ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

MSc Econometrics and Management Science Quantitative Finance

Evaluating large dynamic covariance matrix and volatility forecasts for the S&P 500 constituents

Master's Thesis Date final version: September 13, 2021

Supervisor O. KLEEN Second assessor S.O.R. LÖNN Student Dani S. KLEIN ID number 480687

Abstract

In this research, I evaluate multiple dynamic one-month-ahead volatility and covariance matrix forecasts for the S&P 500 constituents in the period January 2000 — December 2020. Covariance matrix predictions are made using a model combining Dynamic Conditional Correlations and nonlinear shrinkage, and its improved version that exploits open-high-low-close price data. Furthermore, I propose Variance Forecast Shrinkage (VFS) and a picking forecast, which combine the theoretically optimal iterated and often practically optimal scaled myopic (GJR-)GARCH model forecasts. Out-of-sample empirical results show that risk-adjusted returns of GMV portfolios, based on the covariance forecasts, almost triple compared with a simple 1/N portfolio. However, the VFS and picking forecasts do not improve upon either iterated or scaled predictions, although they show that none of the latter two is superior.

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

Contents

1	Introduction	2							
2	Literature Review	3							
3	Data	5							
	3.1 Index performances	5							
	3.2 Constituent performances	7							
4	Methodology	9							
	4.1 DCC	9							
	4.2 Nonlinear shrinkage	11							
	4.3 DCC-NL	12							
	4.4 Regularized returns	13							
	4.5 Volatility proxies	14							
	4.5.1 Estimating overnight variance proportion	16							
	4.6 Covariance matrix forecasts	17							
	4.7 VFS	19							
	4.7.1 Predictive accuracy as uncertainty measure	21							
	4.7.2 Kurtosis as uncertainty measure	22							
	4.8 Performance measures	23							
5	Results	24							
	5.1 In-sample fits	24							
	5.2 Volatility forecasts	27							
	5.3 VFS1	28							
	5.4 Portfolio performances	29							
	5.5 VFS2	33							
6	Conclusion & Discussion	35							
Α	Optimal weight COHLC-proxy	39							
в	Portfolio performance <i>p</i> -values 40								

1 Introduction

Much research has been conducted on improving covariance matrix estimation and prediction, because it is useful in a wide variety of research fields. Focusing on Markowitz portfolio selection, one can create a Global Minimum Variance (GMV) portfolio, given an accurate forecast of the covariance matrix. Empirical research has namely shown that stocks with low variances tend to obtain higher (risk-adjusted) returns than more volatile equities (see e.g., Jagannathan and Ma (2003)). This phenomenon is also called the low-volatility anomaly. A GMV portfolio is therefore not only useful for investors that want to lower their risks. However, the intuitive and easy-to-calculate sample covariance matrix starts lacking when the considered stock universe becomes larger, due to the high number of parameters that need to be estimated. For this reason, Ledoit and Wolf (2004) propose linear shrinkage. In this procedure, a linear combination of the sample covariance matrix and a shrinkage target is taken as estimator, with weights that minimize a certain loss function.

Recently, a more useful method for larger dimensions, called nonlinear shrinkage, was proposed by Ledoit and Wolf (2012). This approach generalizes the idea of shrinking the sample covariance matrix by a common shrinkage intensity. It uses the spectral decomposition of the sample covariance matrix, and is developed in such a way that it yields a well-conditioned covariance matrix, even in case the dimension size exceeds the number of observations (N > T). Ledoit and Wolf (2017a) show that the theoretical advantage over linear shrinkage also holds empirically. Engle et al. (2019) combine the Dynamic Conditional Correlations (DCC) model of Engle (2002) with nonlinear shrinkage, creating the possibility to capture the time-dependence of large covariance matrices in the DCC-NonLinear shrinkage model (DCC-NL). In a comparison of several multivariate GARCH models, Engle et al. (2019) show that lower volatilities and higher Sharpe ratios than the 1/N portfolio are obtained when using the DCC-NL covariance matrix forecast in a GMV portfolio. Another striking result is that it significantly outperforms the portfolio of the DCC estimator.

This research focuses on evaluating large dynamic covariance matrix forecasts of S&P 500 constituent returns, and their use in GMV portfolios. Also, an attempt is made to improve variance predictions by combining iterated and scaled myopic forecasts. Ghysels et al. (2019) namely show that, although iterated GARCH forecasts are theoretically optimal, scaled GARCH forecasts tend to be more accurate in practice. Therefore, this paper proposes Variance Forecast Shrinkage (VFS) and a picking forecast. They determine optimal weights on the iterated and scaled forecasts by using historical predictive accuracies and estimated kurtoses as explanatory variables. In this way, ex ante and asset-dependent optimal weights are estimated at each point in time.

