Erasmus University Rotterdam

Erasmus School of Economics

MSc Econometrics and Management Science

Quantitative Finance

Constructing pooled forecasts based on leading recession indicators to predict multiple-day variance

Date final version version: July 4, 2022

Author	*Supervisor*
Y.M. Corper	O. Kleen
Student number	*Second assessor*
502767	A. Tetereva

Abstract

This paper describes the analysis of the predictive power of both the long term Treasury yield spread and the NFCI regarding the construction of a pooled model that forecasts multiple-day variance. The individual models that are used in this pooled model are a HAR model, two CAViaR models and two GARCHX models. The used data contains stock data of eight U.S. stocks and data of both the long term Treasury yield spread and the NFCI. Based on two performance measures (QLIKE and MSE) and multiple Diebold-Mariano tests, it is found that the pooled model, which is based on both indicators, is a very strong competitor for the used benchmark models, especially for tech- and financial stocks. These findings open the door for further improvement of the constructed pooled model, as this may lead to forecasts that are significantly more accurate than those of the benchmark models.

Contents

1 Introduction

Forecasting variance of stocks is a subject which gets a lot of attention in the literature. Variance of stocks is changing all the time and this change can be due to a lot of factors. One of these factors may be the National Financial Conditions Index (NFCI). The higher the value of this index, the worse the financial conditions in the U.S. are. Another factor that may impact variance is the long term Treasury yield spread, which is the ten-year minus three-month Treasury yield. This yield spread is a very prominent stress indicator in the U.S. and it can be linked to the probability of a recession starting in about a year. The lower the value of this yield spread, the higher the probability of a recession in about a year is. As these two indicators give information about the state of the economy in the U.S. economy, their values may have an impact on the variance of stocks. This impact may also be different among different stocks. Namely, stocks can be divided into different types, such as growth stocks, value stocks, tech stocks, financial stocks and many more. All these type of stocks may be affected differently by the yield spread and the NFCI in terms of their variance. This paper will focus on linking the values of these two leading indicators to variance forecasts, so that future predictions of variance can be based on current values of these two indicators.

In this paper the variance of different types of stocks will be forecasted using five different models. This includes two CAViaR type of models, a HAR model and two GARCHX models. Each of these models is suitable for the purpose of this research. The HAR model was first introduced by Corsi (2009). This model is a simple variance model with an additive structure. It makes predictions based on lagged values of the dependent variable. This model is a GARCH type of model. A potential problem of this type of models is that if the shape of the distribution changes over time, these models can produce variance forecasts that are very inefficient due to wrong distribution assumptions. A class of models which solves this problem are the CAViaR type of models. These models were first introduced by Engle and Manganelli (2004). This class of models does not require any assumptions of the underlying distribution, but models the quantiles directly in an autoregressive way. A least squares regression of variance on the calculated quantiles will be performed in order to forecast variance, as is also done in Taylor (2005). The final model which will be used in this research is the GARCHX model, which was introduced by Hwang and Satchell (2005). The GARCHX model is an extended version of the GARCH model, which is able to incorporate exogenous variables. This characteristic makes the model very useful for the purpose of this research, as will be further explained in Section 4.

The forecasts produced by each of these models will be weighted to get as close as possible to the true value of variance, resulting in a pooled forecast. In the paper by Kourentzes et al. (2019) it is shown that pooled forecasts are beneficial in terms of forecast accuracy compared to single model forecasting. By doing this weighting of the models a lot of times, a time series of the weights of each model can be constructed. By use of regressions, a possible relation between the weights of the different type of models and the leading indicators may be found. In this way also relations between the weights of models and weights of the same model belonging to other stocks may be established. Furthermore, Chinn and Kucko (2015) show that the yield curve has predictive power as predictor of future economic activity. Their finding may suggest that the yield spread also has predictive power regarding pooled variance forecasts. In a paper by Opschoor et al. (2014) a direct link between a U.S. Financial Conditions Index (FCI) and both volatilities and correlations is found. The findings of this paper may suggest that the NFCI has predictive power regarding pooled variance forecasts. A benchmark approach of weighting the different models is looking at the historical mean of each weight. This will be called a mean weighted pooled model. The main research question is as follows:

"Can the level of both the long term Treasury yield spread and the NFCI contribute to constructing pooled forecasts that are significantly more accurate than both equally weighted- and mean weighted pooled forecasts regarding multiple-day variance?"

This main research question will be answered with the help of the following subquestions:

1. Does constructing pooled forecasts based on the inverse forecast error of individual models lead to forecasts that are significantly more accurate than the forecasts of the individual models?
2. How can a relation between weights of models and the leading indicators be found?
3. What are other explanatory variables that can make the forecasts of the pooled model even more accurate?

This research is relevant in a couple of ways. First of all, this research is academically relevant as there is no literature yet which describes a direct link between weights of a pooled forecast and leading indicators like the NFCI and the long term Treasury yield spread. Another point which highlights the relevance of this research is the fact that accurately forecasting variance is very important, for example for investors, firms and even policy makers. This research contributes to possible improvements of variance forecasting.

The data that is used in this research consists of log returns of 8 different U.S. stocks. These log returns have been created by using daily stock prices. During this research, the demeaned log returns will be used. The stocks that are considered are Apple, Amazon, Microsoft, Ebay, Goldman Sachs, JP Morgan, Bank of America and Netflix. These stocks have been chosen as they can be grouped into different types of stocks. Apple, Ebay and Microsoft are viewed as tech stocks. Goldman Sachs, JP Morgan and Bank of America are considered as financial stocks. Amazon and Netflix are considered as growth stocks.

Based on multiple Diebold-Mariano tests, it is found that the pooled forecasts, that are constructed based on the values of the two leading indicators, are on average not significantly more accurate than the forecasts of the benchmark models. On the other hand, the constructed pooled forecasts are also not significantly less accurate than the forecasts of the benchmark models. These results indicate that the two used leading indicators have predictive power regarding pooled variance forecasts. It is also found that adding lagged weights of other stocks and the lagged actual variance helps to create pooled forecasts that are a little bit more accurate, but this difference is not significant.

The remainder of this paper is as follows. Section 2 will be used to discuss the literature of all the individual models, the pooled forecasts and the leading indicators. Section 3 will be used to explain which stock data is used, the characteristics of the stock data will be mentioned and the data of the long term Treasury yield spread and the NFCI will be shown. Section 4 will be used to explain the methods and techniques used for all the individual models, how these individual forecasts are pooled into one forecast and how the values of the leading indicators and other explanatory variables are incorporated into the forecasts. The results following these methods and techniques are provided in Section 5. Finally, Section 6 discusses the main findings, provides recommendations and concludes the paper.

2 Literature

There have been quite a few implementations for forecasting variance in the existing literature. This section summarizes the most relevant and important aspects of the scientific literature related to HAR models, CAViaR models and GARCHX models. This section also highlights literature regarding pooled forecasts and the predictive power of both yield spreads and the NFCI.

Simplistic Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) type of models suffer from not having any long term memory. This is due to the exponential decline in the autocorrelations. As described in Corsi (2009), this problem may be solved by adopting fractional difference operators leading to FIGARCH or ARFIMA models. However, these multiplicative models are known to be extremely difficult to identify and estimate. Corsi (2009) provided a solution to this issue with the introduction of the HAR model. This model is a simple multi-component variance model with an additive structure. Corsi (2009) identifies three primary variance components from his assumptions: the short-term variance with daily or higher dealing frequency, the medium-term variance, typically made of portfolio managers who rebalance their positions weekly, and the long-term variance with a characteristic time of one or more months.

As just mentioned, GARCH type of models are a well established class of models for forecasting variance. This kind of models rely on assumptions of the shape of the conditional distribution, for example they assume a normal distribution. A potential problem for GARCH type of models is that if the shape of the distribution varies over time, this type of models will produce inefficient variance forecasts, because of wrong distribution assumptions. To overcome this problem, a CAViaR type of model is also used in this paper. CAViaR type of models were first introduced by Engle and Manganelli (2004). This kind of models do not need any assumption of an underlying distribution, but model the quantiles directly in an autoregressive nature. CAViaR type of models were designed to model and forecast Value at Risk. However, Pearson and Tukey (1965) discovered a constant relation between the ratio of standard deviation and symmetric quantiles. Taylor (2005) uses this finding to construct a least squares regression of variance on the symmetric quantiles to model the variance. The use of quantile regressions to estimate nonlinear quantile methods is supported by White (1994). Taylor (2005) was the first to use CAViaR type of models for predicting variance. He shows that the Asymmetric Slope CAViaR model performs best in comparison to several GARCH type of models and CAViaR type of models. Therefore this type of CAViaR model will be used during this research, both with a 90% symmetric quantile interval and a 95% symmetric quantile interval as these perform best in his research.

The final model that is considered in this research is the GARCHX model. This model is similar to a GARCH model, but incorporates exogenous variables like the cross-sectional market variance. The model was first introduced by Hwang and Satchell (2005). They conclude in their paper that daily return variance can be better specified with a GARCHX model, compared to regular GARCH models. However, they mention that the GARCHX model does not necessarily outperform more conventional GARCH models in terms of forecasting. In a paper by Emenike Kalu and Okwuchukwu (2014) the impact of macroeconomic variables on stock market return variance in Nigeria using a GARCHX model is investigated. They conclude that the macroeconomic variables have a significant influence on return variance. This finding supports the idea of possible predictive power of leading indicators regarding variance. In a paper by Yeasin et al. (2020) a comparison between a regular GARCH model and a GARCHX model is made. From this comparison supremacy of using exogenous factors in variance modelling is concluded. Therefore the GARCHX model will be used during this research, instead of a regular GARCH model.

The forecasts produced by the five different models will be used to construct pooled forecasts. Kourentzes et al. (2019) show that pooled forecasts are beneficial in terms of forecast accuracy in contrast to choosing a single model for forecasting. They show in their paper that by combining forecasts, forecasts errors can be reduced. Also modelling uncertainty is reduced as not a single model has to be chosen. Newbold and Harvey (2002) also argue that pooling forecasts may be very beneficial. According to a paper by Duncan et al. (2001), simple pooling methods are more accurate than complex pooling methods. In their paper it is also shown that pooled forecasts are more accurate than forecasts of individual models. The performance of the pooled forecasts will be compared to two benchmark models, of which one is an equally weighted pooled forecast. This equally weighted model is used as a benchmark as 1/n portfolios are tough competitors, as shown in DeMiguel et al. (2009).

This research is about whether the long term Treasury yield spread and the NFCI can contribute to constructing pooled forecasts that are significantly more accurate than forecasts produced by both a mean weighted pooled forecast and an equally weighted pooled forecast. The long term Treasury yield spread can only contribute to this if it has predictive power. In the literature there are numerous papers which highlight the predictive power of leading indicators in general and of the yield spread specifically. Clements and Galvão (2009) show that leading indicators have significant predictive ability for output growth at horizons up to one year. Furthermore, Chinn and Kucko (2015) show that the yield curve has predictive power as predictor of future economic activity, even after accounting for other leading indicators. In a paper by Hännikäinen (2017) it is found that the slope of the yield curve has more predictive power when inflation is highly persistent. This may also be a useful finding for this research. Finally, a paper by Christiansen (2013) shows that U.S. yield spreads act as leading indicator for severe simultaneous recessions. This finding again highlights the potential of predictive power of the long term Treasury yield spread regarding variance.

Next to the long term Treasury yield spread, the predictive power of the NFCI regarding variance will be investigated. In a paper by Amburgey and McCracken (2022) it is shown that there is predictive content of the NFCI for conditional quantiles of U.S. real GDP growth. In a paper by Opschoor et al. (2014) a direct link between an U.S. Financial Conditions Index (FCI) and both volatilities and correlations is found. The findings of these papers may suggest that the NFCI has predictive power regarding variance. The findings may also suggest that the NFCI has predictive power regarding the weights given to the models that are contained in a pooled forecast.

3 Data

The data used in this paper consists of 8 different U.S. stocks: Apple (AAPL), Amazon (AMZN), Microsoft (MSFT), Ebay (EBAY), Goldman Sachs (GS), JP Morgan (JPM), Bank of America (BAC) and Netflix (NFLX). These specific stocks are considered as they can be divided into three different types of stocks, based on the used sample period. Apple, Ebay and Microsoft are viewed as tech stocks. Goldman Sachs, JP Morgan and Bank of America are considered financial stocks. Finally, Amazon and Netflix are considered growth stocks. For all of the stocks the estimation sample consists of 10 years of daily data, from 01/01/2000 until 31/12/2009. For the Netflix stock there is only data available from 23/05/2002 onwards. Therefore there are 1917 observations, instead of 2515 for the other stocks, available for the estimation of the models for Netflix.

For each of the stocks daily stock prices have been collected via the database of WRDS. These daily stock prices can be transformed into daily log returns as follows: $r_t^* = ln(\frac{P_t}{P_{t-1}})$, where P_t is the stock price at time t. For all of the stocks, the mean of these log returns has been subtracted. This demeaning is done as this is beneficial for estimating model parameters that do not involve a mean, as otherwise a bias would occur. The variance forecasting methods are applied to the resultant error terms: $r_t = r_t^*$ - μ_t. Where μ_t is the mean of the log returns of the past 10 years at time t. From now on these demeaned log returns r_t will be referred to as the log returns.

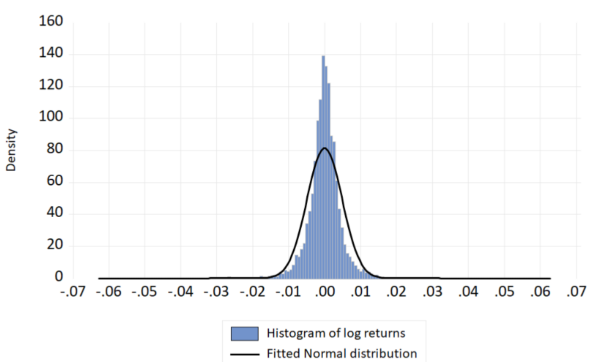
3.1 Descriptive statistics of individual stocks

Table 1 shows descriptive statistics for all of the stocks. From this table it can be concluded that for all stocks it holds that the skewness is different from zero. Furthermore, it can be seen that for all stocks the log returns exhibit excess kurtosis. The two observations of skewness being different from zero and excess kurtosis indicate that the log returns of the stocks are likely to be non-normally distributed. This is confirmed by the Jarque-Bera test that is performed. The p-values that are shown confirm that none of the individual stocks is normally distributed, as they are smaller than 0.05. This finding is also confirmed by the two subfigures shown in Figure 1, where the distribution of the log returns of Goldman Sachs and Amazon are shown against a fitted normal distribution. The two subfigures show that the log returns of both Goldman Sachs and Amazon are non-normally distributed. The descriptive statistics show that the normal distribution is likely to fail in capturing the underlying statistical properties of the time series, as the statistics indicate that the unconditional distribution of log returns deviates from standard Gaussian assumptions. Therefore, only models that can account for data exhibiting excess kurtosis and skewness are considered in this research.

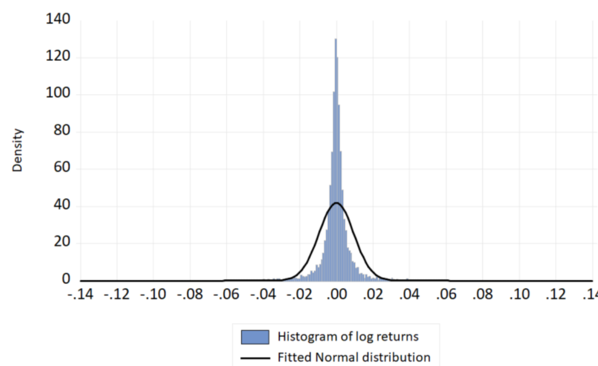
Table 1: Descriptive statistics of the demeaned log returns of all stocks from 04/01/2000 until 31/12/2021.

