	-1.41	2.32**	1.73**	2.42**
AR-1	0.92	1.53*	-0.7	-0.75	1.41*	-	2.41**	1.98**	2.55**
SVM	-2.37	-2.28	-2.47	-2.49	-2.32	-2.41	-	-2.72	0.31
RF	-1.88	-1.59	-2.15	-2.17	-1.73	-1.98	2.72**	-	1.57*
LSTM	-2.48	-2.34	-2.62	-2.62	-2.42	-2.55	-0.31	-1.57	-

Table 65: KS11 Diebold-Mariano Test statistics, k = 5 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.45*	-1.41	-1.67	1.1	0.34	3.73**	0.47	2.14**
C-HAR	-1.45	-	-1.55	-1.59	-0.15	-0.3	3.25**	-0.21	1.89**
WLS	1.41*	1.55*	-	-1.12	1.27	0.6	3.77**	1.06	2.32**
RR	1.67**	1.59*	1.12	-	1.42*	0.9	4.17**	2.24**	2.47**
HAR-J	-1.1	0.15	-1.27	-1.42	-	-0.32	3.19**	-0.17	1.89**
AR-1	-0.34	0.3	-0.6	-0.9	0.32	-	3**	0.01	1.88**
SVM	-3.73	-3.25	-3.77	-4.17	-3.19	-3	-	-4.52	-0.09
RF	-0.47	0.21	-1.06	-2.24	0.17	-0.01	4.52**	-	2.02**
LSTM	-2.14	-1.89	-2.32	-2.47	-1.89	-1.88	0.09	-2.02	-

Table 66: KS11 Diebold-Mariano Test statistics, k = 22 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.37	-0.83	1.67**	-0.11	1.57*	1.81**	-1.03	0.49
C-HAR	0.37	-	-1.15	1.24	0.38	1.25	1.68**	-1.5	0.51
WLS	0.83	1.15	-	1.31*	0.85	1.3*	1.67**	0.34	0.54
RR	-1.67	-1.24	-1.31	-	-1.63	1.13	1.83**	-1.49	0.38
HAR-J	0.11	-0.38	-0.85	1.63*	-	1.54*	1.8**	-1.05	0.49
AR-1	-1.57	-1.25	-1.3	-1.13	-1.54	-	1.87**	-1.44	0.35
SVM	-1.81	-1.68	-1.67	-1.83	-1.8	-1.87	-	-1.75	-0.03
RF	1.03	1.5*	-0.34	1.49*	1.05	1.44*	1.75**	-	0.53
LSTM	-0.49	-0.51	-0.54	-0.38	-0.49	-0.35	0.03	-0.53	-

Table 67: KSE Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	2.33**	2.2**	3.6**	-0.65	2.52**	4.73**	2.06**	3.17**
C-HAR	-2.33	-	1.21	-1.6	-2.3	1.23	4.68**	-0.44	2.81**
WLS	-2.2	-1.21	-	-2.06	-2.19	-0.88	5.14**	-1.51	2.14**
RR	-3.6	1.6*	2.06**	-	-3.01	2.16**	4.71**	1.32*	3.05**
HAR-J	0.65	2.3**	2.19**	3.01**	-	2.43**	4.7**	2.03**	3.18**
AR-1	-2.52	-1.23	0.88	-2.16	-2.43	-	4.73**	-1.15	2.57**
SVM	-4.73	-4.68	-5.14	-4.71	-4.7	-4.73	-	-5	-0.59
RF	-2.06	0.44	1.51*	-1.32	-2.03	1.15	5**	-	2.85**
LSTM	-3.17	-2.81	-2.14	-3.05	-3.18	-2.57	0.59	-2.85	-

Table 68: KSE Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.16	7.74**	6.3**	0.61	1.96**	7.77**	-1.2	5.03**
C-HAR	-1.16	-	7.37**	2.73**	-1.11	1.67**	7.96**	-2.95	4.88**
WLS	-7.74	-7.37	-	-6.39	-7.18	-2.26	3.48**	-7.59	1.79**
RR	-6.3	-2.73	6.39**	-	-4.51	1.27	7.6**	-4.02	4.43**
HAR-J	-0.61	1.11	7.18**	4.51**	-	1.78**	7.95**	-1.96	4.96**
AR-1	-1.96	-1.67	2.26**	-1.27	-1.78	-	4.51**	-2.04	2.76**
SVM	-7.77	-7.96	-3.48	-7.6	-7.95	-4.51	-	-8.57	-0.45
RF	1.2	2.95**	7.59**	4.02**	1.96**	2.04**	8.57**	-	5.2**
LSTM	-5.03	-4.88	-1.79	-4.43	-4.96	-2.76	0.45	-5.2	-

Table 69: KSE Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.08	1.16	1.55*	-1.1	1.46*	1.64*	-0.16	0.29
C-HAR	1.08	-	1.18	1.49*	-1.02	1.43*	1.62*	0.3	0.31
WLS	-1.16	-1.18	-	1.72**	-1.15	1.49*	1.69**	-0.47	0.26
RR	-1.55	-1.49	-1.72	-	-1.42	0.93	1.67**	-0.99	0.2
HAR-J	1.1	1.02	1.15	1.42*	-	1.4*	1.59*	0.97	0.32
AR-1	-1.46	-1.43	-1.49	-0.93	-1.4	-	1.67**	-1.07	0.16
SVM	-1.64	-1.62	-1.69	-1.67	-1.59	-1.67	-	-1.43	-0.03
RF	0.16	-0.3	0.47	0.99	-0.97	1.07	1.43*	-	0.3
LSTM	-0.29	-0.31	-0.26	-0.2	-0.32	-0.16	0.03	-0.3	-