De Nard et al. (2020) even further improve the DCC-NL estimator by making use of Open-High-Low-Close (OHLC) price data. They exploit more efficient daily volatility proxies than the squared daily returns, all found under the assumption that stock prices follow a geometric Brownian motion with or without drift. These estimators were introduced by Parkinson (1980), Garman and Klass (1980), and Rogers and Satchell (1991). However, none of them use all available data to estimate the volatility in one day (considering close-to-close as a day): *both* the previous- and current-day close price, and the high, low, and open price. Therefore, I propose the COHLC-proxy, which *does* use all above values. It also stays similar to the COHL-proxy, the third proposed estimator by Garman and Klass (1980). Compared with this second most efficient proxy, a theoretical efficiency improvement of 12.1% is found for the COHLC-proxy. Furthermore, this research considers assetdependent estimation of the proportion of variance realized overnight, which is required for the proxy. It ensures that the daily volatility estimates of stocks with relatively different closed-market trading activities remain comparable.

In this research, empirical results are obtained using OHLC price data of all S&P 500 constituents over the period January 1995 — December 2020, where the first five years are used as a burn-in period. Sharpe ratios of GMV portfolios, based on the covariance forecasts, are shown to be approximately three times as high as for the 1/N portfolio. Also, GJR-GARCH seems to be of added value in-sample, while no improvement can be found in the out-of-sample forecasts, compared with regular GARCH forecasts. Furthermore, for OHLC-based volatility proxies, in-sample estimates of the proportion of variance realized overnight should be determined asset- and time-dependent. Large differences are namely found among the stock return variances, and an increasing pattern of relative trading activity during closed markets is found over time. I also show that the use of OHLC-based volatility proxies decreases portfolio volatility and increases Sharpe ratios significantly.

It is also shown that the VFS and picking forecasts do not seem to be more accurate than either iterated or scaled predictions. However, their implied weights show that none of the latter two is superior to one another in the total cross-section, or over the entire sample period. Furthermore, no significant performance differences can be found between the VFS forecasts based on the kurtosis and those based on the historical predictive accuracy. Lastly, one can conclude that there is no need to use a t-distribution instead of a normal distribution for the GARCH models.

This paper proceeds as follows. Section 2 discusses the literature relevant for this research. Section 3 elaborates on the data that is used. Section 4 explains the methods considered for covariance matrix and volatility prediction, and how they are evaluated. In Section 5, the results of the models and forecasts are given and interpreted. Section 6 concludes and discusses limitations of the research and possibilities for future research.

2 Literature Review

DeMiguel et al. (2009) focus on the empirical performance of a set of methods in terms of portfolio Sharpe ratios compared with the simple 1/N portfolio. They strikingly show that the 1/N portfolio is hard to outperform. However, they do not consider any dynamic models to forecast a next period's covariance matrix. Furthermore, they use low-frequency data (monthly) with arguably large moving windows, which also causes the methods to fail to capture the dynamic structures. Therefore, portfolios based on dynamic higher-frequency models might be able to outperform the 1/N portfolio.

Engle et al. (2019) account for conditional heteroskedasticity with the DCC model, and counteract the curse of dimensionality with nonlinear shrinkage. They show a predictive improvement over DCC in an empirical and simulation study, which becomes even larger as the asset space grows. Also, higher (risk-adjusted) GMV portfolio returns than the 1/N portfolio are obtained. With DCC-NL, nonlinear shrinkage is applied to the unconditional correlation matrix of the DCC model, where previously just the sample correlation matrix was used. Ledoit and Wolf (2020) derive an analytical method to determine the optimal individual shrinkage intensities in the nonlinear shrinkage method by using random matrix theory.

Hautsch et al. (2015) use intraday data of the constituents of the S&P 500 to construct GMV portfolios. They show that high-frequency data yields better covariance matrix estimators than low-frequency data. However, this conclusion is only drawn for horizons up to one month, making their research especially useful for short-term investors and traders. Andersen et al. (1999) also show that forecasts are improved when using high-frequency data for intraday to one-month horizons.

In a study to exploit OHLC price data, Parkinson (1980), Garman and Klass (1980) and Rogers and Satchell (1991) propose new volatility proxies that can theoretically improve the efficiency of the squared return proxy with factors up to 8.4. These improved proxies can be used to replace the squared return in the GARCH models, as shown by Chou (2005). De Nard et al. (2020) improve the DCC-NL estimator by using the OHLC data instead of only daily returns. They also use this intraday data to introduce so-called regularized returns, which are used for a more accurate correlation matrix estimator.