	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Mean	1.67E-04	-2.13E-05	-5.97E-06	2.28E-05	7.36E-06	2.09E-05	4.52E-05	4.90E-06
Median	-5.73E-04	-7.39E-05	-3.28E-04	-5.18E-05	1.63E-05	-5.02E-05	-5.18E-05	-6.90E-04
Standard deviation	0.0290	0.0095	0.0169	0.0085	0.0049	0.0079	0.0141	0.0260
Skewness	42.88	1.02	27.15	37.77	0.30	10.67	9.31	47.68
Kurtosis	2098.99	37.59	931.37	2244.06	18.69	419.38	390.36	2848.06
P-value JB test	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	5535	5535	5535	5535	5535	5535	5535	4937

Notes: P-values of the Jarque-Bera test statistics are given. A p-value smaller than 0.05 indicates that the corresponding log returns are not normally distributed.



(a) Distribution of log returns GS.



(b) Distribution of log returns AMZN.

Figure 1: Distribution of log returns of Goldman Sachs against a fitted Normal distribution (left) and distribution of log returns of Amazon against a fitted Normal distribution (right).

Table 2 shows the correlation matrix for the individual stocks. The correlations are calculated based on the log returns ranging from 23/05/2002 - 31/12/2021. What can be seen from this table is that the highest cross-correlation is equal to 0.48, which is not a very high correlation. It can also be seen that all stocks have a relatively low correlation with NFLX. The stocks GS and BAC have a relatively high cross-correlation. This can be explained as these stocks can be seen as financial stocks during the period 23/05/2002 - 31/12/2021. Finally, it can be seen that the correlation between BAC and JPM and the correlation between GS and JPM are very small. This is surprising as these are the same type of stocks and it was expected that these stocks would be more correlated. However, this low correlation between JPM and both GS and BAC may be due to the fact that when looking at the historical data, JPM has recovered much faster to the peak pre-crisis levels after the financial crisis, while the other two stocks have been hovering around lower levels for a longer time.

Table 2: Correlation matrix of the individual stocks based on the log returns ranging from 23/05/2002 until 31/12/2021.

	AAPL	AMZN	BAC	EBAY	GS	JPM	MSFT	NFLX
AAPL	1.000000	0.079902	0.047340	0.041945	0.081837	0.011911	0.073293	0.001623
AMZN	0.079902	1.000000	0.127700	0.181139	0.307850	0.050496	0.228785	**-**0.009901
BAC	0.047340	0.127700	1.000000	0.119923	0.482570	0.021321	0.173671	**-**0.005554
EBAY	0.041945	0.181139	0.119923	1.000000	0.164216	0.014168	0.112537	0.005819
GS	0.081837	0.307850	0.482570	0.164216	1.000000	0.046455	0.296414	**-**0.009496
JPM	0.011911	0.050496	0.021321	0.014168	0.046455	1.000000	**-**0.001831	0.007643
MSFT	0.073293	0.228785	0.173671	0.112537	0.296414	**-**0.001831	1.000000	**-**0.000154
NFLX	0.001623	**-**0.009901	**-**0.005554	0.005819	**-**0.009496	0.007643	**-**0.000154	1.000000

Notes: For clarity, minus signs have been highlighted.

3.2 Proxy of actual variance

The performance of the models can only be tested by comparing the forecasted variance of the models against actual variance. A proxy of actual variance is constructed following the paper of Garman and Klass (1980). In this paper they describe the best analytic scale-invariant estimator of variance for stock i at time t as:

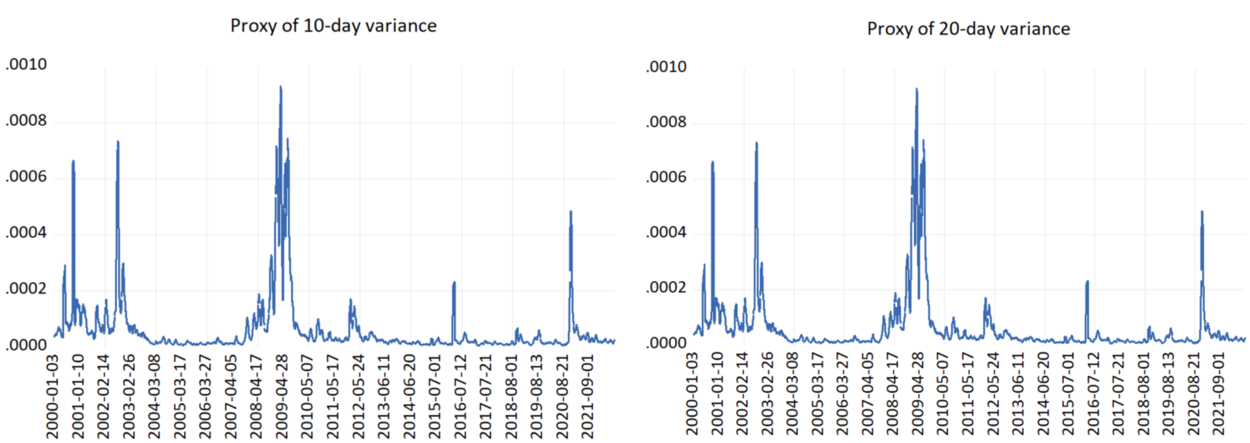
$$\sigma_{i,t}^2 = 0.5(u_{i,t} - d_{i,t})^2 - (2\ln(2) - 1)c_{i,t}^2 \quad (1)$$

Where $u_{i,t}$ is the normalized high of the log price of stock i during day t, calculated by subtracting the opening log price of the stock from the stock high (also log price) during the day. Furthermore, $d_{i,t}$ is the normalized low of the log price of stock i during day t, calculated by subtracting the stock low during the day (log price) from the opening log price of the stock. Finally, $c_{i,t}$ is the normalized closing log price of the stock, calculated by subtracting the opening log price of the stock from closing log price of the stock.

This calculated proxy of variance does not take into account the overnight return. In order to make sure that the proxy takes this overnight return into account, the proxy needs to be scaled. The scaling term looks as follows for stock i at time t:

$$\frac{\mu_{r^2_{i,t}}}{\mu_{\sigma^2_{i,t}}} \tag{2}$$

Where $\mu_{r^2_{i,t}}$ is the average squared return of stock i of the past 10 years at time t and $\mu_{\hat{\sigma}^2_{i,t}}$ is the average variance proxy of stock i of the past 10 years at time t. In this research both the 10-day proxy of variance and the 20-day proxy of variance will be used. The k-day proxy of variance for stock i at time t is calculated as $\sigma^2_{i,t,k} = \sum_{j=1}^{k} \sigma^2_{i,t+j}$, where $\sigma^2_{i,t+j}$ is the scaled proxy of variance on day $t + j$. As an example, Figure 2 shows both the 10-day proxy of variance of JPM (left) and the 20-day proxy of variance of JPM (right) from 01/01/2000 until 31/12/2021.



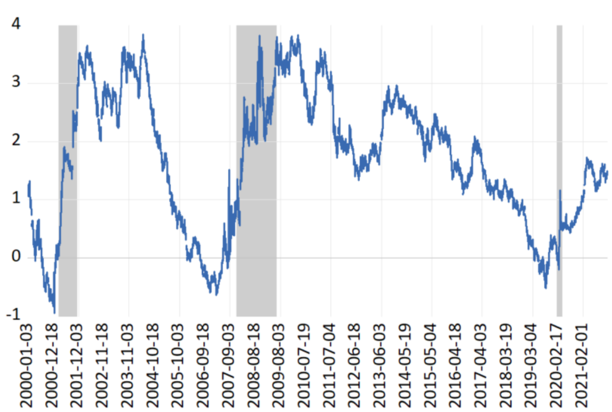
(a) Proxy of 10-day variance JPM.

(b) Proxy of 20-day variance JPM.

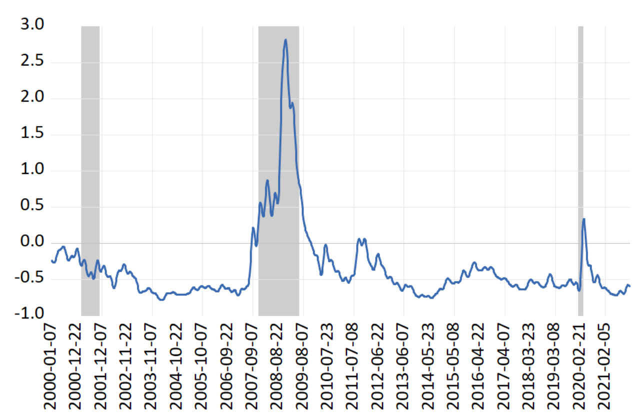
Figure 2: Proxy of the 10-day variance (left) and the 20-day variance (right) of JPM.

3.3 Leading indicators

This research is about the predictive power of leading indicators regarding variance. The considered indicators in this research are the long term Treasury yield spread and the National Financial Conditions Index. Figure 3a shows the value of the long term Treasury yield spread (ten-year minus three-month Treasury yields) in percentages from 01/01/2000 until 31/12/2021. The grey areas shown in this figure indicate U.S. recessions. The three grey areas can be linked to nine eleven, the financial crises and the outbreak of the Coronavirus. When looking at this figure it can be seen that when this long term Treasury yield spread drops (sharply) below zero percent, within about a year from that point a U.S. recession started. Figure 3b shows the value of the NFCI from 01/01/2000 until 31/12/2021. Also in this subfigure the grey areas indicate U.S. recessions. From this figure it can be seen that the NCFI has negative values most of the time. Only during recessions the NFCI really becomes positive.



(a) Value of long term Treasury yield spread.



(b) Value of National Financial Conditions Index.

Figure 3: Daily value of the long term Treasury yield spread (left) and the daily value of the NFCI (right).

4 Methodology

This section discusses the models and techniques that are used throughout this research. For all the models that will be introduced in this section it holds that they are estimated using an estimation sample of 10 years, starting with the sample from 01/01/2000 - 31/12/2009. Only the models belonging to the Netflix (NFLX) stock are estimated with a smaller sample, namely starting with the sample from 23/05/2002 - 31/12/2009. However, it holds for all stocks that a moving window will be used to construct forecasts for all the models from 01/01/2010 - 31/12/2021. This moving window will move up a place on a daily basis.

4.1 HAR model

The Heterogeneous AutoRegressive (HAR) model is an AR type of model of additive variance estimates with different horizons, see Corsi (2009). The HAR model is built on the Heterogeneous Market Hypothesis, first presented by Muller et al. (1995). This hypothesis tries to explain the empirical observation of a strong positive correlation between variance and market presence by the assumption that the market is composed of heterogeneous traders as stated by Muller et al. (1995). Under this hypothesis different agents are likely to disagree on prices and therefore execute their transactions under different market situations, creating an increasing amount of variance for increasing market presence. This heterogeneity may arise due to various reasons and this model focuses on one of them, namely the difference in the time horizon. The financial market consists of different kinds of participants, who each react to news within different time horizons.

The HAR model considered in this research will use three averages of lags of the 10- and 20-day variance. These three average lags are the 10- and 20-day variance of the trading day before, five trading days ago (one trading week) and 20 trading days ago (four trading weeks). The m-day lag is calculated as follows: ${\sigma^2_{i,t,k}}^{(m)} = \frac{1}{m}\sum_{j=1}^{m} \sigma^2_{i,t-j,k}$. Using these three lags, the HAR model will produce a forecast of the 10- and 20-day variance. Therefore the HAR model used in this research looks as follows for stock i, day t and number of days k:

$$\hat{\sigma}^2_{i,t,k} = \omega + \beta_1 {\sigma^2_{i,t,k}}^{(1)} + \beta_2 {\sigma^2_{i,t,k}}^{(5)} + \beta_3 {\sigma^2_{i,t,k}}^{(20)} + \epsilon_{i,t,k} \tag{3}$$

Where ω is a constant, β_1, β_2 and β_3 are coefficients and ${\sigma^2_{i,t,k}}^{(1)}, {\sigma^2_{i,t,k}}^{(5)}$ and ${\sigma^2_{i,t,k}}^{(20)}$ are average lags of the k-day multiple-day variance. Finally, $\epsilon_{i,t,k}$ is an error term.

4.2 Asymmetric Slope CAViaR model

Conditional AutoRegressive Value at Risk (CAViaR) models were first introduced by Engle and Manganelli (2004). These models make use of autoregressive modelling of conditional quantiles. These models first compute the Value at Risk in a semi-parametric way. However, Pearson and Tukey (1965) discovered a constant relation between the ratio of standard deviation and symmetric quantiles. Taylor (2005) uses this finding to construct a least squares regression of variance on the symmetric quantiles to model the variance. The use of quantile regressions to estimate nonlinear quantile methods is supported by White (1994). Taylor (2005) was the first to use CAViaR type of models for predicting variance.

These models do not rely on any assumption of distribution. This means that the conditional distribution over time is not fixed. The assumption that the conditional distribution of financial returns changes over time is very plausible, as economic conditions and financial markets evolve. CAViaR models directly estimate the quantiles of series of financial returns with an autoregressive model. The CAViaR type of model that is used in this research is the Asymmetric Slope CAViaR model, given by equation 4. The Asymmetric Slope CAViaR model is the same as in the paper of Engle and Manganelli (2004) and looks as follows:

$$Q_t(\theta) = \omega + \alpha Q_{t-1}(\theta) + \beta_1 \max(r_{t-1}, 0) - \beta_2 \min(r_{t-1}, 0) \tag{4}$$

Where $Q_t(\theta)$ is the conditional θ quantile and r_{t-1} is the log return at time $t-1$. Furthermore, ω is a constant and α, β_1, β_2 are the coefficients for the parameters. The model is able to handle a different reaction for positive and negative returns. For this CAViaR model two symmetric intervals will be generated: a 90% interval and a 95% interval. These specific intervals are chosen as these intervals performed the best in multi-step-ahead forecasting in the paper of Taylor (2005). The parameters of the model are estimated using the Quantile Regression Minimization (QR) Sum given in the following equation:

$$Min(\sum_{t|r_t>=Q_t(\theta)} \theta|r_t - Q_t(\theta)| + \sum_{t|r_t<Q_t(\theta)} (1-\theta)|r_t - Q_t(\theta)|) \tag{5}$$

In order to find a minimum, a value for $Q_0(\theta)$ is needed. A historical quantile of the sample is used to estimate this value. The parameters of the CAViaR models are optimized by using a specific procedure, similar to the procedure mentioned by Engle and Manganelli (2004). Namely, 4×10^5 vectors of parameters from a uniform random number generator between 0 and 1 are generated. For each vector the QR Sum in equation 5 is calculated and the 15 vectors with the lowest value of the QR Sum are used to optimize the parameters with a quasi-newton algorithm. Subsequently, the vector with the lowest QR sum will be chosen as the parameter vector, which is then used to make a forecast of the quantiles. For all optimization the code of Engle and Manganelli (2004), which was transformed to a code in R-studio by Buczyński, Chlebus, et al. (2017) was used as a base.

In order to calculate the k-period realized variance, a couple of assumptions need to be made. Namely, the conditional mean of the log returns is constant over the k days and there is no autocorrelation between successive shocks. The realized multiperiod variance of stock i can be calculated as was also mentioned in Section 3 and looks as follows:

$$\sigma^2_{i,t,k} = \sum_{j=1}^{k} \sigma^2_{i,t+j} \tag{6}$$

Making use of the one-step-ahead quantile forecasts estimated by the CAViaR model, forecasts of k-period variance can be made. As described by Taylor (2005), a least squares regression of the k-period variance on the square of the interval between symmetric quantile estimates can be performed. This looks as follows:

$$\hat{\sigma}^2_{i,t,k} = \alpha_k + \beta_k(\hat{Q}_{t+1}(1-\theta) - \hat{Q}_{t+1}(\theta))^2 \tag{7}$$

Where α_k and β_k are parameters which are estimated by the least squares regression of the k-period realized variance on the square of the interval between symmetric quantile estimates. In this research, k will be equal to both 10 and 20. Therefore the CAViaR model produces 10- and 20-day variance forecasts.