Table 70: MXX Diebold-Mariano Test statistics, k = 1 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0.27	0.08	1.69**	-1.02	1.92**	3.48**	2.14**	2.99**
C-HAR	-0.27	-	-0.28	1.85**	-0.91	1.86**	3.55**	2.05**	2.97**
WLS	-0.08	0.28	-	2.21**	-0.86	1.88**	3.52**	2.12**	2.99**
RR	-1.69	-1.85	-2.21	-	-2.31	1.66**	3.5**	1.71**	2.8**
HAR-J	1.02	0.91	0.86	2.31**	-	2.1**	3.5**	2.17**	3.03**
AR-1	-1.92	-1.86	-1.88	-1.66	-2.1	-	2.91**	-0.02	1.98**
SVM	-3.48	-3.55	-3.52	-3.5	-3.5	-2.91	-	-3.48	-0.49
RF	-2.14	-2.05	-2.12	-1.71	-2.17	0.02	3.48**	-	2.03**
LSTM	-2.99	-2.97	-2.99	-2.8	-3.03	-1.98	0.49	-2.03	-

Table 71: MXX Diebold-Mariano Test statistics, k = 5 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-2.02	-2.7	-2.41	-2.03	1.54*	3.97**	1.17	1.57*
C-HAR	2.02**	-	0.68	0.22	0.9	1.87**	4.59**	2.57**	1.97**
WLS	2.7**	-0.68	-	-0.77	0.07	1.74**	4.26**	2.06**	1.89**
RR	2.41**	-0.22	0.77	-	0.78	1.85**	4.54**	2.55**	1.97**
HAR-J	2.03**	-0.9	-0.07	-0.78	-	1.82**	4.22**	1.99**	1.89**
AR-1	-1.54	-1.87	-1.74	-1.85	-1.82	-	1.18	-1.09	-0.01
SVM	-3.97	-4.59	-4.26	-4.54	-4.22	-1.18	-	-5.13	-1.01
RF	-1.17	-2.57	-2.06	-2.55	-1.99	1.09	5.13**	-	1.08
LSTM	-1.57	-1.97	-1.89	-1.97	-1.89	0.01	1.01	-1.08	-

Table 72: MXX Diebold-Mariano Test statistics, k = 22 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.04	0.68	1.94**	-1.59	0.42	3.04**	2.74**	0.94
C-HAR	1.04	-	1.73**	1.89**	0.11	0.68	2.98**	2.72**	0.97
WLS	-0.68	-1.73	-	1.9**	-1.56	0.11	3.03**	2.78**	0.92
RR	-1.94	-1.89	-1.9	-	-1.93	-0.82	3.12**	2.43**	0.83
HAR-J	1.59*	-0.11	1.56*	1.93**	-	0.8	3.01**	2.78**	0.97
AR-1	-0.42	-0.68	-0.11	0.82	-0.8	-	2.92**	2**	0.92
SVM	-3.04	-2.98	-3.03	-3.12	-3.01	-2.92	-	-2.98	-0.15
RF	-2.74	-2.72	-2.78	-2.43	-2.78	-2	2.98**	-	0.63
LSTM	-0.94	-0.97	-0.92	-0.83	-0.97	-0.92	0.15	-0.63	-

Table 73: N225 Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.01	-1.83	1.48*	-1.13	1.48*	3.66**	2.97**	3.2**
C-HAR	0.01	-	-0.87	0.61	-0.59	1.29*	3.61**	2.83**	3.21**
WLS	1.83**	0.87	-	2.21**	0.74	1.7**	3.65**	3.02**	3.24**
RR	-1.48	-0.61	-2.21	-	-1.69	1.17	3.66**	2.93**	3.16**
HAR-J	1.13	0.59	-0.74	1.69**	-	1.76**	3.65**	2.98**	3.22**
AR-1	-1.48	-1.29	-1.7	-1.17	-1.76	-	3.57**	2.21**	3.06**
SVM	-3.66	-3.61	-3.65	-3.66	-3.65	-3.57	-	-3.75	-0.25
RF	-2.97	-2.83	-3.02	-2.93	-2.98	-2.21	3.75**	-	2.5**
LSTM	-3.2	-3.21	-3.24	-3.16	-3.22	-3.06	0.25	-2.5	-

Table 74: N225 Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0.22	0.55	1.59*	-0.05	1.96**	4.54**	1.88**	3.27**
C-HAR	-0.22	-	0.61	0.93	-0.32	1.9**	4.47**	1.7**	3.26**
WLS	-0.55	-0.61	-	0.4	-0.63	1.77**	4.33**	1.35*	3.26**
RR	-1.59	-0.93	-0.4	-	-1.42	1.67**	4.64**	1.6*	3.16**
HAR-J	0.05	0.32	0.63	1.42*	-	1.96**	4.52**	1.86**	3.28**
AR-1	-1.96	-1.9	-1.77	-1.67	-1.96	-	3.33**	-0.83	2.37**
SVM	-4.54	-4.47	-4.33	-4.64	-4.52	-3.33	-	-4.61	-0.1
RF	-1.88	-1.7	-1.35	-1.6	-1.86	0.83	4.61**	-	2.82**
LSTM	-3.27	-3.26	-3.26	-3.16	-3.28	-2.37	0.1	-2.82	-

Table 75: N225 Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.92	1.36*	1.02	-0.73	-1.33	1.38*	0	0.27
C-HAR	0.92	-	1.07	0.95	-0.47	0.74	1.21	1.07	0.47
WLS	-1.36	-1.07	-	-0.22	-0.82	-1.35	1.35*	-1.02	0.21
RR	-1.02	-0.95	0.22	-	-0.77	-1.28	1.43*	-0.59	0.22
HAR-J	0.73	0.47	0.82	0.77	-	0.63	1.02	0.78	0.57
AR-1	1.33*	-0.74	1.35*	1.28	-0.63	-	1.39*	0.71	0.32
SVM	-1.38	-1.21	-1.35	-1.43	-1.02	-1.39	-	-1.3	-0.04
RF	0	-1.07	1.02	0.59	-0.78	-0.71	1.3*	-	0.27
LSTM	-0.27	-0.47	-0.21	-0.22	-0.57	-0.32	0.04	-0.27	-