Engle et al. (2019) state that DCC-NL might also be improved by changing the GARCH model to a model that incorporates a leverage effect, for example. This allows the variance to react differently after upward or downward shocks. Hansen and Lunde (2005) compare the performance of 330 ARCH-type models using ten years of intraday International Business Machines Corp(IBM) return data. They find that out-of-sample performance is improved compared with GARCH when accommodating this effect. Therefore, this research also considers the GJR-GARCH model of Glosten et al. (1993).

Next to that, this research aims to improve variance forecasts by generalizing the idea that one should either use scaled myopic or iterated forecasts. Ghysels et al. (2019) show that a scaled forecast is preferred when considering one-month-ahead forecasts of the S&P 500 index. Yet, this finding does not need to hold for the individual equities within the index. For this reason, VFS is proposed. It uses a measure of forecast uncertainty to determine weights for an optimal linear combination of the scaled and iterated variance predictions.

3 Data

The data used in this research comes from the CRSP database, and contains daily information of all S&P 500 constituents over the period January 1995 — December 2020. This period is chosen because of the availability of open prices since 1995, which are necessary for OHLC-based models. Also, it contains periods of high overall volatility and recessions, yielding valid performance comparisons.

Because the index composition changes over time, the dataset does not contain 500 but 1162 stocks, with T = 6547 days of information each. Since it also includes the companies that have gone bankrupt in this period, there is no survivorship bias. Furthermore, at each point in time, only companies that are in the S&P 500 at that time are considered, to rule out other look-ahead biases. Another reason for this, is the fact that constituents of the S&P 500 have high market capitalizations by construction, which is often accompanied by high liquidity and data availability. Lastly, at each point in time, stocks that have not been traded for at least five years are left out. In this way, the considered methods can be estimated using a moving window of five years. It causes the average number of considered stocks to decrease by only $\pm 5\%$, to 476.

For each day-stock combination, the dataset contains the holding period return of that day, which is already adjusted for, e.g., dividends or stock splits. Furthermore, it consists of the open, high, low, and close prices, which are used to improve volatility proxies. Next to that, for the sake of completeness, delisting returns are retrieved to derive returns especially from companies that were still in the S&P 500 while being delisted.

The dataset also contains ten companies with two share classes, whose returns tend to be highly correlated in practice. Therefore, of each such pair, if simultaneously occurring, the stock with the shortest history in the S&P 500 is removed. If these numbers coincide, the stock with the highest average bid-ask spread is removed. This causes ten stocks to be fully removed from the data, while the other two pairs of stock do not trade simultaneously at any point in time. Eventually, N = 1152 stocks remain. Next to that, very rarely, an asset's return is missing on a specific date while being in the S&P 500, wherefore it is replaced by a zero. This only has a small impact, since the largest number of missing observations of a stock's return is just six.

As is common in practice, portfolios in this research are rebalanced monthly, which ensures that turnovers do not become too large. Furthermore, for simplicity, a month is assumed to be a period of 21 trading days. This causes the dataset to contain 311 full 'months', wherefore the last 16 observations are dropped.

3.1 Index performances

Figure 1a shows the cumulative returns of the equally-weighted (EW) and value-weighted (VW) S&P 500, while Figure 1b shows daily EW returns over time. One can also observe periods of recessions and high-volatility regimes by means of the shaded areas.



Figure 1: Simple returns of the EW and VW S&P 500 indices over January 1995 — December 2020. The blue-shaded areas represent U.S. recessions¹. The red-shaded areas represent high-volatility regimes from a bivariate Markov Switching model fitted to the EW returns. The purple-shaded areas are their intersects.

Firstly, the cumulative returns seem to be higher for EW than for VW. This slightly better performance is also reflected by the higher mean returns and Sharpe ratios for EW in multiple sub-periods, shown in Table 1. According to Plyakha et al. (2012), the relatively good performance of the EW S&P 500 index can be explained by its higher exposure to the well-known risk factors market, size, and value. Yet, they argue it can also be explained by the monthly rebalancing of the EW portfolio, which takes advantage of, e.g., reversal at a monthly frequency. DeMiguel et al. (2009) also find higher Sharpe ratios for EW than VW in all six of their considered datasets. Therefore, in this research, the 1/N portfolio is considered the benchmark. Although the 1/N and EW portfolio are similarly defined, this paper considers EW to be based on *all* S&P 500 stocks, and 1/N only on the stocks considered by the GMV portfolios at each point in time. Note that this causes the 1/N portfolio to consist out of only \pm 5% less stocks than EW. However, comparisons between 1/N and GMV portfolio performances remain valid because of it.