4.3 GARCHX model

The GARCHX model introduced by Hwang and Satchell (2005) is an extended version of the GARCH model, as it is capable of incorporating the effect of exogenous variables. This model incorporates for example cross-sectional market variance, as is done by Hwang and Satchell (2005). In this research two different exogenous variables will be used, resulting in two different GARCHX models. The first of these exogenous variables is the 10- or 20-day proxy of variance, which were introduced in Section 3. The exogenous variable in the second GARCHX model will be the CBOE Volatility Index, also known as the VIX. This second exogenous variable is introduced as it is very close to the cross-sectional market volatility, which is used in the paper by Hwang and Satchell (2005). The GARCHX models that will be used in this research are both GARCHX(1,1) models, meaning that only one lag of the squared log returns and one lag of the variance is used. This results in the following expression for the GARCHX model for stock i:

$$\hat{\sigma}^2_{i,t,k} = \omega + \alpha r^2_{t-1} + \beta \sigma^2_{i,t-1} + \lambda \sigma^2_{i,t-1,k} + \eta_{i,t,k} \tag{8}$$

Where ω, α, β and λ are parameters. Furthermore, r^2_{t-1} is the squared log return at time $t-1$, $\sigma^2_{i,t-1}$ represents the variance of r^2_{t-1} and $\sigma^2_{i,t-1,k}$ is the value of the k-day proxy of variance at time $t-1$ in the first GARCHX model, and it represents the last value of the VIX in the other GARCHX model. Finally, $\eta_{i,t,k}$ represents the error term. The parameters of the GARCHX models have been estimated using Quasi Maximum Likelihood.

4.4 Constructing pooled models

As mentioned at the beginning of Section 4, each of the models will produce forecasts from 01/01/2010 - 31/12/2021 using a daily moving window of the estimation sample. Each of the models will produce 10- and 20-day forecasts of variance during this period. The forecasts of each model will be weighted in order to form pooled forecasts. These pooled forecasts look as follows:

$$\begin{aligned}\hat{\sigma}^2_{i,t,k,pooled} = \omega_{HAR,t}\hat{\sigma}^2_{i,t,k,HAR} + \omega_{Sym-CAViaR90,t}\hat{\sigma}^2_{i,t,k,Sym-CAViaR90} \\ + \omega_{Sym-CAViaR95,t}\hat{\sigma}^2_{i,t,k,Sym-CAViaR95} + \omega_{GARCHX,t}\hat{\sigma}^2_{i,t,k,GARCHX} \\ + \omega_{GARCHX-VIX,t}\hat{\sigma}^2_{i,t,k,GARCHX-VIX}\end{aligned} \tag{9}$$

Where $\omega_{h,t}$ represents the weight of model h at time t and $\hat{\sigma}^2_{i,t,k,h}$ is the forecast of k-day variance of model h at time t for model i. The weight of each model lies somewhere between

0 and 1. The weights of the individual models are determined by looking at the inverse of the absolute forecast error. At each point in time the inverse of the absolute forecast error will be calculated, represented by $e_{h,t}$, where t indicates the point in time an h indicates the model. For example, the inverse of the absolute forecast error for model h at time t looks as follows:

$$e_{h,t} = (|(\sigma^2_{i,t,k} - \hat{\sigma}^2_{i,t,k,h})|)^{-1} \tag{10}$$

This absolute forecast error is calculated for each of the individual models. After this, the weight of model h at time t is calculated as:

$$\omega_{h,t} = \frac{e_{h,t}}{\sum_{j=1}^{5} e_{j,t}} \tag{11}$$

For each stock, this weighting will be done from 01/01/2010 until 31/12/2021, resulting in a time series of weights for each of the individual models. This metric is chosen to assign the weights of the individual models as this metric makes sure that at each point in time a model with a large absolute forecast error receives a small weight, and a model with a small absolute forecast error receives a large weight. Furthermore, this method is used as according to a paper by Duncan et al. (2001), relatively simple pooling methods are more accurate than complex pooling methods.

4.5 Predictive power leading indicators

In order to investigate the predictive power of the leading indicators, the time series of weights that are obtained in Section 4.4 will be regressed on values of the leading indicators. This will be done both for the weights constructed using the 10- and 20-day forecasts. These weights are obtained from 01/01/2010 - 31/12/2021. For each stock, the weights of each individual model will be regressed on the lagged values of the two leading indicators (equation 12). At each point in time in the out-of-sample period, running from 01/01/2013 - 31/12/2021, this regression is done using weights and lagged values of the past 3 years at that point in time. This is the reason why the out-of-sample period starts from 01/01/2013, instead of 01/01/2010. Also for each stock the regressions will be expanded with extra exogenous variables. Namely, the weights of each model will be regressed on the lagged values of the two leading indicators, on lagged weights of other stocks corresponding to the same model and lagged values of the actual variance (equation 13). For the weight of model h of stock i at time t these regressions looks as follows:

$$\omega_{i,h,t} = \alpha + \zeta_1 yieldspread_{t-1} + \zeta_2 nfci_{t-1} + \epsilon_{h,t} \tag{12}$$

$$\begin{aligned}\omega_{i,h,t} = \gamma &+ \beta_1 yieldspread_{t-1} + \beta_2 nfci_{t-1} + \beta_3\omega_{e,h,t-1} + \beta_4\omega_{f,h,t-1} + \beta_5\omega_{g,h,t-1} \\ &+\beta_6\omega_{j,h,t-1} + \beta_7\omega_{l,h,t-1} + \beta_8\omega_{m,h,t-1} + \beta_9\omega_{n,h,t-1} + \beta_{10}\sigma^2_{i,t-1,k} + \eta_{h,t}\end{aligned} \tag{13}$$

Where α and γ are constants. Furthermore ζ_1, ζ_2 and $\beta_1 - \beta_{10}$ are coefficients and $\epsilon_{h,t}$ and $\eta_{h,t}$ are error terms. Also $\omega_{e,h,t-1}, \omega_{f,h,t-1}, \omega_{g,h,t-1}, \omega_{j,h,t-1}, \omega_{l,h,t-1}, \omega_{m,h,t-1}$ and $\omega_{n,h,t-1}$ are lagged weights of the respective model of the other seven stocks. Finally, $yieldspread_{t-1}$ and $nfci_{t-1}$ are the lagged values of the two leading indicators and $\sigma^2_{i,t-1,k}$ is the lagged actual variance.

During the out-of-sample period, running from 01/01/2013 - 31/12/2021, the values of the coefficients found in the regressions will be used to come up with a suggested value of the weight of a model, given the last observed value of the leading indicators, lagged weight values of other stocks and the lagged value of the actual variance. During the out-of-sample period, this will be done for each model, resulting in a pooled forecast consisting of weights that are determined based on the value of leading indicators, lagged weight values of other stocks and the lagged value of the actual variance.

As mentioned earlier, this will be done for different types of stocks. Therefore it may be found that the values of the coefficients are different among different types of stocks. This may have an impact on the accuracy of the forecasts among different types of stocks. Due to the coefficients coming from the regressions, it may happen that the weights do not add up to one. If the total weight is larger than one, the difference between this sum and one is calculated. The weight of the most accurate model, in terms of forecasting accuracy over the past 3 years at that point in time, is then decreased by this number. If the total weight is smaller than one, the difference between this sum and one is calculated. The weight of the most accurate model, in terms of forecasting accuracy, is then increased by this number.

In order to determine the predictive power of both of the leading indicators (and of the lagged weight values of other stocks and the lagged value of the actual variance), the performance of the pooled forecast that is constructed based on the lagged values of the indicators will be compared to two benchmark models. The first benchmark model will be a pooled model which assigns an equal weight of $\frac{1}{5}$ to each of the models. This model is called the equally weighted pooled model. This will also be done for the out-of-sample period from 01/01/2013 - 31/12/2021. This results in the following model:

$$\begin{aligned}\hat{\sigma}^2_{i,t,k,equally-weigthed} = \frac{1}{5}\hat{\sigma}^2_{i,t,k,HAR} + \frac{1}{5}\hat{\sigma}^2_{i,t,k,Sym-CAViaR90} + \frac{1}{5}\hat{\sigma}^2_{i,t,k,Sym-CAViaR95} \\ + \frac{1}{5}\hat{\sigma}^2_{i,t,k,GARCHX} + \frac{1}{5}\hat{\sigma}^2_{i,t,k,GARCHX-VIX}\end{aligned} \tag{14}$$

The second benchmark model will also be a pooled model. This second benchmark model assigns weights to each of the individual models, equal to the mean weight of the past 3 years of the corresponding model. Again, it may happen that the weights do not add up to one. If this is the case, the same methodology described in the previous paragraph is used in order to get the total weights of the individual models at each point in time equal to one.

4.6 Performance measures

In order to evaluate the performance of each model two performance measures are used. The first performance measure is the QLIKE loss function, which looks as follows:

$$QLIKE = \frac{1}{n}\sum_{t=1}^{n}(\frac{\sigma^2_{i,t,k}}{\hat{\sigma}^2_{i,t,k}} - log(\frac{\sigma^2_{i,t,k}}{\hat{\sigma}^2_{i,t,k}}) - 1) \tag{15}$$

Where $\sigma^2_{i,t,k}$ represents the actual value of the k-day variance of stock i at time t and $\hat{\sigma}^2_{i,t,k}$ represents the predicted value of the k-day variance of stock i at that same point in time. The second performance measure is the average squared error, also known as the Mean Squared Error (MSE). The formula of the MSE looks as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\sigma^2_{i,t,k} - \hat{\sigma}^2_{i,t,k})^2 \tag{16}$$

Where again σ_t represents the actual value of the realized variance at time t and $\hat{\sigma_t}$ represents the predicted value of the realized variance at that same point in time.

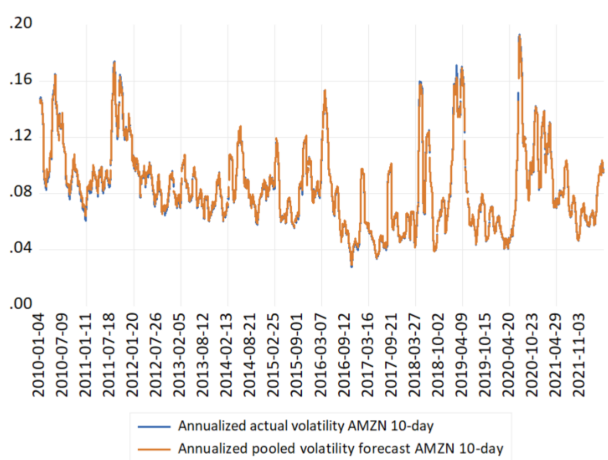
In order to test whether the forecasts of one model are significantly more accurate than the forecasts of another model, the Diebold-Mariano test will be used. The Diebold-Mariano test compares the forecast errors of two models and tests whether the forecast errors of one model are significantly larger or smaller than the forecast errors of the other model. The null hypothesis of this test is that the two models have the same forecast accuracy. As the test is performed as a two-sided test, the alternative hypothesis is that the two models have different levels of accuracy. The test statistic of the Diebold-Mariano test is based on Newey-West standard errors. In order to determine whether the difference between the forecast errors is significant, the p-value of the DM test statistic is inspected. If the p-value is smaller than 0.05, there is a significant difference between the forecast errors of the two models that are being tested.

5 Results

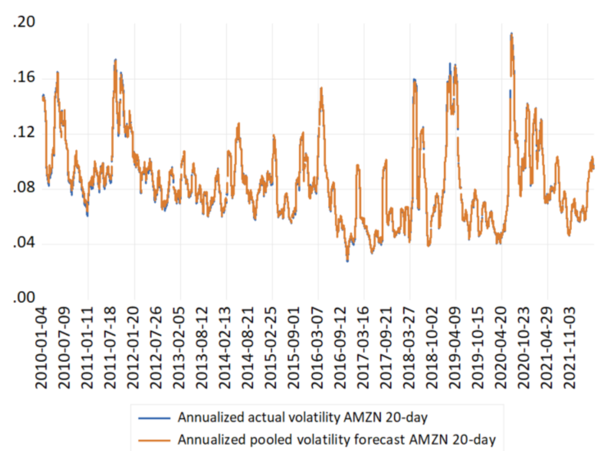
5.1 Performance of individual- and pooled models

As mentioned in Section 4, the individual models will produce forecasts from 01/01/2010 - 31/12/2021. This period is used to assign weights to each of the models and to construct a forecast-error weighted pooled forecast given these weights. The values of the QLIKE and MSE performance measures for each of the individual models and for the forecast-error weighted pooled model can be seen in Table 3 for both the 10- and 20-day variance forecasts. In this table the values of the QLIKE and MSE are explicitly mentioned for the forecast-error weighted pooled model, while for the individual models the ratio of their value of the QLIKE and MSE performance measure compared to the value of the forecast-error weighted pooled model is shown. The values in bold show the lowest value of the performance measure for each stock. What can be concluded from this table is that in 14 out of the 16 cases the forecast-error weighted pooled model has the lowest value for the QLIKE performance measure and in 15 out of the 16 cases this model has the lowest value for the MSE performance measure. This finding shows that pooling individual models is beneficial and leads to an improvement of the accuracy of the variance forecasts. Only for the 10-day variance of the AAPL stock the forecast-error weighted pooled model does not have the lowest value for both performance measures. Especially the QLIKE performance measure is much lower for the individual models compared to the forecast-error weighted pooled model regarding the 10-day variance of the AAPL stock, showing ratio's between 0.31 and 0.53.

Based on both performance measures, it can be seen that on average both the HAR model and both GARCHX models outperform the two CAViaR models. This finding shows that the CAViaR models produce variance forecasts that are not as accurate as the forecasts of the other individual models. Although this is the case, using the CAViaR models as part of the forecast-error weighted pooled model still leads to more accurate forecasts than those of the HAR model and the two GARCHX models, based on Table 3. This shows that incorporating the CAViaR models is still very useful. Figure 4 shows the annualized actual volatility against the forecast-error weighted pooled forecast of the annualized volatility of both the 10-day (Figure 4a) and the 20-day annualized volatility (Figure 4b) for AMZN from 01/01/2010 until 31/12/2021. This annualized scaling and the transformation from variance to volatility in this figure has been done for improved clarity of the given values. From these figures it can be seen that the forecast-error weighted pooled forecast is very close to the actual volatility for both the 10- and 20-day volatility. This also holds for all other stocks. This is in line with the findings in Table 3.



(a) Actual annualized volatility against forecast-error weighted pooled forecast 10-day annualized volatility.



(b) Actual annualized volatility against forecast-error weighted pooled forecast 20-day annualized volatility.

Figure 4: Actual annualized volatility against forecast-error weighted pooled forecast of the 10-day annualized volatility (left) and of the 20-day annualized volatility (right) for AMZN.

Table 3: Values of the QLIKE and MSE performance measures of all individual models and the forecast-error weighted pooled model, both for 10- and 20-day variance.