Table 76: NSEI Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.81	-0.3	-0.62	1	0.97	2.85**	0.48	1.05
C-HAR	0.81	-	1.14	0.69	1	0.98	2.75**	1.26	1.19
WLS	0.3	-1.14	-	-0.5	1	0.88	2.85**	0.97	1.09
RR	0.62	-0.69	0.5	-	1	0.93	2.89**	1.05	1.12
HAR-J	-1	-1	-1	-1	-	-1	-0.99	-1	-0.99
AR-1	-0.97	-0.98	-0.88	-0.93	1	-	0.56	-0.77	0.18
SVM	-2.85	-2.75	-2.85	-2.89	0.99	-0.56	-	-3.26	-0.21
RF	-0.48	-1.26	-0.97	-1.05	1	0.77	3.26**	-	0.96
LSTM	-1.05	-1.19	-1.09	-1.12	0.99	-0.18	0.21	-0.96	-

Table 77: NSEI Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.82	1.28	-1.26	1	1	1.05	-2.55	-0.19
C-HAR	1.82**	-	2.81**	2.23**	1	1.02	7.71**	-4.77	1.91**
WLS	-1.28	-2.81	-	-1.85	1	0.92	-0.6	-3.34	-1.36
RR	1.26	-2.23	1.85**	-	1	1	2.79**	-3.44	0.33
HAR-J	-1	-1	-1	-1	-	-1	-1	-1.01	-1
AR-1	-1	-1.02	-0.92	-1	1	-	-0.94	-1.04	-0.98
SVM	-1.05	-7.71	0.6	-2.79	1	0.94	-	-9.02	-1.31
RF	2.55**	4.77**	3.34**	3.44**	1.01	1.04	9.02**	-	3.01**
LSTM	0.19	-1.91	1.36*	-0.33	1	0.98	1.31*	-3.01	-

Table 78: NSEI Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.06	0.55	1.2	1.02	0.96	-0.52	-1.58	0.01
C-HAR	-1.06	-	-1.03	-1.01	0.87	-1.84	-1.04	-1.09	-0.11
WLS	-0.55	1.03	-	1.05	1	0.94	-1.06	-1.95	0.01
RR	-1.2	1.01	-1.05	-	0.98	0.91	-1.16	-1.63	0.01
HAR-J	-1.02	-0.87	-1	-0.98	-	-1.18	-1	-1.05	-0.13
AR-1	-0.96	1.84**	-0.94	-0.91	1.18	-	-0.94	-1	-0.09
SVM	0.52	1.04	1.06	1.16	1	0.94	-	-1.77	0.02
RF	1.58*	1.09	1.95**	1.63*	1.05	1	1.77**	-	0.02
LSTM	-0.01	0.11	-0.01	-0.01	0.13	0.09	-0.02	-0.02	-

Table 79: OMXC20 Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.04	-0.89	-0.93	1.01	1.01	-0.85	-1.04	-0.41
C-HAR	-1.04	-	-1.04	-1.04	-1.07	-1.14	-1.03	-1.04	-1
WLS	0.89	1.04	-	0.6	0.99	1	-0.68	-1.32	0.05
RR	0.93	1.04	-0.6	-	1	1	-0.67	-1.18	-0.02
HAR-J	-1.01	1.07	-0.99	-1	-	1.01	-0.99	-1.01	-0.92
AR-1	-1.01	1.14	-1	-1	-1.01	-	-0.99	-1.01	-0.95
SVM	0.85	1.03	0.68	0.67	0.99	0.99	-	-1.74	0.18
RF	1.04	1.04	1.32*	1.18	1.01	1.01	1.74**	-	0.66
LSTM	0.41	1	-0.05	0.02	0.92	0.95	-0.18	-0.66	-

Table 80: OMXC20 Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.11	2.14**	-1.04	1	1.01	-1.04	-0.97	0.2
C-HAR	-1.11	-	-1.07	-1.11	-0.76	0.87	-1.1	-1.1	-0.98
WLS	-2.14	1.07	-	-1.48	0.96	0.98	-1.37	-1.32	-0.31
RR	1.04	1.11	1.48*	-	1.01	1.01	-1.03	-0.84	1.23
HAR-J	-1	0.76	-0.96	-1.01	-	1.02	-1.01	-1	-0.89
AR-1	-1.01	-0.87	-0.98	-1.01	-1.02	-	-1.01	-1.01	-0.94
SVM	1.04	1.1	1.37*	1.03	1.01	1.01	-	1.46*	1.91**
RF	0.97	1.1	1.32*	0.84	1	1.01	-1.46	-	1.82**
LSTM	-0.2	0.98	0.31	-1.23	0.89	0.94	-1.91	-1.82	-

Table 81: OMXC20 Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0.52	1.02	1.22	0.56	0.81	0.29	-1.47	0
C-HAR	-0.52	-	-0.2	-0.27	0.64	0.89	-0.24	-0.85	-0.01
WLS	-1.02	0.2	-	-0.41	0.38	0.79	-0.28	-1.28	0
RR	-1.22	0.27	0.41	-	0.39	0.76	-0.2	-1.41	0
HAR-J	-0.56	-0.64	-0.38	-0.39	-	0.92	-0.34	-0.8	-0.01
AR-1	-0.81	-0.89	-0.79	-0.76	-0.92	-	-0.71	-0.88	-0.05
SVM	-0.29	0.24	0.28	0.2	0.34	0.71	-	-1	0
RF	1.47*	0.85	1.28*	1.41*	0.8	0.88	1	-	0.01
LSTM	0	0.01	0	0	0.01	0.05	0	-0.01	-

Table 82: OMXHPI Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.11	1.29*	-0.86	0.99	0.99	-0.81	-1.03	-0.09
C-HAR	-1.11	-	-1.1	-1.07	0.58	0.97	-1.07	-1.09	-0.65
WLS	-1.29	1.1	-	-0.95	0.98	0.99	-0.92	-1.06	-0.14
RR	0.86	1.07	0.95	-	0.97	0.99	0.92	-1.23	0.07
HAR-J	-0.99	-0.58	-0.98	-0.97	-	0.99	-0.97	-1	-0.66
AR-1	-0.99	-0.97	-0.99	-0.99	-0.99	-	-0.99	-0.99	-0.92
SVM	0.81	1.07	0.92	-0.92	0.97	0.99	-	-1.2	0.04
RF	1.03	1.09	1.06	1.23	1	0.99	1.2	-	0.27
LSTM	0.09	0.65	0.14	-0.07	0.66	0.92	-0.04	-0.27	-

Table 83: OMXHPI Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.15	1.31*	-1.02	1	0.99	-1.1	-1.1	-1.07
C-HAR	-1.15	-	-0.99	-1.11	-1.69	0.98	-1.13	-1.13	-1.12
WLS	-1.31	0.99	-	-1.19	0.55	0.98	-1.2	-1.2	-1.19
RR	1.02	1.11	1.19	-	1.01	1	-1.26	-1.25	-1.16
HAR-J	-1	1.69**	-0.55	-1.01	-	0.99	-1.04	-1.04	-1.03
AR-1	-0.99	-0.98	-0.98	-1	-0.99	-	-1	-1	-1
SVM	1.1	1.13	1.2	1.26	1.04	1	-	-0.1	0.42
RF	1.1	1.13	1.2	1.25	1.04	1	0.1	-	0.44
LSTM	1.07	1.12	1.19	1.16	1.03	1	-0.42	-0.44	-

Table 84: OMXHPI Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.07	1.14	0.69	-1.05	-1.01	1.17	-1.08	0.01
C-HAR	1.07	-	1.09	1.01	-1	-0.94	1.12	-1.12	0.02
WLS	-1.14	-1.09	-	-1.8	-1.06	-1.03	1.19	-1.1	0.01
RR	-0.69	-1.01	1.8**	-	-1.01	-0.98	1.33*	-1.02	0.01
HAR-J	1.05	1	1.06	1.01	-	-0.82	1.09	0.96	0.03
AR-1	1.01	0.94	1.03	0.98	0.82	-	1.06	0.91	0.03
SVM	-1.17	-1.12	-1.19	-1.33	-1.09	-1.06	-	-1.12	0
RF	1.08	1.12	1.1	1.02	-0.96	-0.91	1.12	-	0.02
LSTM	-0.01	-0.02	-0.01	-0.01	-0.03	-0.03	0	-0.02	-

Table 85: OMXSPI Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.25	-0.44	-0.89	-1.12	0.99	0.33	-1.09	0.16
C-HAR	1.25	-	1.64*	-0.03	-0.98	1	3.09**	-0.61	0.47
WLS	0.44	-1.64	-	-1.04	-1.26	0.99	0.82	-1.32	0.2
RR	0.89	0.03	1.04	-	-1.61	0.99	2.12**	-1.61	0.47
HAR-J	1.12	0.98	1.26	1.61*	-	1	1.95**	1.17	0.71
AR-1	-0.99	-1	-0.99	-0.99	-1	-	-0.95	-1	-0.88
SVM	-0.33	-3.09	-0.82	-2.12	-1.95	0.95	-	-2.49	0.11
RF	1.09	0.61	1.32*	1.61*	-1.17	1	2.49**	-	0.52
LSTM	-0.16	-0.47	-0.2	-0.47	-0.71	0.88	-0.11	-0.52	-

Table 86: OMXSPI Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.19	2.24**	0.87	-1.32	0.3	0.92	0.39	1.69**
C-HAR	0.19	-	1.94**	0.66	-0.79	0.32	0.9	0.64	1.77**
WLS	-2.24	-1.94	-	-1.89	-2.38	-1.09	-1.73	-1.84	-0.83
RR	-0.87	-0.66	1.89**	-	-1.98	0.18	0.61	-0.16	1.45*
HAR-J	1.32*	0.79	2.38**	1.98**	-	0.45	1.89**	1.54*	2.1**
AR-1	-0.3	-0.32	1.09	-0.18	-0.45	-	-0.1	-0.2	0.52
SVM	-0.92	-0.9	1.73**	-0.61	-1.89	0.1	-	-0.5	1.31*
RF	-0.39	-0.64	1.84**	0.16	-1.54	0.2	0.5	-	1.55*
LSTM	-1.69	-1.77	0.83	-1.45	-2.1	-0.52	-1.31	-1.55	-

Table 87: OMXSPI Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0.77	0.83	1.06	0.61	1.28*	1.19	0.8	0.08
C-HAR	-0.77	-	-0.73	1.27	-0.96	1.08	1.38*	-0.72	0.06
WLS	-0.83	0.73	-	1.18	0.28	1.21	1.23	0.68	0.07
RR	-1.06	-1.27	-1.18	-	-1.47	1.03	1.25	-1.26	0.05
HAR-J	-0.61	0.96	-0.28	1.47*	-	1.18	1.31*	0.05	0.07
AR-1	-1.28	-1.08	-1.21	-1.03	-1.18	-	-0.29	-1.17	-0.02
SVM	-1.19	-1.38	-1.23	-1.25	-1.31	0.29	-	-1.26	0
RF	-0.8	0.72	-0.68	1.26	-0.05	1.17	1.26	-	0.07
LSTM	-0.08	-0.06	-0.07	-0.05	-0.07	0.02	0	-0.07	-

Table 88: OSEAX Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.6	-0.94	0.4	-1.06	1.08	2.59**	1.75**	1.58*
C-HAR	0.6	-	-0.12	1.38*	-0.97	1.07	2.48**	1.55*	1.61*
WLS	0.94	0.12	-	2.61**	-0.82	1.07	2.49**	1.61*	1.63*
RR	-0.4	-1.38	-2.61	-	-2.29	1.05	2.44**	1.38*	1.55*
HAR-J	1.06	0.97	0.82	2.29**	-	1.08	2.49**	1.64*	1.65**
AR-1	-1.08	-1.07	-1.07	-1.05	-1.08	-	-0.47	-1	-0.34
SVM	-2.59	-2.48	-2.49	-2.44	-2.49	0.47	-	-2.87	-0.04
RF	-1.75	-1.55	-1.61	-1.38	-1.64	1	2.87**	-	1.2
LSTM	-1.58	-1.61	-1.63	-1.55	-1.65	0.34	0.04	-1.2	-