QLIKE								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
HAR	0.35	5.43	1.14	3.89	2.86	7.97	2.35	0.229
GARCHX	0.33	10.44	1.53	6.26	3.06	6.50	4.09	0.202
GARCHX-VIX	**0.31**	7.63	1.30	5.64	2.56	5.22	3.35	**0.185**
CAViaR 90%	0.53	8.76	1.90	5.04	3.33	4.95	3.08	0.290
CAViaR 95%	0.53	8.76	1.90	5.08	3.34	4.95	3.08	0.290
Forecast-error weighted pooled model	0.064	**0.0019**	**0.011**	**0.0023**	**0.0036**	**0.0026**	**0.0048**	0.060
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
HAR	4.39	8.05	3.11	4.55	4.17	20.03	3.29	4.55
GARCHX	4.23	16.80	3.96	8.46	5.22	25.22	6.55	4.65
GARCHX-VIX	3.67	12.15	3.27	6.85	3.89	19.77	5.38	3.91
CAViaR 90%	7.87	16.92	4.79	8.40	6.40	24.07	6.11	7.62
CAViaR 95%	7.87	16.91	4.80	8.53	6.41	24.11	6.13	7.62
Forecast-error weighted pooled model	**0.0020**	**0.00041**	**0.0016**	**0.00056**	**0.00079**	**0.00029**	**0.0010**	**0.0011**
MSE								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
HAR	**0.99**	2.26	1.46	1.58	2.08	7.97	1.89	5.45
GARCHX	1.54	4.62	2.08	2.84	2.87	2.68	3.43	2.80
GARCHX-VIX	1.41	3.74	1.85	2.54	2.24	2.28	2.76	2.69
CAViaR 90%	1.69	5.14	2.57	3.12	2.97	2.86	3.34	3.50
CAViaR 95%	1.69	5.13	2.57	3.13	3.16	2.88	3.16	3.50
Forecast-error weighted pooled model	9.33E-07	**1.25E-11**	**6.35E-09**	**9.18E-11**	**4.85E-12**	**3.81E-11**	**4.65E-10**	**3.75E-08**
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
HAR	2.36	4.25	1.86	1.98	4.70	13.23	2.78	8.30
GARCHX	4.29	11.09	3.37	4.23	7.09	7.55	6.04	4.09
GARCHX-VIX	3.70	8.56	2.79	3.41	5.17	5.96	4.91	3.58
CAViaR 90%	5.61	13.08	4.02	5.05	8.45	8.76	6.67	6.20
CAViaR 95%	5.61	13.05	4.02	5.08	8.33	8.87	6.68	6.20
Forecast-error weighted pooled model	**1.01E-07**	**1.74E-12**	**1.29E-09**	**1.88E-11**	**5.63E-13**	**5.02E-12**	**8.45E-11**	**9.62E-09**

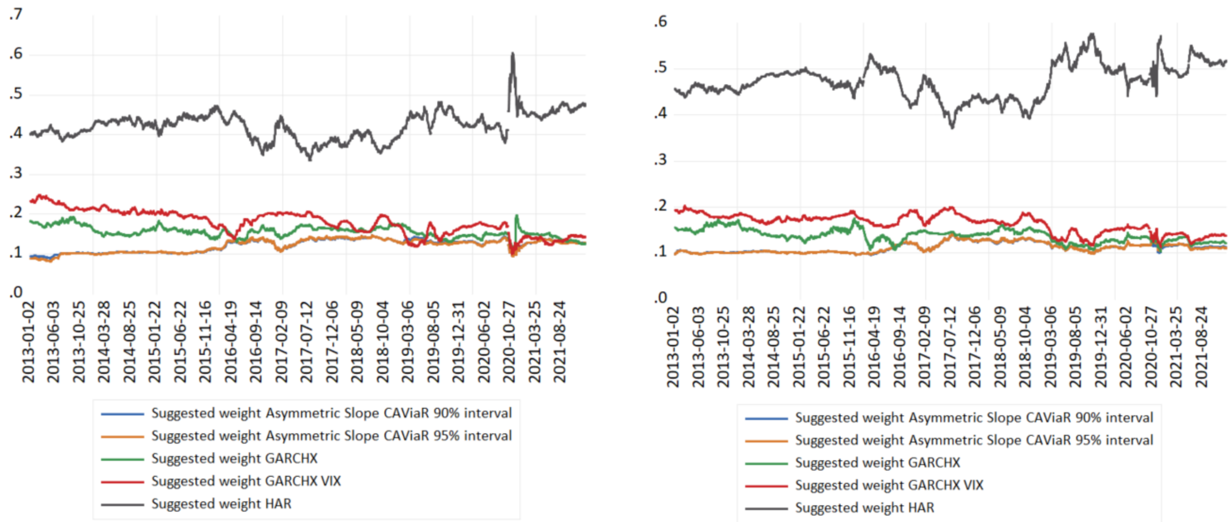
Notes: In this table the values of the QLIKE and MSE performance measures are explicitly mentioned for the forecast-error weighted pooled model, while for the individual models the ratio of their value of the QLIKE and MSE performance measure compared to the value of the forecast-error weighted pooled model is shown. The values in bold show the lowest value of the performance measure for each stock.

5.2 Linking weights to indicators and exogenous variables

The weights that are used to construct the forecast-error weighted pooled forecasts of all stocks are used as dependent variables in two different regressions: one with only the two leading indicators as explanatory variables (equation 12) and one with these two leading indicators, the lagged weights of the respective model of other stocks and the lagged actual variance as explanatory variables (equation 13). The coefficients coming from these regressions are used to construct suggested weights for each individual model of each stock, which results in a pooled forecast for each stock, with the weight of each model based on values of the relevant explanatory variables.

Figure 5 shows plots of the suggested weights for all models of AMZN based on the regressions with only the two leading indicators as explanatory variables for both the 10-day and 20-day variance. As can be seen from these figures, the HAR model receives the largest weight throughout the out-of-sample period. This is in line with the findings in Table 3, where among the individual models the HAR model is one of the most accurate models. Furthermore, it can be seen that the weights given to the two Asymmetric Slope CAViaR models are quite similar throughout the whole out-of-sample period. This is also in line with the findings in Table 3, as both CAViaR models have nearly the same values for the two performance measures. What is also interesting to see is that the weight of the HAR model and the GARCHX-VIX model seem to move in opposite direction. Their time series seem to be somewhat mirrored. These conclusions can also be drawn for these plots of the other stocks, which can be found in Appendix A.

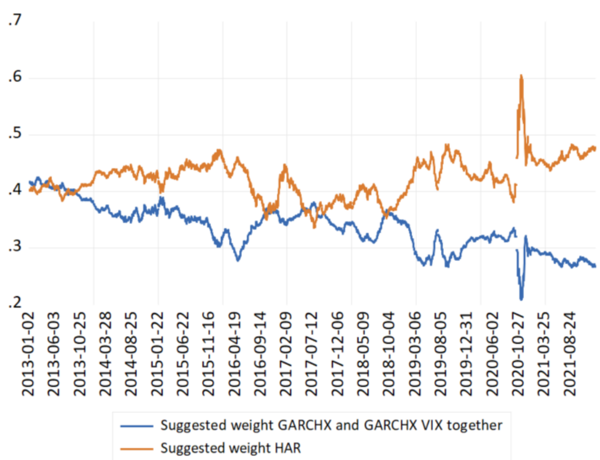
Figure 6 is a follow-up of Figure 5. This figure shows the suggested weights of the two GARCHX models summed up against the suggested weights of the HAR model for the AMZN stock. It can be seen that the suggested weights given to the two GARCHX models are somewhat mirrored compared to the suggested weight of the HAR model. As these suggested weights are based on a regression which only contains the two leading indicators, Figure 6 shows that the effect of both leading indicators are the opposite of each other for both type of models. For almost all other stocks this same result is found. The figures of the other stocks can be found in Appendix B.



(a) Suggested weights for 10-day.

(b) Suggested weights for 20-day.

Figure 5: Suggested weights for all models of AMZN based on regression with only the two leading indicators. Weights corresponding to 10-day variance (left) and 20-day variance (right).



(a) Both GARCHX models against HAR model for 10-day variance.

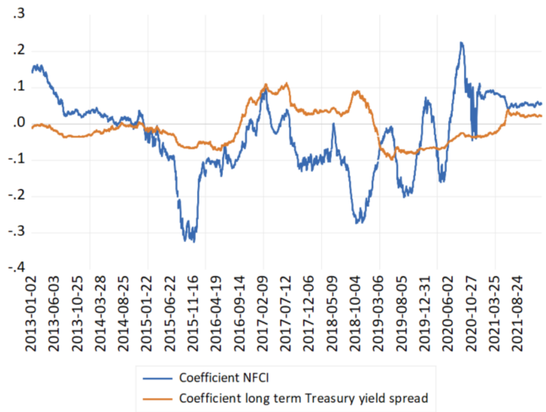
(b) Both GARCHX models against HAR model for 20-day variance.

Figure 6: Suggested weights of both GARCHX models together against the suggested weights of the HAR model of AMZN based on regression with only the two leading indicators. Weights corresponding to 10-day variance (left) and 20-day variance (right).

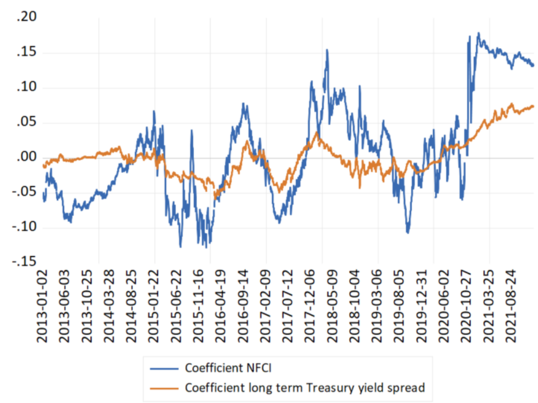
As mentioned earlier, for one of the constructed forecast-error weighted pooled models the weights are only based on the values of the leading indicators. The effect of the leading indicators on the weights of the forecast-error weighted pooled models is not constant, but varies over time. These effects are also not the same among the different stocks. Figure 7 shows plots of the coefficients belonging to each of the leading indicators for the weight of the HAR model. These HAR models are part of the forecast-error weighted pooled model that forecasts the 20-day variance of the AMZN and JPM stock. This are different type of stocks, AMZN being considered a growth stock and JPM considered a financial stock. As can be seen from the figures, the effects of the leading indicators are different over time and different among both stocks.

It can be seen that the coefficient of the NFCI is of a greater magnitude than the coefficient of the long term Treasury yield spread, which is closer to zero on average. As shown in Section 3, the values of the NFCI are on average smaller than those of the long term Treasury yield spread in absolute terms. This shows that in total the importance of the NFCI compared to the long term Treasury yield spread is smaller than that Figure 7 makes it appear. This will be further discussed in Section 5.3. Another interesting finding from these figures is that from the end of 2020, which still is a period that is heavily influenced by the Coronavirus, the coefficient of the long term Treasury yield spread and the NFCI stay relatively constant. As the regressions that determine the coefficients are based on 3-year historical data, this finding suggests that from the end of 2017 onwards the effect of both leading indicators becomes less important in constructing the forecast-error weighted pooled models. The coefficients belonging to the other individual stocks have similar results. These results correspond to the 20-day variance. The results corresponding to the 10-day variance are very similar and are therefore not explicitly shown. The plots of the coefficients of the other stocks can be found in Appendix C.

Table 4 shows the values of both performance measures for both the forecast-error weighted pooled forecasts, where the weights are based on values of the explanatory variables, and the two benchmark models. The values in bold indicate the lowest value of the performance measure for each stock. Based on the table, it can be concluded that the forecast-error weighted pooled model which has the weights based on values of the explanatory variables



(a) Coefficients for AMZN.



(b) Coefficients for JPM.

Figure 7: Coefficients of both leading indicators over time for AMZN and JPM. For both stocks, the coefficients belonging to each of the leading indicators are used to come up with a weight for the HAR model.

Notes: As can be seen from the figures, the effects of the leading indicators are different over time and different among both stocks. Also for the coefficients it holds that they get both positive and negative over time. It can be seen that the coefficient of the NFCI is of a greater magnitude than the coefficient of the long term Treasury yield spread, which is closer to zero on average.

is the best performing model in 28 out of 64 cases. The mean weighted pooled model is the best performing model in 11 out of 64 cases. Finally, the equally weighted pooled model is the best performing model in 25 out of 64 cases. Based on these results, it can be concluded that the forecast-error weighted pooled model which has the weights based on values of the explanatory variables is a very competitive model.

From Table 4 it can also be concluded that, for the regressions which only include the leading indicators as explanatory variables, in most of the cases the forecast-error weighted pooled model, which has the weights based on the values of those indicators, beats the mean weighted pooled model. These findings show that these leading indicators really have predictive power. It can also be seen that, when lagged weights of other stocks and the lagged value of the actual variance are included, the values of the performance measures on average stay the same, showing that these explanatory variables do not add significant predictive power. These results will be further discussed in Section 6.

Table 4: Values of the QLIKE and MSE performance measures of all pooled models, both for 10- and 20-day variance.

QLIKE								
Indicators only								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	**0.027**	0.012	**0.033**	0.015	**0.016**	0.015	0.017	0.019
Mean weighted pooled model	1.012	0.984	1.017	0.995	1.001	0.958	0.987	**0.982**
Equally weighted pooled model	1.055	**0.763**	1.492	**0.988**	1.160	**0.669**	**0.921**	1.193
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	0.447	0.435	**0.253**	**0.332**	0.240	0.687	**0.262**	**0.316**
Mean weighted pooled model	**0.997**	1.005	1.003	1.002	1.002	0.998	1.003	1.002
Equally weighted pooled model	1.020	**0.978**	1.100	1.023	**0.987**	**0.842**	1.020	1.056
All regressors								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	0.028	0.011	**0.034**	**0.015**	**0.016**	0.015	0.017	0.020
Mean weighted pooled model	**0.987**	1.037	1.003	1.039	0.995	0.964	1.020	**0.976**
Equally weighted pooled model	1.029	**0.804**	1.471	1.031	1.154	**0.673**	**0.951**	1.186
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	0.449	0.436	**0.254**	**0.332**	0.240	0.658	**0.262**	**0.315**
Mean weighted pooled model	**0.993**	1.004	1.000	1.000	1.003	1.043	1.001	1.004
Equally weighted pooled model	1.015	**0.977**	1.096	1.022	**0.988**	**0.880**	1.018	1.057
MSE								
Indicators only								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	**6.02E-07**	**1.53E-11**	**9.24E-09**	**1.41E-10**	**4.75E-12**	1.07E-10	4.43E-10	8.29E-08
Mean weighted pooled model	1.129	1.011	1.010	1.012	1.004	0.982	**0.997**	**0.973**
Equally weighted pooled model	1.370	1.224	1.456	1.439	1.403	**0.774**	1.396	1.085
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	**8.65E-06**	2.25E-10	**8.02E-08**	3.04E-09	**1.64E-10**	1.20E-09	8.01E-09	9.32E-07
Mean weighted pooled model	1.024	1.002	1.001	1.001	1.002	1.000	1.003	0.998
Equally weighted pooled model	1.002	**0.995**	1.012	**0.985**	1.017	**0.997**	**0.992**	**0.970**
All regressors								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	**6.20E-07**	**1.35E-11**	**9.23E-09**	**1.16E-10**	5.89E-12	1.08E-10	**3.38E-10**	8.48E-08
Mean weighted pooled model	1.097	1.145	1.012	1.228	**0.810**	0.967	1.308	**0.952**
Equally weighted pooled model	1.330	1.386	1.457	1.747	1.132	**0.762**	1.831	1.060
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	**8.62E-06**	2.28E-10	8.11E-08	3.07E-09	1.66E-10	1.20E-09	8.08E-09	9.25E-07
Mean weighted pooled model	1.029	0.986	**0.990**	0.991	**0.990**	1.000	0.994	1.006
Equally weighted pooled model	1.006	**0.980**	1.001	**0.976**	1.004	**0.998**	**0.983**	**0.978**

Notes: The values of the QLIKE and MSE performance measures are explicitly mentioned for the forecast-error weighted pooled model, while for the two benchmark models the ratio of their value of the QLIKE and MSE performance measure compared to the value of the forecast-error weighted pooled model is shown. The values in bold show the lowest value of the performance measure for each stock.

As the forecast-error weighted pooled model and both the mean weighted- and the equally weighted pooled model seem to be very competitive in terms of forecasting accuracy, Diebold-Mariano tests will be performed to find out whether some models produce more accurate forecasts than other models. The results of these tests for each stock can be found in Table 5, with the p-values in brackets. A negative value of the DM statistic in this table means that the forecasts of the forecast-error weighted pooled model are more accurate than those of the mean weighted pooled forecast/equally weighted pooled forecast. A positive value of the DM statistic means that the opposite is the case. Furthermore, a p-value smaller than 0.05 indicates that the corresponding DM statistic is significant at a 5% level. The bold values in the table indicate that the DM statistic is significant.

Looking at Table 5, it can be seen that for the regressions, which only include the leading indicators as explanatory variables, the forecast-error weighted pooled model has significantly more accurate forecasts in 8 out of the 16 cases when testing the mean weighted pooled model. The forecast-error weighted pooled model has significantly more accurate forecasts in 6 out of the 16 cases when testing the equally weighted pooled model. The mean weighted pooled model has significantly more accurate forecasts in 1 out of the 16 cases and the equally weighted pooled model has significantly more accurate forecasts in none of the 16 cases. When looking at the regressions, which include more explanatory variables, it can be seen that the forecast-error weighted pooled model has significantly more accurate forecasts in 3 out of the 16 cases when testing the mean weighted pooled model. The forecast-error weighted pooled model has significantly more accurate forecasts in 5 out of the 16 cases when testing the equally weighted pooled model. The mean weighted pooled model has significantly more accurate forecasts in 3 out of the 16 cases and the equally weighted pooled model has significantly more accurate forecasts in 1 of the 16 cases. These results indicate that the forecast-error weighted pooled models are very good competitors for the two benchmark models. Furthermore, it can be seen from Table 5 that the forecast-error weighted pooled model is most of the time more accurate than the benchmark models for the tech- and financial stocks. This is an interesting finding, which will be further discussed in Section 6.

Table 5: Values of the DM test statistic between the forecast-error weighted pooled model and the two benchmark pooled models, both for 10- and 20-day variance. Corresponding p-values are shown in brackets.

DM test with mean weighted pooled model								
Indicators only								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
DM statistic	**-2.32 (0.02)**	-1.45 (0.15)	**-5.80 (0.00)**	**-2.98 (0.00)**	-1.15 (0.25)	**2.11 (0.04)**	0.77 (0.44)	1.85 (0.06)
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
DM statistic	-1.63 (0.10)	**-3.30 (0.00)**	**-2.32 (0.02)**	**-4.65 (0.00)**	**-4.17 (0.00)**	-0.15 (0.89)	**-2.54 (0.01)**	0.91 (0.36)
All regressors								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
DM statistic	-0.94 (0.35)	**-2.63 (0.01)**	-0.37 (0.71)	**-2.45 (0.01)**	**2.41 (0.02)**	0.31 (0.75)	**-3.03 (0.00)**	1.09 (0.27)
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
DM statistic	-1.18 (0.24)	**3.53 (0.00)**	**3.07 (0.00)**	1.48 (0.14)	1.41 (0.16)	-0.09 (0.93)	0.84 (0.40)	-0.67 (0.51)
DM test with equally weighted pooled model								
Indicators only								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
DM statistic	**-4.40 (0.00)**	**-3.58 (0.00)**	**-6.32 (0.00)**	**-4.25 (0.00)**	**-2.86 (0.00)**	1.32 (0.19)	**-3.65 (0.00)**	-0.64 (0.53)
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
DM statistic	-0.13 (0.89)	0.69 (0.49)	-0.91 (0.36)	1.47 (0.14)	-1.60 (0.11)	0.19 (0.85)	0.71 (0.48)	1.03 (0.30)
All regressors								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
DM statistic	**-4.98 (0.00)**	**-3.36 (0.00)**	**-4.98 (0.00)**	**-3.66 (0.00)**	1.07 (0.28)	1.12 (0.26)	**-3.44 (0.00)**	-0.40 (0.69)
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
DM statistic	-0.44 (0.66)	**2.03 (0.04)**	-0.07 (0.95)	1.57 (0.12)	-0.43 (0.67)	0.14 (0.89)	0.95 (0.34)	1.09 (0.28)

Notes: A negative value of the DM statistic in this table means that the forecasts of the forecast-error weighted pooled model are more accurate than those of the mean weighted pooled forecast/equally weighted pooled forecast. A positive value of the DM statistic means that the opposite is the case. Furthermore, a p-value smaller than 0.05 indicates that the corresponding DM statistic is significant at a 5% level. The bold values in the table indicate that the DM statistic is significant.

In order to see whether the forecasts of the forecast-error weighted pooled model with only the leading indicators as explanatory variables are significantly more accurate than those of the forecast-error weighted pooled model with more explanatory variables, again Diebold-Mariano tests can be performed. The results of these tests for each stock can be found in Table 6, with the p-values in brackets. A negative value of the DM statistic in this table means that the forecasts of the forecast-error weighted pooled model, which is based on all explanatory variables, are more accurate than those of the forecast-error weighted pooled model with only the leading indicators as explanatory variables. A positive value of the DM statistic means that the opposite is the case. Furthermore, a p-value smaller than 0.05

indicates that the corresponding DM statistic is significant at a 5% level. The bold values in the table indicate that the DM statistic is significant.

Looking at Table 6, it can be seen that in 3 out of the 16 cases the DM statistic is significantly negative. In 3 out of the 16 cases the DM statistic is significantly positive. This finding shows that on average none of the two forecast-error weighted pooled models is more accurate than the other. This is in line with the findings shown in Table 4, where both forecast-error weighted pooled models had somewhat the same values for both performance measures. Figure 8 shows a plot of both pooled forecasts against the actual annualized volatility for the AMZN stock, for the 20-day annualized volatility. The plot is about May 2021. Again, this annualized scaling and the transformation from variance to volatility in this figure has been done for improved clarity of the given values. It can be seen from the plot that both forecast-error weighted pooled models produce comparable forecasts of volatility. From this figure it also seems that the forecast-error weighted pooled model, which is based on all explanatory variables, is a little bit closer to the actual volatility than the forecast-error weighted pooled model with only the leading indicators as explanatory variables. Based on the performed Diebold-Mariano tests shown in Table 6 it can be concluded that this difference is not significant. For the other stocks these results are similar.

Table 6: Values of the DM test statistic between the forecast-error weighted pooled model with all explanatory variables and the forecast-error weighted pooled model with only the leading indicators as explanatory variables, both for 10- and 20-day variance. Corresponding p-values are shown in brackets.

10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
DM statistic	-0.47 (0.64)	**2.57 (0.01)**	0.03 (0.97)	**2.40 (0.02)**	**-2.42 (0.02)**	-0.15 (0.88)	**3.05 (0.00)**	-0.64 (0.52)
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
DM statistic	0.42 (0.68)	**-3.98 (0.00)**	**-3.41 (0.00)**	-1.60 (0.11)	-1.62 (0.11)	0.08 (0.94)	-1.17 (0.24)	0.73 (0.47)

Notes: A negative value of the DM statistic in this table means that the forecasts of the forecast-error weighted pooled model which is based on all explanatory variables are more accurate than those of the forecast-error weighted pooled model with only the leading indicators as explanatory variables. A positive value of the DM statistic means that the opposite is the case. Furthermore, a p-value smaller than 0.05 indicates that the corresponding DM statistic is significant at a 5% level. The bold values in the table indicate that the DM statistic is not significant.

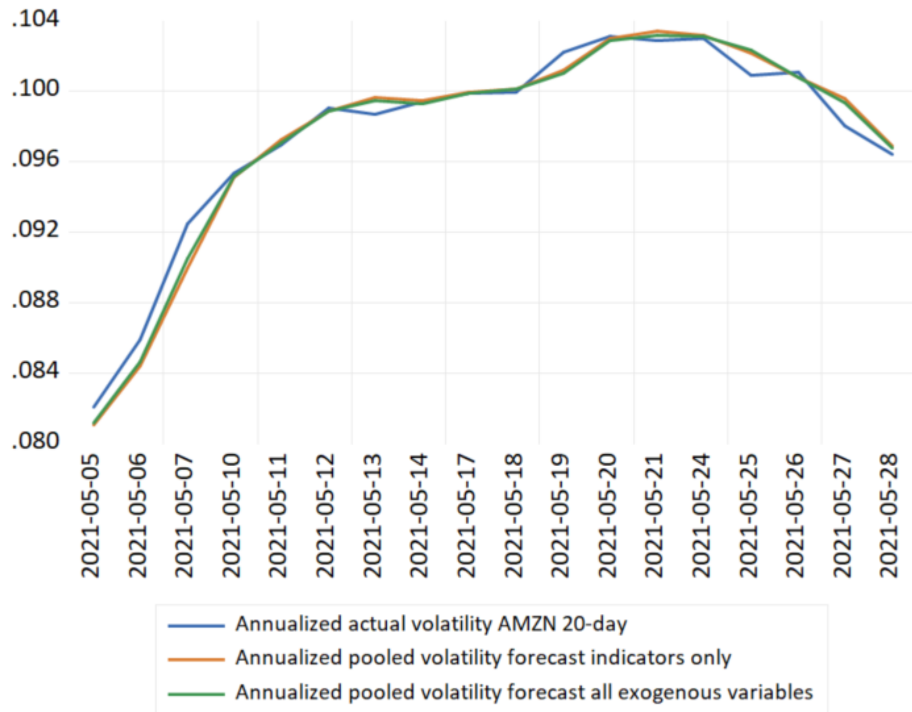


Figure 8: Both pooled forecasts against the actual annualized volatility for the AMZN stock, for the 20-day variance, May 2021.

From Table 5 it can be concluded that both leading indicators have predictive power regarding variance, as in a lot of cases the benchmark models are beaten or the forecast-error weighted pooled model is not significantly beaten by the benchmark models. From Tables 5 and 6 and Figure 8 it can also be concluded that when, next to the two leading indicators, other variables are included in the regressions, the performance of the forecast-error weighted pooled model improves a little bit and moves a little bit closer towards the actual variance (shown as volatility in Figure 8). However, this improvement is not significant. From Figure 8 it can also be seen that when relatively large changes in the level of the actual volatility occur, the forecast-error weighted pooled model, which includes other variables next to the two leading indicators, is closer to the actual volatility than the forecast-error weighted pooled model that only incorporates the two leading indicators. Nevertheless, these results suggest that incorporating other leading indicators or more exogenous variables could improve the performance of the forecast-error weighted pooled models a bit further, which may eventually lead to more accurate variance forecasts. In Section 6 this will be further discussed.

5.3 Performance of leading indicators individually

The previous results have shown the predictive power of the two leading indicators combined, but individual results of the indicators have not been shown yet. The performance of the weights in the forecast-error weighted pooled model have also been determined when the weights are constructed based on only one leading indicator. For both performance measures these results can be found in Table 7. The values of the QLIKE and MSE performance measures are explicitly mentioned for the forecast-error weighted pooled model, while for the two benchmark models the ratio of their value of the QLIKE and MSE performance measure compared to the value of the forecast-error weighted pooled model is shown. The values in bold show the lowest value of the performance measure for each stock.

From Table 7 it can be seen that when only the long term Treasury yield spread is used to construct the weights, the forecast-error weighted pooled model has the lowest value for the performance measures in 10 out of the 32 cases. It can also be seen that when only the NFCI is used to construct the weights, the forecast-error weighted pooled model has the lowest value for the performance measures in 16 out of the 32 cases. These results indicate that, regarding multiple-day variance, the predictive power of the NFCI is a little bit larger than the predictive power of the long term Treasury yield spread.

Although the NFCI performs quite well compared to the long term Treasury yield spread, also incorporating the long term Treasury yield spread into the construction of the forecast-error weighted pooled models is still beneficial, as the values of both performance measures are a little lower in Table 4 compared to Table 7. To conclude this section, it can be seen that the forecast-error weighted pooled models based on one of the two leading indicators perform a little bit worse on average than when both indicators are incorporated. In turn, the forecast-error weighted pooled models based on both leading indicators perform a little bit worse on average than when both indicators and extra explanatory variables are incorporated. This suggests that incorporating more explanatory variables could make the performance of the forecast-error weighted pooled model even more accurate. This will be further discussed in Section 6.

Table 7: Values of the QLIKE and MSE performance measures of all pooled models, both for 10- and 20-day variance. Regressions are based on only one of the two leading indicators.

QLIKE								
Only yield spread								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	0.028	0.012	**0.033**	0.015	0.016	0.015	0.017	0.019
Mean weighted pooled model	**0.993**	1.006	1.015	0.999	**0.987**	0.963	0.981	**0.985**
Equally weighted pooled model	1.035	**0.779**	1.489	**0.992**	1.144	**0.672**	**0.915**	1.196
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	0.446	0.435	**0.254**	0.333	0.240	0.687	**0.262**	**0.316**
Mean weighted pooled model	**0.998**	1.006	1.002	**1.000**	1.001	0.999	1.003	1.002
Equally weighted pooled model	1.020	**0.979**	1.098	1.021	**0.986**	**0.843**	1.020	1.055
Only NFCI								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	**0.027**	0.012	**0.033**	0.015	**0.015**	0.015	0.017	0.019
Mean weighted pooled model	1.014	0.982	1.013	0.991	1.009	0.979	1.000	**0.991**
Equally weighted pooled model	1.057	**0.761**	1.485	**0.984**	1.170	**0.684**	**0.933**	1.203
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	0.447	0.437	**0.253**	**0.332**	0.240	0.689	**0.263**	**0.316**
Mean weighted pooled model	**0.996**	1.001	1.003	1.003	1.001	0.996	1.000	1.001
Equally weighted pooled model	1.018	**0.974**	1.099	1.024	**0.986**	**0.840**	1.018	1.054
MSE								
Only yield spread								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	**6.80E-07**	**1.53E-11**	**9.25E-09**	**1.42E-10**	4.78E-12	1.06E-10	4.45E-10	8.16E-08
Mean weighted pooled model	1.001	1.015	1.009	1.005	**0.998**	0.990	**0.993**	**0.988**
Equally weighted pooled model	1.214	1.230	1.453	1.430	1.393	**0.780**	1.390	1.101
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	8.87E-06	2.25E-10	**8.02E-08**	3.04E-09	**1.64E-10**	1.20E-09	8.02E-09	9.31E-07
Mean weighted pooled model	1.001	1.001	1.001	1.000	1.001	1.000	1.001	0.999
Equally weighted pooled model	**0.978**	**0.995**	1.012	**0.984**	1.016	**0.997**	**0.990**	**0.971**
Only NFCI								
10-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	**6.05E-07**	**1.54E-11**	**9.28E-09**	**1.42E-10**	**4.74E-12**	1.07E-10	**4.41E-10**	8.29E-08
Mean weighted pooled model	1.124	1.007	1.005	1.004	1.005	0.978	1.002	**0.973**
Equally weighted pooled model	1.363	1.219	1.449	1.428	1.404	**0.771**	1.403	1.085
20-day	AAPL	AMZN	EBAY	MSFT	GS	JPM	BAC	NFLX
Forecast-error weighted pooled model	**8.67E-06**	2.25E-10	**8.02E-08**	3.04E-09	**1.64E-10**	1.20E-09	8.01E-09	9.32E-07
Mean weighted pooled model	1.023	1.001	1.001	1.001	1.001	1.000	1.002	0.998
Equally weighted pooled model	1.001	**0.995**	1.012	**0.985**	1.016	**0.997**	**0.991**	**0.970**

Notes: The values of the QLIKE and MSE performance measures are explicitly mentioned for the forecast-error weighted pooled model, while for the two benchmark models the ratio of their value of the QLIKE and MSE performance measure compared to the value of the forecast-error weighted pooled model is shown. The values in bold show the lowest value of the performance measure for each stock.