Table 89: OSEAX Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.05	-1.17	-0.3	-0.67	1.01	5.23**	-1.4	2.13**
C-HAR	1.05	-	0.34	1.34*	0.83	1.01	5.75**	-0.58	2.5**
WLS	1.17	-0.34	-	1.35*	0.14	1.01	5.52**	-0.84	2.45**
RR	0.3	-1.34	-1.35	-	-0.65	1	6.03**	-2.02	2.21**
HAR-J	0.67	-0.83	-0.14	0.65	-	1	5.46**	-0.86	2.4**
AR-1	-1.01	-1.01	-1.01	-1	-1	-	-0.9	-1.01	-0.91
SVM	-5.23	-5.75	-5.52	-6.03	-5.46	0.9	-	-5.83	-0.35
RF	1.4*	0.58	0.84	2.02**	0.86	1.01	5.83**	-	2.62**
LSTM	-2.13	-2.5	-2.45	-2.21	-2.4	0.91	0.35	-2.62	-

Table 90: OSEAX Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.97	1.6*	2.44**	-1.8	2.1**	3.46**	1.34*	2.28**
C-HAR	0.97	-	1.46*	2.35**	0.12	2.16**	3.44**	1.35*	2.31**
WLS	-1.6	-1.46	-	2.52**	-1.83	1.96**	3.51**	1.29*	2.21**
RR	-2.44	-2.35	-2.52	-	-2.58	1.04	3.5**	0.85	2.04**
HAR-J	1.8**	-0.12	1.83**	2.58**	-	2.27**	3.46**	1.38*	2.3**
AR-1	-2.1	-2.16	-1.96	-1.04	-2.27	-	3.43**	0.57	1.94**
SVM	-3.46	-3.44	-3.51	-3.5	-3.46	-3.43	-	-3.67	-0.38
RF	-1.34	-1.35	-1.29	-0.85	-1.38	-0.57	3.67**	-	1.66**
LSTM	-2.28	-2.31	-2.21	-2.04	-2.3	-1.94	0.38	-1.66	-

Table 91: RUT Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.02	-1.97	1.93**	-1.61	0.61	3.7**	3.25**	3.25**
C-HAR	1.02	-	-0.33	1.63*	0.51	0.92	3.69**	3.29**	3.27**
WLS	1.97**	0.33	-	2.19**	1.34*	1.41*	3.69**	3.23**	3.27**
RR	-1.93	-1.63	-2.19	-	-1.94	-0.21	3.71**	3.2**	3.21**
HAR-J	1.61*	-0.51	-1.34	1.94**	-	0.91	3.69**	3.26**	3.26**
AR-1	-0.61	-0.92	-1.41	0.21	-0.91	-	3.62**	2.82**	3.22**
SVM	-3.7	-3.69	-3.69	-3.71	-3.69	-3.62	-	-3.52	-0.21
RF	-3.25	-3.29	-3.23	-3.2	-3.26	-2.82	3.52**	-	2.63**
LSTM	-3.25	-3.27	-3.27	-3.21	-3.26	-3.22	0.21	-2.63	-

Table 92: RUT Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.65**	1.98**	-0.12	1.31*	0.49	6.45**	-1.4	2.87**
C-HAR	-1.65	-	2.01**	-0.73	-1.66	-0.2	5.9**	-1.6	2.72**
WLS	-1.98	-2.01	-	-1.55	-1.99	-1.49	3.59**	-1.89	1.8**
RR	0.12	0.73	1.55*	-	0.28	0.55	7.24**	-1.89	2.88**
HAR-J	-1.31	1.66**	1.99**	-0.28	-	0.34	6.35**	-1.44	2.84**
AR-1	-0.49	0.2	1.49*	-0.55	-0.34	-	6.28**	-1.47	2.8**
SVM	-6.45	-5.9	-3.59	-7.24	-6.35	-6.28	-	-7.03	-0.71
RF	1.4*	1.6*	1.89**	1.89**	1.44*	1.47*	7.03**	-	3.25**
LSTM	-2.87	-2.72	-1.8	-2.88	-2.84	-2.8	0.71	-3.25	-

Table 93: RUT Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.91	0.55	0.74	-1.91	0.52	1.85**	0.02	0.18
C-HAR	1.91**	-	2.02**	1.71**	1.52*	1.68**	2.07**	1.42*	0.29
WLS	-0.55	-2.02	-	0.23	-2	0.36	1.59*	-0.21	0.16
RR	-0.74	-1.71	-0.23	-	-1.72	0.1	2.11**	-0.9	0.15
HAR-J	1.91**	-1.52	2**	1.72**	-	1.56*	2.09**	1.39*	0.27
AR-1	-0.52	-1.68	-0.36	-0.1	-1.56	-	1.16	-0.32	0.15
SVM	-1.85	-2.07	-1.59	-2.11	-2.09	-1.16	-	-2.45	-0.03
RF	-0.02	-1.42	0.21	0.9	-1.39	0.32	2.45**	-	0.17
LSTM	-0.18	-0.29	-0.16	-0.15	-0.27	-0.15	0.03	-0.17	-