6 Discussion & Conclusion

Forecasting variance is important for a lot of investors, companies and even governments. Due to this importance, many classes of models have been developed over the years. As each of these classes have their own advantages and disadvantages, this paper focuses on combining those models into a pooled model in order to get more accurate forecasts due to this diversification. The following three classes are considered: HAR type of models, GARCH type of models and CAViaR type of models. The weight that each individual model receives within the pooled model at each point in time is determined by the long term Treasury yield spread and the NFCI. The first pooled model that is used as a benchmark assigns a weight to each model based on the mean weight of the past three years. The second pooled model that is used as a benchmark assigns an equal weight to each of the individual models. The main research question of this paper is therefore as follows: *"Can the level of both the long term Treasury yield spread and the NFCI contribute to constructing pooled forecasts that are significantly more accurate than both equally weighted- and mean weighted pooled forecasts regarding multiple-day variance?"*

In order to answer this main research question, an accurate proxy of actual variance needs to be established first. This proxy is constructed following the paper of Garman and Klass (1980), where their proxy incorporates highs and lows during the trading day to get an even more accurate proxy of actual variance. This proxy is converted to multiple-day variance in order to answer the main research question, namely to 10- and 20-day variance. On the basis of this proxy of variance, pooled models are constructed for each of the stocks by assigning a weight to an individual model based on the size of the forecast error at each point in time. If an individual model has a large forecast error compared to the other individual models, this individual model receives a small weight, and the other way around if an individual model has a small forecast error compared to the other individual models. The construction of this forecast-error weighted pooled model is done from 01/01/2010 until 31/12/2021. It is found that this pooling of models is beneficial for all investigated stocks, as the forecast-error weighted pooled models have lower values of the QLIKE and MSE performance measures compared to the individual models.

Based on regressions that use 3-year historical data, a relation between the weight assigned to an individual model and the leading indicators can be found, which is used to come up with a suggested weight of this individual model. As each individual model receives a suggested weight, a pooled forecast of variance can be formed. It is found that the formed forecast-error weighted pooled models produce accurate forecasts and that they are really good competitors for the benchmark models. This shows that the used leading indicators really have predictive power. Especially for financial- and tech stocks the forecast-error weighted pooled model performs really well compared to the benchmark models. For these type of stocks further improvement of the forecast-error weighted pooled model could let this pooled model beat the benchmark models. Only for the Netflix stock the forecast-error weighted pooled model is not able to produce forecasts that are significantly more accurate than those of the benchmark models. As Netflix is purely a growth stock, this finding might indicate that for growth stocks forecast-error weighted pooled models are less effective for the prediction of variance than for other type of stocks.

The performance of the constructed pooled model has also been determined when the weights are based on only one of the two leading indicators. Results indicate that the NFCI contains more predictive power than the long term Treasury yield spread. However, it is found that incorporating both leading indicators leads to more accurate variance forecasts compared to incorporating only one indicator. In order to get even more accurate forecasts, some extra variables are included regarding the construction of the weights of the individual models. The extra variables are the lagged actual variance and lagged weights of the same individual model of the other stocks. Including these extra variables has led to forecasts that are a little bit more accurate than when only the two leading indicators are incorporated. However, this improvement is not significant. This has been tested using a Diebold-Mariano test. However, other findings suggest that including extra variables might lead to even more accurate forecasts of variance, maybe even forecasts that are significantly more accurate. Furthermore, it is found that when the weights are based on only the two indicators, the weight assigned to the HAR model and the two weights of given to the GARCHX models added up move in opposite direction, indicating that the effect of both leading indicators are the opposite for the two type of models.

Based on all findings of this paper, it can be concluded that both the long term Treasury yield spread and the NFCI can construct weights leading to a pooled forecast that are very strong competitors for both benchmark models. On average, the produced forecasts are not significantly more accurate than those of the benchmark models. However, further improvement of the models could lead to forecasts that are significantly more accurate. This answers the main research question. This conclusion especially holds for the considered tech- and financial stocks.

Based on the findings and used methods in this paper, there are some suggestions for further research. Regarding the indicators used in this paper, a first suggestion would be to expand the set of leading indicators and to investigate whether other leading indicators also have predictive power regarding the prediction of variance, and maybe even contain more predictive power than the two leading indicators used in this paper. An example of other indicators is the Sahm Rule Recession indicator, which signals the start of a recession when the three-month moving average of the national unemployment rate rises by 0.50 percentage points or more relative to its low during the previous 12 months. Other examples are the unemployment rate, nominal GDP growth and the Conference Board Leading Economic Index. The set of recession indicators in the U.S. is very large, which provides enough room for investigation.

Another suggestion for further research is incorporating more exogenous variables in the construction of the weights. It might be interesting to incorporate both the skewness and kurtosis of the distribution of the weight itself over the past three years. Also using the market variance might lead to more accurate forecasts. Furthermore, this paper incorporates three types of models. As there are many other model classes and techniques, these can also be tested in order to get to more accurate forecasts or new insights. A final suggestion for further research would be that the way in which the weights are initially determined can be done in other ways too. In this paper the inverse of the absolute forecast error is used. Other methods that can be used are Bayesian Model Averaging or shrinkage methods like Least Absolute Shrinkage and Selection Operator (LASSO) or Elastic Net.

References

Amburgey, A., & McCracken, M. W. (2022). On the real-time predictive content of financial conditions indices for growth. *FRB St. Louis Working Paper*.

Buczyński, M., Chlebus, M. et al. (2017). *Is caviar model really so good in value at risk forecasting? evidence from evaluation of a quality of value-at-risk forecasts obtained based on the: Garch (1, 1), garch-t (1, 1), garch-st (1, 1), qml-garch (1, 1), caviar and the historical simulation models depending on the stability of financial markets* (tech. rep.).

Chinn, M., & Kucko, K. (2015). The predictive power of the yield curve across countries and time. *International Finance*, *18(2)*, 129–156.

Christiansen, C. (2013). Predicting severe simultaneous recessions using yield spreads as leading indicators. *Journal of International Money and Finance*, *32*, 1032–1043.

Clements, M. P., & Galvão, A. B. (2009). Forecasting us output growth using leading indicators: An appraisal using midas models. *Journal of Applied Econometrics*, *24*, 1187–1206.

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

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, *22*(5), 1915–1953.

Duncan, G. T., Gorr, W. L., & Szczypula, J. (2001). Forecasting analogous time series. *In Principles of forecasting*, 195–213.

Emenike Kalu, O., & Okwuchukwu, O. (2014). Stock market return volatility and macroeconomic variables in nigeria. *International journal of empirical finance*, *2*(2), 75–82.

Engle, R. F., & Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of business & economic statistics*, *22*(4), 367–381.

Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, *53*(1), 67–78.

Hännikäinen, J. (2017). When does the yield curve contain predictive power? evidence from a data-rich environment. *International Journal of Forecasting*, *33(4)*, 1044–1064.

Hwang, S., & Satchell, S. E. (2005). Garch model with cross-sectional volatility: Garchx models. *Applied Financial Economics*, *15*(3), 203–216.

Kourentzes, N., Barrow, D., & Petropoulos, F. (2019). Another look at forecast selection and combination: Evidence from forecast pooling. *International Journal of Production Economics*, *209*, 226–235.

Muller, U. A., Dacorogna, M. M., Dave, R. D., Pictet, O. V., Olsen, R. B., & Ward, J. R. (1995). Fractals and intrinsic time – a challenge to econometricians.

Newbold, P., & Harvey, D. I. (2002). Forecast combination and encompassing. *A companion to economic forecasting*, *1, 620*.

Opschoor, A., van Dijk, D., & van der Wel, M. (2014). Predicting volatility and correlations with financial conditions indexes. *Journal of Empirical Finance*, *29*, 435–447.

Pearson, E. S., & Tukey, J. W. (1965). Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika*, *52*(3/4), 533–546.

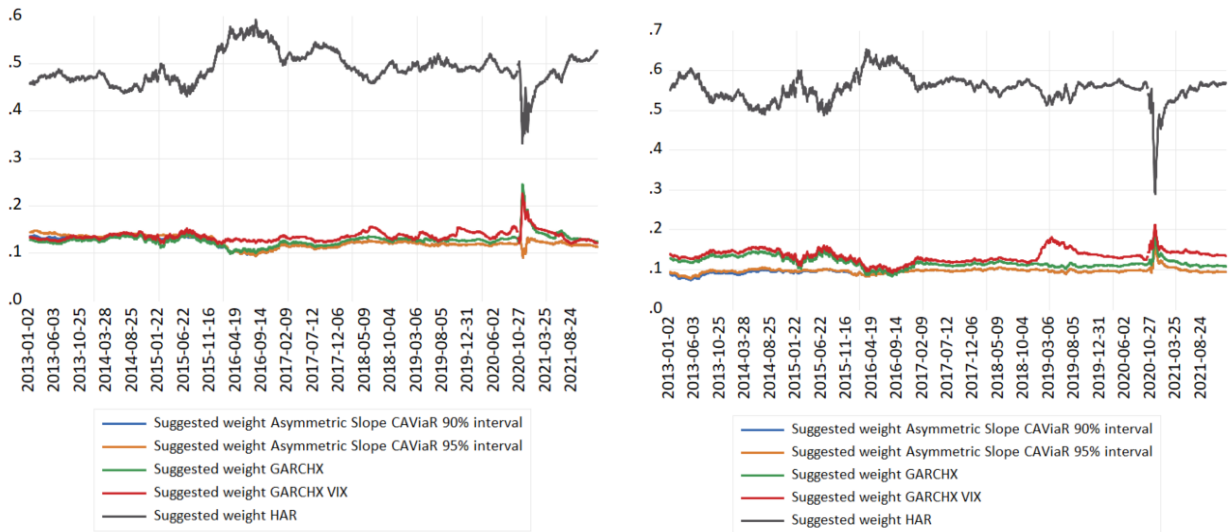
Taylor, J. W. (2005). Generating volatility forecasts from value at risk estimates. *Management Science*, *51*(5), 712–725.

White, H. (1994). Estimation, inference and specification analysis. *Cambridge University Press, Cambridge, UK.*

Yeasin, M., Singh, K. N., Lama, A., & Paul, R. K. (2020). Modelling volatility influenced by exogenous factors using an improved garch-x model. *Journal of the Indian Society of Agricultural Statistics*, *74*, 209–216.

A Suggested weights for all individual stocks

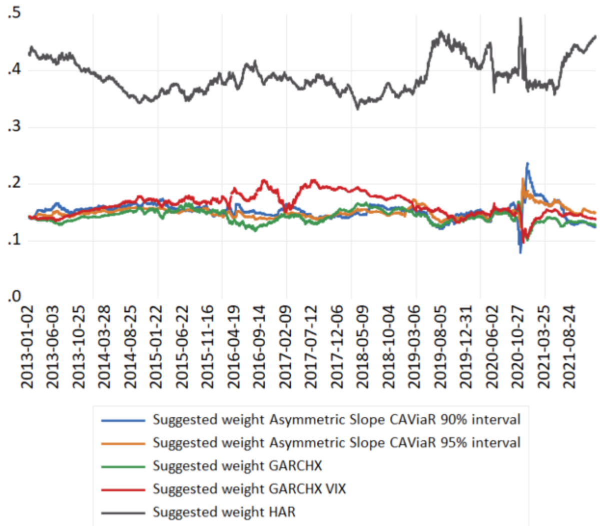
This appendix shows the suggested weights for all five individual models of based on regression with only the two leading indicators. These suggested weights are shown for each stock individually. Both the suggested weights of the 10-day variance and the 20-day variance are shown for each stock.



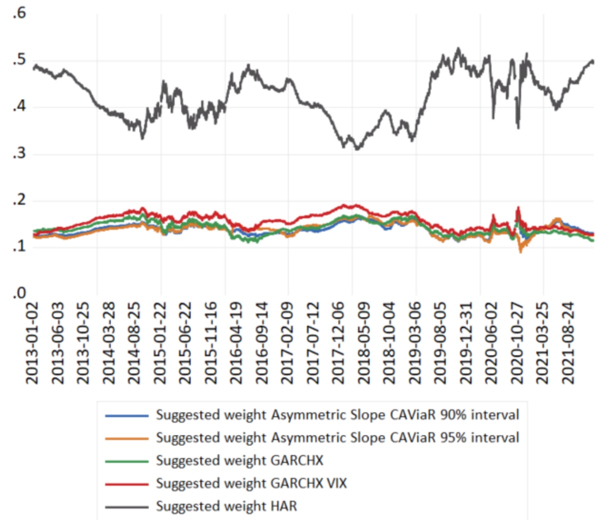
(a) Suggested weight of all models for 10-day.

(b) Suggested weight of all models for 20-day.

Figure 9: Suggested weights for all models of AAPL based on regression with only the two leading indicators. Weights of 10-day variance (left) and 20-day variance (right).

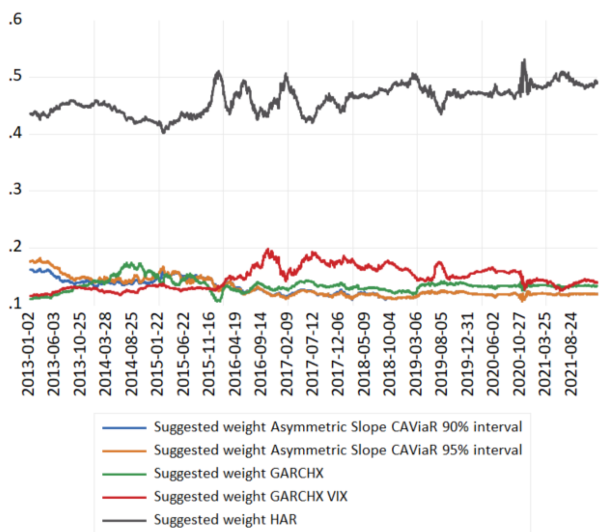


(a) Suggested weight of all models for 10-day.

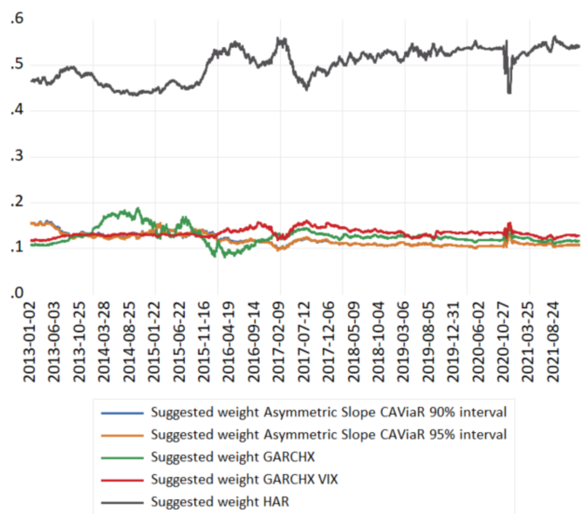


(b) Suggested weight of all models for 20-day.

Figure 10: Suggested weights for all models of BAC based on regression with only the two leading indicators. Weights of 10-day variance (left) and 20-day variance (right).

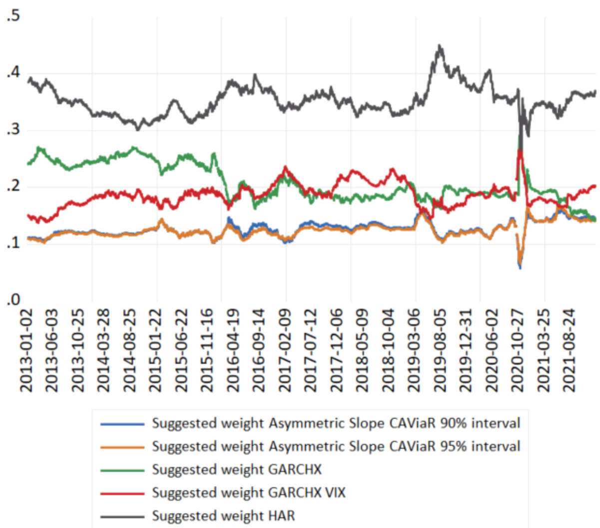


(a) Suggested weight of all models for 10-day.