Table 94: SMSI Diebold-Mariano Test statistics, k = 1 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.21	1.23	0.45	-1.71	1.26	3.66**	-0.97	2.51**
C-HAR	1.21	-	1.38*	1.45*	-0.72	1.3*	3.76**	0.36	2.78**
WLS	-1.23	-1.38	-	-1.15	-1.54	1.17	2.94**	-1.27	2.05**
RR	-0.45	-1.45	1.15	-	-2.12	1.21	3.6**	-1.19	2.49**
HAR-J	1.71**	0.72	1.54*	2.12**	-	1.34*	3.83**	0.95	2.85**
AR-1	-1.26	-1.3	-1.17	-1.21	-1.34	-	1.26	-1.24	0.87
SVM	-3.66	-3.76	-2.94	-3.6	-3.83	-1.26	-	-3.94	-0.25
RF	0.97	-0.36	1.27	1.19	-0.95	1.24	3.94**	-	2.74**
LSTM	-2.51	-2.78	-2.05	-2.49	-2.85	-0.87	0.25	-2.74	-

Table 95: SMSI Diebold-Mariano Test statistics, k = 5 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.69	1.71**	2.23**	-1.48	1.25	3.9**	-2.77	1.71**
C-HAR	0.69	-	1.58*	1.47*	-1.03	1.22	4.53**	-3.06	2.08**
WLS	-1.71	-1.58	-	-1.58	-1.71	0.93	-0.44	-1.96	-0.37
RR	-2.23	-1.47	1.58*	-	-1.98	1.17	1.81**	-2.71	1.03
HAR-J	1.48*	1.03	1.71**	1.98**	-	1.27	5.76**	-4.14	2.26**
AR-1	-1.25	-1.22	-0.93	-1.17	-1.27	-	-0.76	-1.37	-0.74
SVM	-3.9	-4.53	0.44	-1.81	-5.76	0.76	-	-6.17	0.03
RF	2.77**	3.06**	1.96**	2.71**	4.14**	1.37*	6.17**	-	2.79**
LSTM	-1.71	-2.08	0.37	-1.03	-2.26	0.74	-0.03	-2.79	-

Table 96: SMSI Diebold-Mariano Test statistics, k = 22 * = p < 0.1, ** = p < 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.89	0.06	1.62*	-0.49	2.1**	3.34**	1.4*	2.07**
C-HAR	0.89	-	0.8	1.8**	0.7	2.17**	3.3**	1.38*	2.13**
WLS	-0.06	-0.8	-	2.23**	-0.37	2.37**	3.37**	1.19	2.06**
RR	-1.62	-1.8	-2.23	-	-1.57	2.42**	3.47**	0.21	1.79**
HAR-J	0.49	-0.7	0.37	1.57*	-	2.05**	3.27**	1.32*	2.11**
AR-1	-2.1	-2.17	-2.37	-2.42	-2.05	-	3.54**	-1.15	1.37*
SVM	-3.34	-3.3	-3.37	-3.47	-3.27	-3.54	-	-3.5	-0.49
RF	-1.4	-1.38	-1.19	-0.21	-1.32	1.15	3.5**	-	1.73**
LSTM	-2.07	-2.13	-2.06	-1.79	-2.11	-1.37	0.49	-1.73	-

Table 97: SPXCASH Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.02	-1.1	0.93	0.85	2.54**	5.36**	3.07**	5.03**
C-HAR	1.02	-	0.2	1.25	1.63*	2.6**	5.37**	3.23**	5.07**
WLS	1.1	-0.2	-	1.52*	1.4*	2.78**	5.36**	3.12**	5.05**
RR	-0.93	-1.25	-1.52	-	-0.51	2.94**	5.38**	2.89**	4.99**
HAR-J	-0.85	-1.63	-1.4	0.51	-	2.36**	5.34**	2.97**	5.01**
AR-1	-2.54	-2.6	-2.78	-2.94	-2.36	-	5.28**	1.19	4.73**
SVM	-5.36	-5.37	-5.36	-5.38	-5.34	-5.28	-	-5.47	-0.98
RF	-3.07	-3.23	-3.12	-2.89	-2.97	-1.19	5.47**	-	4.46**
LSTM	-5.03	-5.07	-5.05	-4.99	-5.01	-4.73	0.98	-4.46	-

Table 98: SPXCASH Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	2.81**	0.74	-1.12	0.36	1.96**	6.46**	1.68**	5.88**
C-HAR	-2.81	-	-1.19	-2.37	-2.11	1.35*	6.34**	1.15	5.69**
WLS	-0.74	1.19	-	-1.16	-0.46	1.51*	6.23**	1.3*	5.71**
RR	1.12	2.37**	1.16	-	1.53*	2.99**	6.85**	2.66**	6.07**
HAR-J	-0.36	2.11**	0.46	-1.53	-	1.9**	6.54**	1.73**	5.88**
AR-1	-1.96	-1.35	-1.51	-2.99	-1.9	-	6.56**	-0.04	5.33**
SVM	-6.46	-6.34	-6.23	-6.85	-6.54	-6.56	-	-6.51	-0.47
RF	-1.68	-1.15	-1.3	-2.66	-1.73	0.04	6.51**	-	5.33**
LSTM	-5.88	-5.69	-5.71	-6.07	-5.88	-5.33	0.47	-5.33	-

Table 99: SPXCASH Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-1.01	-0.16	1.87**	-1.04	2.23**	2.96**	1.05	1.78**
C-HAR	1.01	-	1.13	1.73**	0.64	2.23**	2.86**	1.05	1.87**
WLS	0.16	-1.13	-	2.07**	-0.91	2.46**	2.94**	0.97	1.79**
RR	-1.87	-1.73	-2.07	-	-1.72	2.4**	3.01**	0.41	1.61*
HAR-J	1.04	-0.64	0.91	1.72**	-	2.17**	2.88**	1.06	1.85**
AR-1	-2.23	-2.23	-2.46	-2.4	-2.17	-	2.91**	-0.52	1.39*
SVM	-2.96	-2.86	-2.94	-3.01	-2.88	-2.91	-	-3.35	-0.43
RF	-1.05	-1.05	-0.97	-0.41	-1.06	0.52	3.35**	-	1.44*
LSTM	-1.78	-1.87	-1.79	-1.61	-1.85	-1.39	0.43	-1.44	-

Table 100: SPXFUT Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.35	-1.48	0.36	0.45	2.19**	5.13**	2.64**	4.99**
C-HAR	0.35	-	-0.66	0.43	0.96	2**	5.13**	2.65**	5.01**
WLS	1.48*	0.66	-	1.03	1.46*	2.34**	5.11**	2.67**	5.03**
RR	-0.36	-0.43	-1.03	-	0.07	2.66**	5.14**	2.56**	4.98**
HAR-J	-0.45	-0.96	-1.46	-0.07	-	1.81**	5.11**	2.52**	4.98**
AR-1	-2.19	-2	-2.34	-2.66	-1.81	-	5.05**	1.59*	4.79**
SVM	-5.13	-5.13	-5.11	-5.14	-5.11	-5.05	-	-5.25	-0.56
RF	-2.64	-2.65	-2.67	-2.56	-2.52	-1.59	5.25**	-	4.39**
LSTM	-4.99	-5.01	-5.03	-4.98	-4.98	-4.79	0.56	-4.39	-

Table 101: SPXFUT Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	3.3**	-1.08	-0.31	-0.52	2.04**	6.53**	2.58**	4.95**
C-HAR	-3.3	-	-2.94	-2.25	-2.95	1	6.37**	1.52*	4.71**
WLS	1.08	2.94**	-	0.43	0.87	2.01**	6.37**	2.45**	5.02**
RR	0.31	2.25**	-0.43	-	0.23	2.31**	6.74**	3.04**	4.97**
HAR-J	0.52	2.95**	-0.87	-0.23	-	2.06**	6.55**	2.67**	4.96**
AR-1	-2.04	-1	-2.01	-2.31	-2.06	-	6.59**	0.65	4.52**
SVM	-6.53	-6.37	-6.37	-6.74	-6.55	-6.59	-	-6.98	-0.73
RF	-2.58	-1.52	-2.45	-3.04	-2.67	-0.65	6.98**	-	4.34**
LSTM	-4.95	-4.71	-5.02	-4.97	-4.96	-4.52	0.73	-4.34	-

Table 102: SPXFUT Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.1	-1.9	2.88**	-0.7	1.52*	2.98**	1.58*	1.95**
C-HAR	0.1	-	-0.43	2.08**	-0.37	1.42*	2.93**	1.44*	1.95**
WLS	1.9**	0.43	-	2.74**	0.46	1.69**	2.96**	1.62*	1.99**
RR	-2.88	-2.08	-2.74	-	-2.54	-0.71	2.97**	0.93	1.66**
HAR-J	0.7	0.37	-0.46	2.54**	-	1.64*	2.95**	1.54*	1.98**
AR-1	-1.52	-1.42	-1.69	0.71	-1.64	-	2.85**	0.9	1.73**
SVM	-2.98	-2.93	-2.96	-2.97	-2.95	-2.85	-	-3.31	-0.61
RF	-1.58	-1.44	-1.62	-0.93	-1.54	-0.9	3.31**	-	1.32*
LSTM	-1.95	-1.95	-1.99	-1.66	-1.98	-1.73	0.61	-1.32	-

Table 103: SSEC Diebold-Mariano Test statistics, k = 1 * = $p < 0.1$, ** = $p < 0.05$

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.26	-1.78	2.67**	1.03	2.26**	3.5**	2.36**	2.59**
C-HAR	-1.26	-	-1.67	-0.63	-1.17	0.32	3.48**	2.25**	2.44**
WLS	1.78**	1.67**	-	2.42**	1.65**	2.38**	3.49**	2.4**	2.63**
RR	-2.67	0.63	-2.42	-	-0.15	1.47*	3.49**	2.27**	2.51**
HAR-J	-1.03	1.17	-1.65	0.15	-	1.14	3.48**	2.31**	2.52**
AR-1	-2.26	-0.32	-2.38	-1.47	-1.14	-	3.36**	1.73**	2.39**
SVM	-3.5	-3.48	-3.49	-3.49	-3.48	-3.36	-	-3.81	-1.08
RF	-2.36	-2.25	-2.4	-2.27	-2.31	-1.73	3.81**	-	1.62*
LSTM	-2.59	-2.44	-2.63	-2.51	-2.52	-2.39	1.08	-1.62	-

Table 104: SSEC Diebold-Mariano Test statistics, k = 5 * = $p < 0.1$, ** = $p < 0.05$

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.58	0.77	3.88**	1.3*	4.87**	6.83**	0.07	2.67**
C-HAR	0.58	-	1.63*	2.59**	1.68**	4.83**	6.66**	0.39	2.76**
WLS	-0.77	-1.63	-	1.8**	-0.3	4.5**	6.55**	-0.41	2.6**
RR	-3.88	-2.59	-1.8	-	-2.98	4.66**	6.89**	-2.52	2.17**
HAR-J	-1.3	-1.68	0.3	2.98**	-	4.9**	6.78**	-0.38	2.6**
AR-1	-4.87	-4.83	-4.5	-4.66	-4.9	-	5.99**	-4.43	0.59
SVM	-6.83	-6.66	-6.55	-6.89	-6.78	-5.99	-	-7.15	-2.79
RF	-0.07	-0.39	0.41	2.52**	0.38	4.43**	7.15**	-	2.66**
LSTM	-2.67	-2.76	-2.6	-2.17	-2.6	-0.59	2.79**	-2.66	-

Table 105: SSEC Diebold-Mariano Test statistics, k = 22 * = $p < 0.1$, ** = $p < 0.05$

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.58*	0.46	1.79**	-0.23	1.44*	2.57**	2.01**	0.64
C-HAR	-1.58	-	-1.67	-1.46	-1.1	-0.44	2.79**	-0.27	0.45
WLS	-0.46	1.67**	-	1.64*	-0.27	1.22	2.61**	2.09**	0.63
RR	-1.79	1.46*	-1.64	-	-0.66	0.87	2.58**	1.83**	0.59
HAR-J	0.23	1.1	0.27	0.66	-	1.93**	2.28**	1.31*	0.67
AR-1	-1.44	0.44	-1.22	-0.87	-1.93	-	2.12**	0.4	0.54
SVM	-2.57	-2.79	-2.61	-2.58	-2.28	-2.12	-	-2.71	-0.1
RF	-2.01	0.27	-2.09	-1.83	-1.31	-0.4	2.71**	-	0.48
LSTM	-0.64	-0.45	-0.63	-0.59	-0.67	-0.54	0.1	-0.48	-