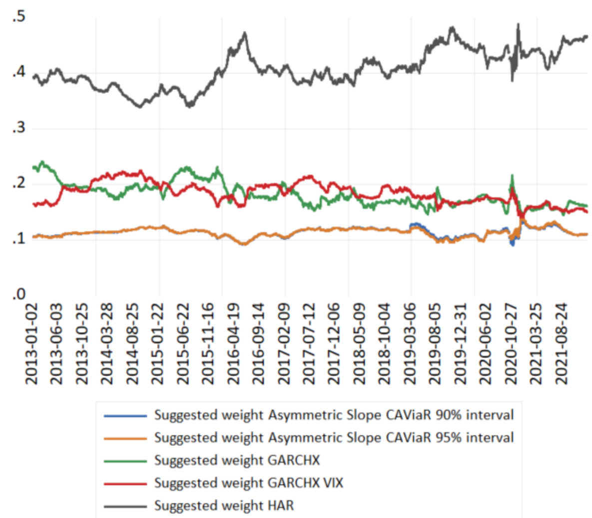


(b) Suggested weight of all models for 20-day.

Figure 11: Suggested weights for all models of EBAY based on regression with only the two leading indicators. Weights of 10-day variance (left) and 20-day variance (right).

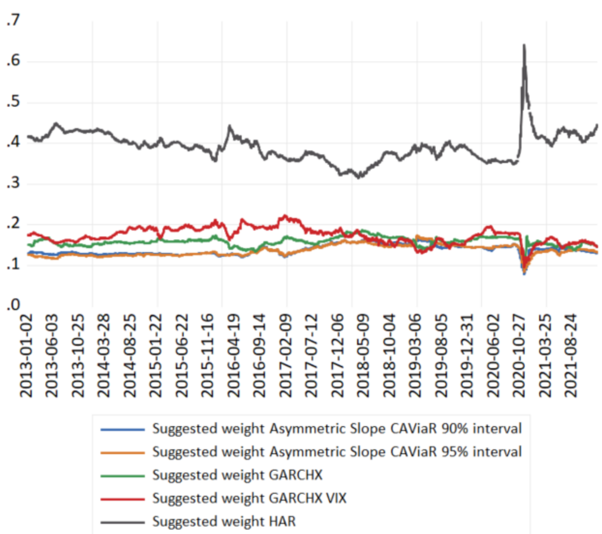


(a) Suggested weight of all models for 10-day.

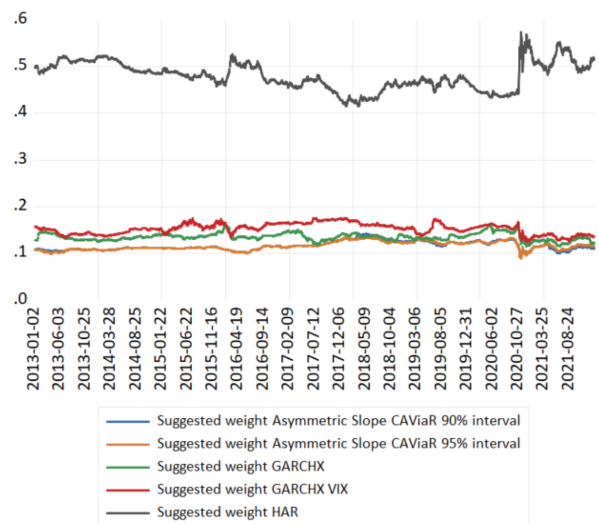


(b) Suggested weight of all models for 20-day.

Figure 12: Suggested weights for all models of GS based on regression with only the two leading indicators. Weights of 10-day variance (left) and 20-day variance (right).

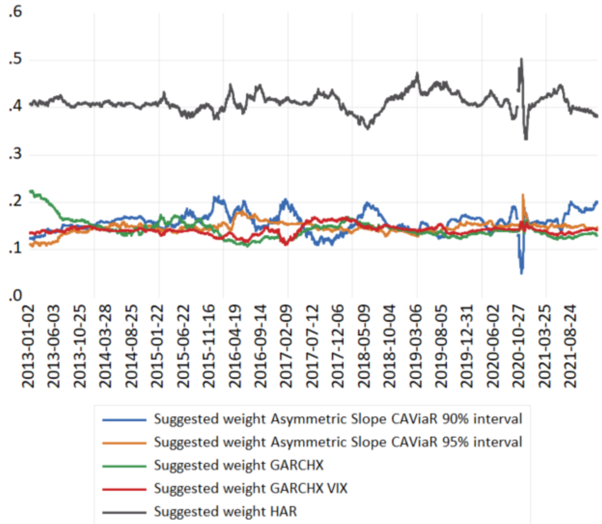


(a) Suggested weight of all models for 10-day.

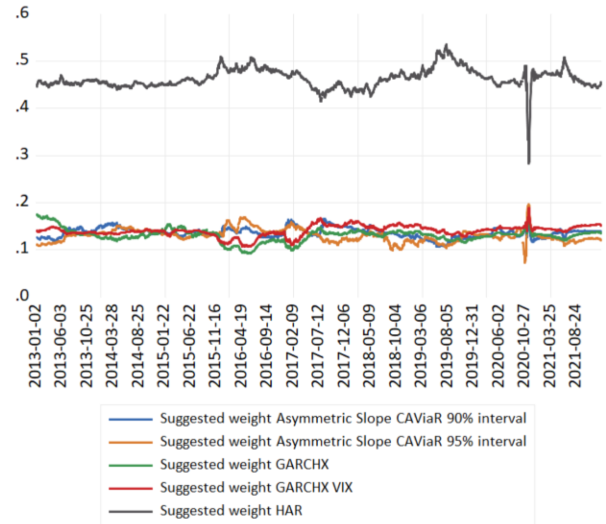


(b) Suggested weight of all models for 20-day.

Figure 13: Suggested weights for all models of JPM based on regression with only the two leading indicators. Weights of 10-day variance (left) and 20-day variance (right).

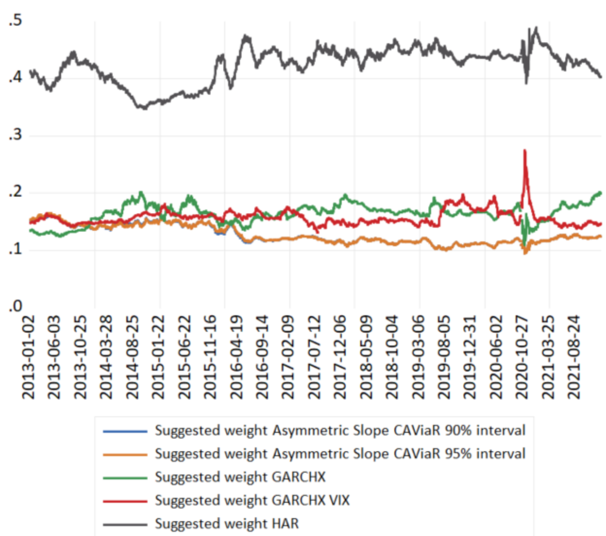


(a) Suggested weight of all models for 10-day.

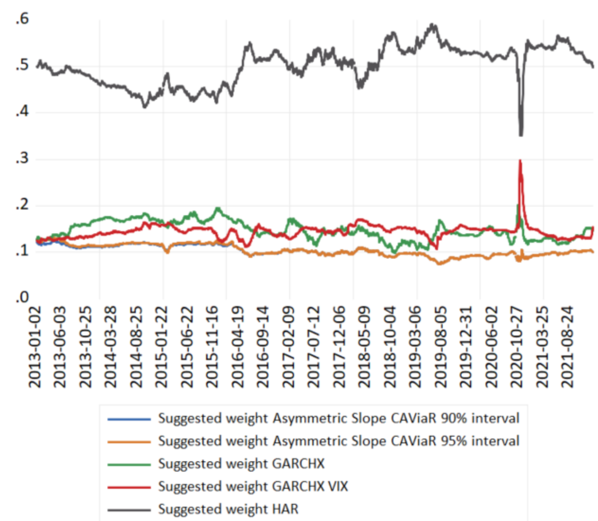


(b) Suggested weight of all models for 20-day.

Figure 14: Suggested weights for all models of MSFT based on regression with only the two leading indicators. Weights of 10-day variance (left) and 20-day variance (right).



(a) Suggested weight of all models for 10-day.

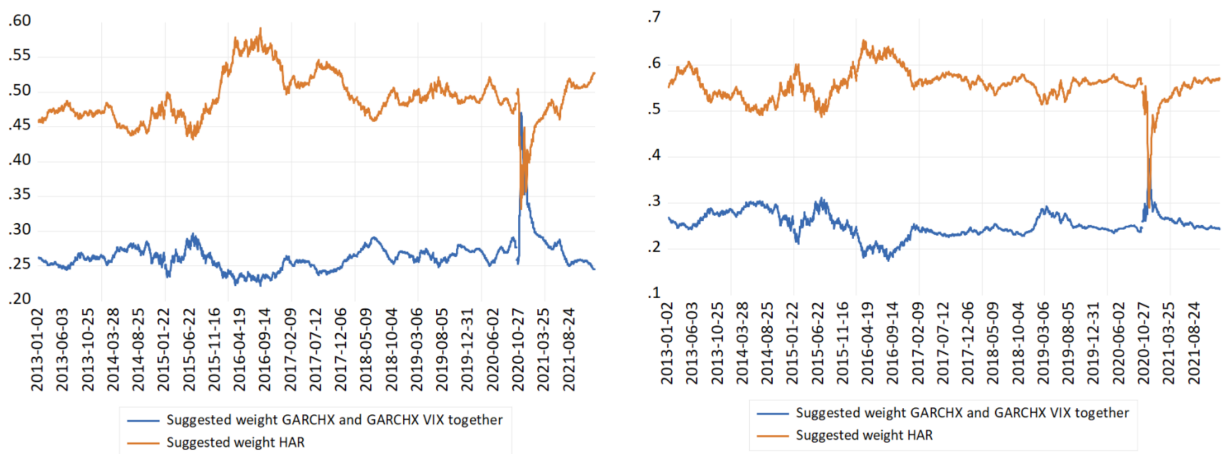


(b) Suggested weight of all models for 20-day.

Figure 15: Suggested weights for all models of NFLX based on regression with only the two leading indicators. Weights of 10-day variance (left) and 20-day variance (right).

B Weights both GARCHX models against HAR-RV

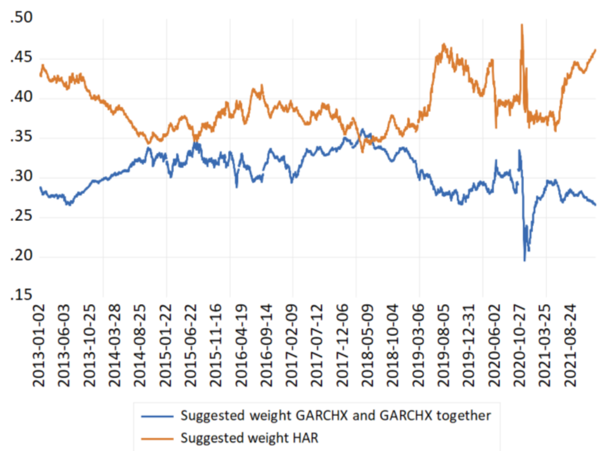
This appendix shows the suggested weights of both GARCHX models summed up against the suggested weights of the HAR-RV model of all individual stocks. These suggested weights are based on regressions with only the two leading indicators. Both suggested weights of the 10-day variance and the 20-day variance are shown for each stock.



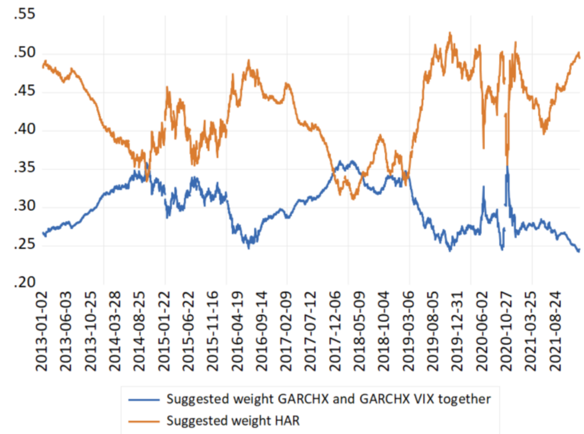
(a) Both GARCHX models against HAR-RV model for 10 day.

(b) Both GARCHX models against HAR-RV model for 20 day.

Figure 16: Suggested weights of both GARCHX models together against the suggested weights of the HAR-RV model of AAPL based on regression with only the two leading indicators. Weights corresponding to 10-day variance (left) and 20-day variance (right).

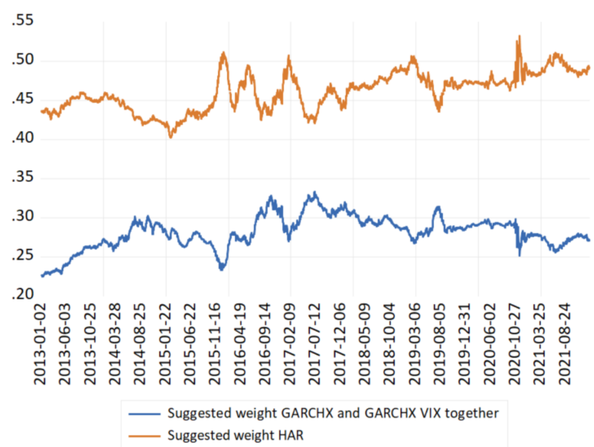


(a) Both GARCHX models against HAR-RV model for 10 day.

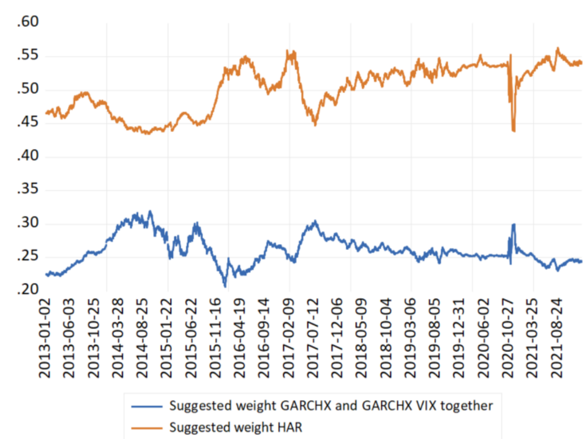


(b) Both GARCHX models against HAR-RV model for 20 day.

Figure 17: Suggested weights of both GARCHX models together against the suggested weights of the HAR-RV model of BAC based on regression with only the two leading indicators. Weights corresponding to 10-day variance (left) and 20-day variance (right).

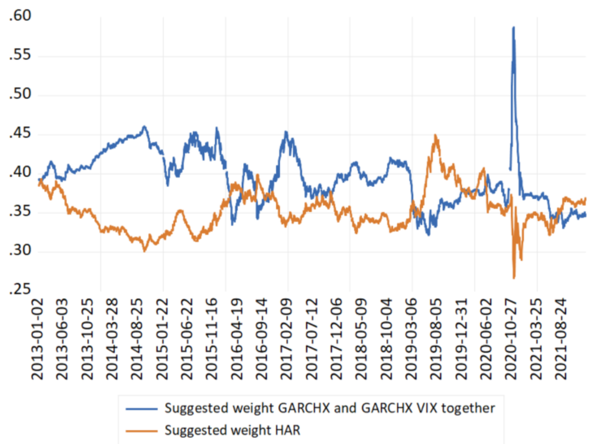


(a) Both GARCHX models against HAR-RV model for 10 day.

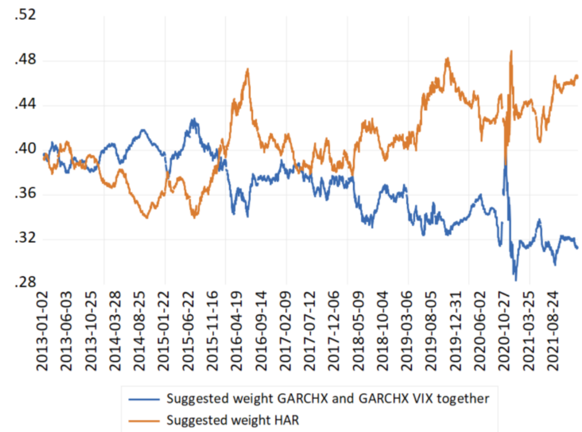


(b) Both GARCHX models against HAR-RV model for 20 day.

Figure 18: Suggested weights of both GARCHX models together against the suggested weights of the HAR-RV model of EBAY based on regression with only the two leading indicators. Weights corresponding to 10-day variance (left) and 20-day variance (right).

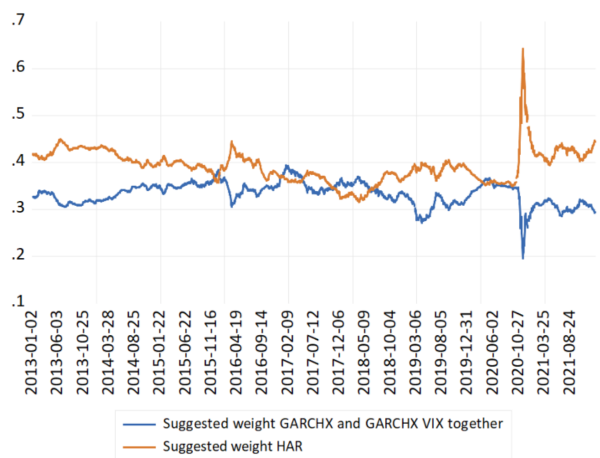


(a) Both GARCHX models against HAR-RV model for 10 day.

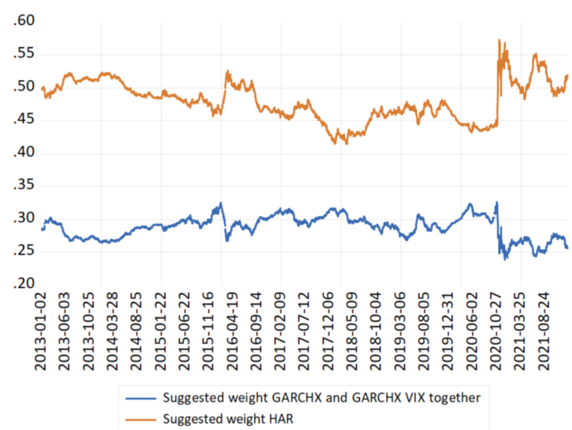


(b) Both GARCHX models against HAR-RV model for 20 day.

Figure 19: Suggested weights of both GARCHX models together against the suggested weights of the HAR-RV model of GS based on regression with only the two leading indicators. Weights corresponding to 10-day variance (left) and 20-day variance (right).

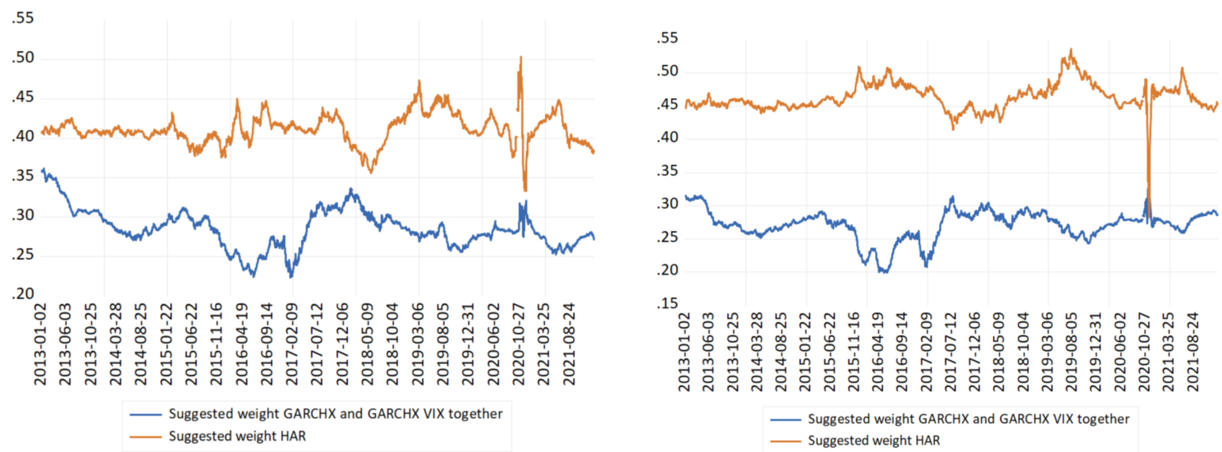


(a) Both GARCHX models against HAR-RV model for 10 day.



(b) Both GARCHX models against HAR-RV model for 20 day.

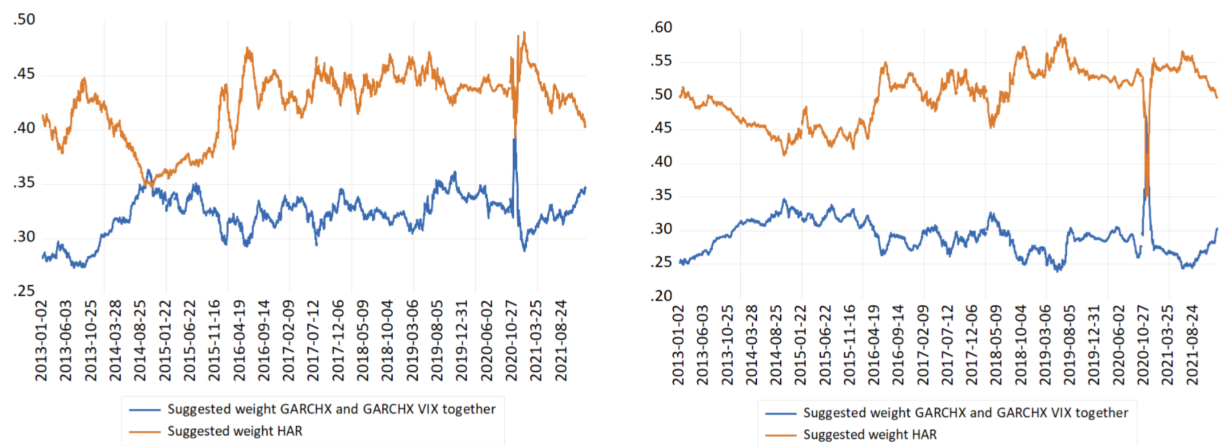
Figure 20: Suggested weights of both GARCHX models together against the suggested weights of the HAR-RV model of JPM based on regression with only the two leading indicators. Weights corresponding to 10-day variance (left) and 20-day variance (right).



(a) Both GARCHX models against HAR-RV model for 10 day.

(b) Both GARCHX models against HAR-RV model for 20 day.

Figure 21: Suggested weights of both GARCHX models together against the suggested weights of the HAR-RV model of MSFT based on regression with only the two leading indicators. Weights corresponding to 10-day variance (left) and 20-day variance (right).



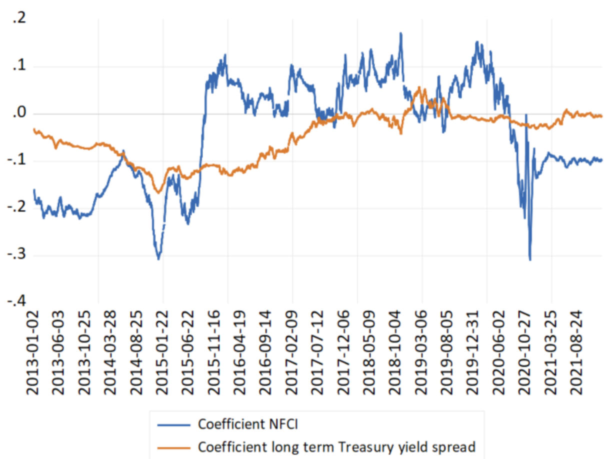
(a) Both GARCHX models against HAR-RV model for 10 day.

(b) Both GARCHX models against HAR-RV model for 20 day.

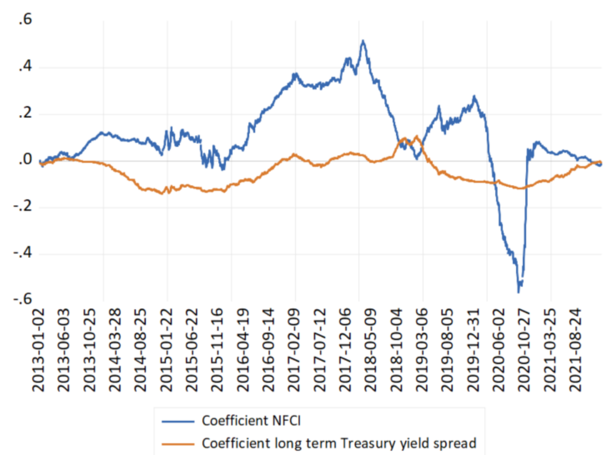
Figure 22: Suggested weights of both GARCHX models together against the suggested weights of the HAR-RV model of NFLX based on regression with only the two leading indicators. Weights corresponding to 10-day variance (left) and 20-day variance (right).

C Coefficients of leading indicators

This appendix shows plots of the coefficients of both leading indicators over time for AAPL, BAC, EBAY, GS, MSFT and NFLX. For all stocks, these coefficients belonging to each of the leading indicators are used to come up with a weight for the HAR-RV model, which is used in the forecast-error weighted pooled model.

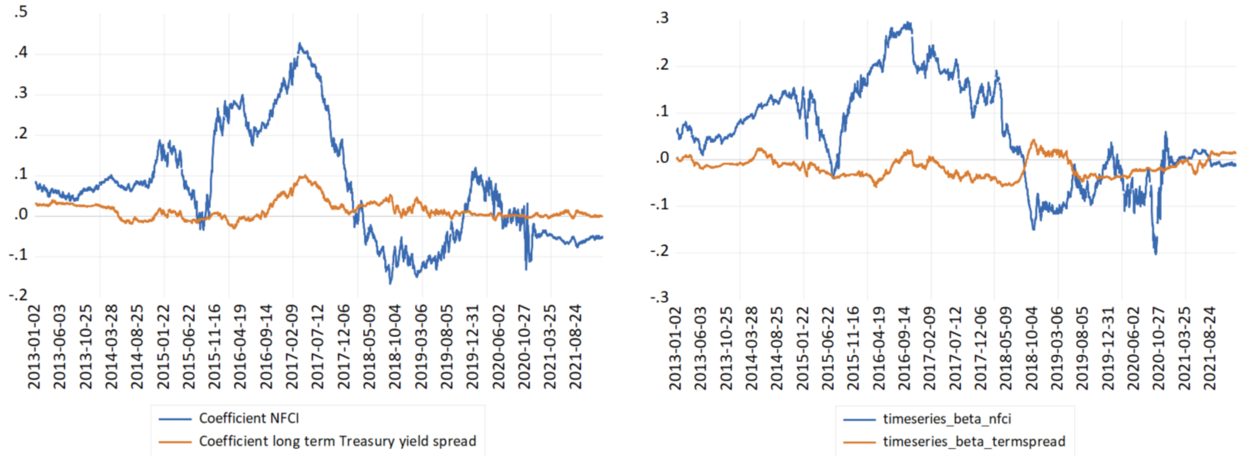


(a) Coefficients for AAPL.



(b) Coefficients for BAC.

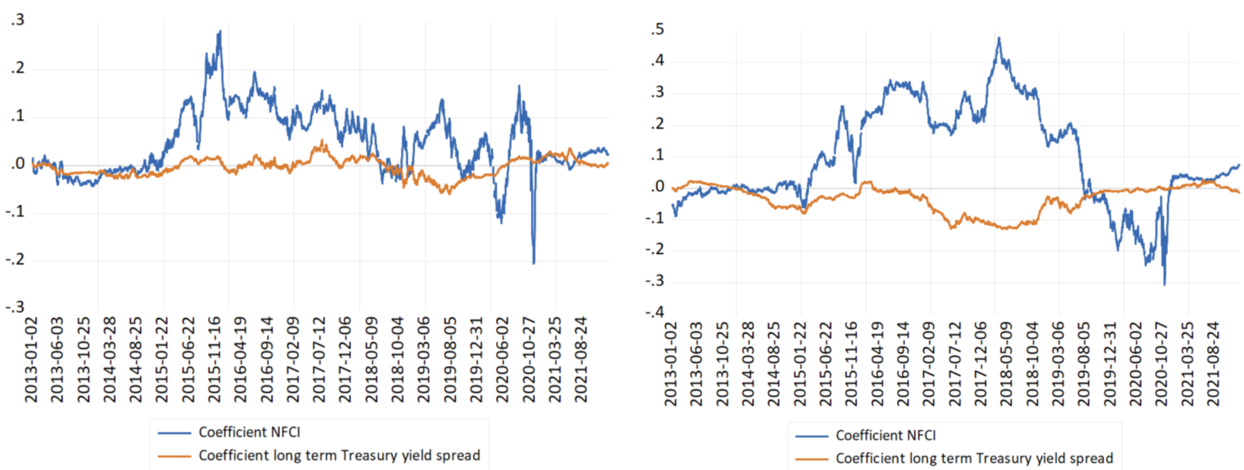
Figure 23: Coefficients of both leading indicators over time for AAPL and BAC. For these stocks, the coefficients belonging to each of the leading indicators are used to come up with a weight for the HAR-RV model.



(a) Coefficients for EBAY.

(b) Coefficients for GS.

Figure 24: Coefficients of both leading indicators over time for EBAY and GS. For these stocks, the coefficients belonging to each of the leading indicators are used to come up with a weight for the HAR-RV model.



(a) Coefficients for MSFT.

(b) Coefficients for NFLX.

Figure 25: Coefficients of both leading indicators over time for MSFT and NFLX. For these stocks, the coefficients belonging to each of the leading indicators are used to come up with a weight for the HAR-RV model.

D Short description of programming files

This appendix provides short descriptions of the programming files that are used during this thesis. The programming files are provided in a separate zip-file.

1. CAViaR models master thesis: provides the code which is used to get forecasts of the CAViaR models.
2. caviar: Optimization of the code of Engle and Manganelli (2004) which was transformed to a code in R-studio by Buczyński, Chlebus, et al. (2017), this was used as a base.
3. GARCHX model master thesis: provides the code which is used to get forecasts of the GARCHX model which uses the lagged multiple-day variance as exogenous variable.
4. GARCHX model vix master thesis: provides the code which is used to get forecasts of the GARCHX model which uses the VIX as exogenous variable.
5. HAR model master thesis: provides the code which is used to get forecasts of the HAR model.
6. Weighting the individual models: provides the code which is used to assign weights to each of the individual models based on the inverse of he absolute forecast error.
7. Regressions to get suggested weights: provides the code which is used to run regressions with all different set of regressors to come up with suggested weights for all individual models based on those regressors.
8. DM tests between pooled model and benchmark models: provides the code which is used to run DM tests between the forecast-error weighted pooled model and both benchmark models.
9. DM tests between pooled models: provides the code which is used to run DM tests between the forecast-error weighted pooled model which is based on only the two leading indicators and the forecast-error weighted pooled model which is based on the two leading indicators and other explanatory variables.