Table 106: SSMI Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.94**	-1.18	-1.55	1.37*	1.58*	3.03**	2.26**	2.67**
C-HAR	-1.94	-	-1.96	-2.02	-1.85	-0.71	3.18**	1.93**	2.1**
WLS	1.18	1.96**	-	0.61	1.61*	1.57*	3.04**	2.31**	2.72**
RR	1.55*	2.02**	-0.61	-	1.51*	1.66**	3.05**	2.31**	2.69**
HAR-J	-1.37	1.85**	-1.61	-1.51	-	1.49*	3.01**	2.2**	2.65**
AR-1	-1.58	0.71	-1.57	-1.66	-1.49	-	2.81**	1.4*	2.29**
SVM	-3.03	-3.18	-3.04	-3.05	-3.01	-2.81	-	-3.42	-0.19
RF	-2.26	-1.93	-2.31	-2.31	-2.2	-1.4	3.42**	-	1.69**
LSTM	-2.67	-2.1	-2.72	-2.69	-2.65	-2.29	0.19	-1.69	-

Table 107: SSMI Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.27	1.74**	-0.73	1.02	0.57	3.66**	-0.97	1.44*
C-HAR	-1.27	-	0.3	-1.97	-0.56	0.09	4.63**	-3.19	1.18
WLS	-1.74	-0.3	-	-1.49	-1.8	-0.1	2.39**	-1.58	1.12
RR	0.73	1.97**	1.49*	-	0.93	0.68	3.97**	-0.79	1.47*
HAR-J	-1.02	0.56	1.8**	-0.93	-	0.38	3.17**	-1.07	1.36*
AR-1	-0.57	-0.09	0.1	-0.68	-0.38	-	2.12**	-0.85	1.09
SVM	-3.66	-4.63	-2.39	-3.97	-3.17	-2.12	-	-4.8	-0.07
RF	0.97	3.19**	1.58*	0.79	1.07	0.85	4.8**	-	1.59*
LSTM	-1.44	-1.18	-1.12	-1.47	-1.36	-1.09	0.07	-1.59	-

Table 108: SSMI Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	-0.3	0.38	0.42	-0.96	1.62*	1.4*	0.7	0.29
C-HAR	0.3	-	0.37	1.61*	-0.51	1.47*	1.51*	1.16	0.3
WLS	-0.38	-0.37	-	-0.29	-0.44	0.27	0.74	-0.12	0.22
RR	-0.42	-1.61	0.29	-	-1.16	1.42*	1.47*	0.84	0.28
HAR-J	0.96	0.51	0.44	1.16	-	1.55*	1.46*	1	0.31
AR-1	-1.62	-1.47	-0.27	-1.42	-1.55	-	1.09	-0.95	0.15
SVM	-1.4	-1.51	-0.74	-1.47	-1.46	-1.09	-	-1.58	-0.1
RF	-0.7	-1.16	0.12	-0.84	-1	0.95	1.58*	-	0.24
LSTM	-0.29	-0.3	-0.22	-0.28	-0.31	-0.15	0.1	-0.24	-

Table 109: STOXX50E Diebold-Mariano Test statistics, k = 1 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	1.71**	0.1	0.85	0.57	1.17	3.63**	2.22**	2.66**
C-HAR	-1.71	-	-1.29	-1.49	-2.22	0.51	3.55**	1.6*	2.43**
WLS	-0.1	1.29*	-	0.19	0.18	1.15	3.54**	2.08**	2.66**
RR	-0.85	1.49*	-0.19	-	-0.1	1.03	3.64**	2.16**	2.61**
HAR-J	-0.57	2.22**	-0.18	0.1	-	1.02	3.59**	2.09**	2.62**
AR-1	-1.17	-0.51	-1.15	-1.03	-1.02	-	3.58**	0.69	2.18**
SVM	-3.63	-3.55	-3.54	-3.64	-3.59	-3.58	-	-3.97	-0.35
RF	-2.22	-1.6	-2.08	-2.16	-2.09	-0.69	3.97**	-	1.99**
LSTM	-2.66	-2.43	-2.66	-2.61	-2.62	-2.18	0.35	-1.99	-

Table 110: STOXX50E Diebold-Mariano Test statistics, k = 5 * = p $<$ 0.1, ** = p $<$ 0.05

	HAR	C-HAR	WLS	RR	HAR-J	AR-1	SVM	RF	LSTM
HAR	-	0.75	2.71**	2.74**	0.37	1.1	4.71**	-2.41	2.56**
C-HAR	-0.75	-	2.68**	0.85	-0.69	1.03	4.63**	-2.46	2.47**
WLS	-2.71	-2.68	-	-2.61	-2.71	-1.38	-1.48	-2.94	-1.16
RR	-2.74	-0.85	2.61**	-	-1.02	1	4.65**	-2.76	2.37**
HAR-J	-0.37	0.69	2.71**	1.02	-	1.06	4.48**	-2.17	2.51**
AR-1	-1.1	-1.03	1.38*	-1	-1.06	-	0.39	-1.36	0.48
SVM	-4.71	-4.63	1.48*	-4.65	-4.48	-0.39	-	-5.14	0.26
RF	2.41**	2.46**	2.94**	2.76**	2.17**	1.36*	5.14**	-	3.12**
LSTM	-2.56	-2.47	1.16	-2.37	-2.51	-0.48	-0.26	-3.12	-

Table 111: STOXX50E Diebold-Mariano Test statistics, k = 22 * = p $<$ 0.1, ** = p $<$ 0.05