EW	Full period	Expansions	Recessions	Low-vol.	High-vol.
Ann. Mean %	13.675	14.138	10.080	17.813	-14.252
Ann. Vol. $\%$	20.470	16.621	39.083	15.087	41.315
Ann. Sharpe	0.668	0.851	0.258	1.181	-0.345
VW	Full period	Expansions	Recessions	Low-vol.	High-vol.
Ann. Mean %	11.908	13.098	2.662	15.928	-15.224
Ann. Vol. $\%$	19.131	16.385	33.512	15.030	36.168
Ann. Sharpe	0.622	0.799	0.079	1.060	-0.421

Table 1: Performance measures of EW and VW S&P 500 index over January 1995 — December 2020 and different distinctive time periods.

¹Coming from the National Bureau of Economic Research.

Secondly, the returns tend to be very low in the high-volatility regimes, especially during the global financial crisis from 2007 to 2009, which contains the most extreme daily returns in the dataset. In the corresponding recession, the EW cumulative return $(R_{\text{cum},t})$ went from 421.80% in December 2007 to 140.42% at its low in March 2009, indicating a loss of 53.92% in only 16 months.

3.2 Constituent performances

To understand the importance of diversification and low stock variances, this section discusses individual performances of the stocks in the dataset. In Figure 2a, one can see the distribution of average annualized returns of constituents while being in the S&P 500. The average is 10.54%, with a maximum of 1144.30%, belonging to the company Pet Inc., that performed well in January 1995, but was acquired on February 8, 1995. Since most high average annualized returns in the dataset are from companies that were in the S&P 500 briefly (or acquired quickly), the figure only shows values up to \pm 80%. From the companies that have been a constituent for the whole sample period, Apple Inc. has the highest average annualized return (26.48%). Furthermore, $\frac{299}{1152} = 25.95\%$ of the companies have a negative average return, while $\frac{737}{1152} = 63.98\%$ do not outperform the EW portfolio. However, for the Sharpe ratio, we see that $\frac{933}{1152} = 80.99\%$ do not outperform the EW portfolio. This finding underlines the importance of diversification, since the probability that a portfolio contains a stock like Apple Inc. becomes greater when expanding it.



Figure 2: Histograms of the annualized average returns and Sharpe ratios in the S&P 500-period of all the constituents over January 1995 — December 2020. The black lines represent the averages, while the red lines represent the EW portfolio.

A perfect example of the low-volatility anomaly can be found in Figure 3, which shows each stock's mean and volatility during their S&P 500-time plotted against each other. This anomaly states that stocks with lower volatilities tend to outperform stocks with higher volatilities. A simple linear regression shows us that the expected return on average decreases with the standard deviation of the return. To be more precise, the fitted model is $\hat{\mu}_i = 34.65 - 0.53\sigma_i$ for i = 1, ..., N. This means that a stock with volatility $\sigma = 30\%$ is expected to have an average annualized return of

 $\mu = 18.75\%$, while $\sigma = 50\%$ coincides with only $\mu = 8.15\%$. One can also see that the stocks with the highest Sharpe ratios are more often low-volatility stocks than stocks with high means.

As indicated by Figure 4, variances are often clustered, while the returns themselves are not. Therefore, one can predict variances much better. Together with the low-volatility anomaly, this suggests that it could pay off to forecast the covariance matrix of stock returns, and use it to find GMV portfolio weights, because a lower variance is expected to yield higher (risk-adjusted) returns.



Figure 3: Standard deviations of daily returns plotted against average returns during January 1995 — December 2020. Each circle represents a stock, and its size grows with the corresponding Sharpe ratio. The fitted regression line has an intercept and slope of respectively 34.65 and -0.53.



Figure 4: Sample autocorrelations of (squared) daily EW returns over January 1995 — December 2020, with blue lines representing 95% confidence intervals when no serial correlation is assumed.

4 Methodology

In this section, the DCC model and its estimation procedure are discussed first. Next, the concept behind and origin of nonlinear shrinkage are clarified. Subsequently, combining the latter two subsections, the DCC-NL model is given. Afterwards, the extension of the DCC-NL to the DCC-NL-OHLC model by means of OHLC price data and regularized returns is explained. Lastly, different forecast types are discussed, VFS is proposed, and the forecast performance measures are given.

4.1 DCC

The basis of all methods considered in this research is the DCC model, proposed by Engle (2002). It is a multivariate GARCH model that generalizes the Constant Conditional Correlations (CCC) model of Bollerslev (1990). Both models assume a univariate GARCH model (see Bollerslev (1986)) for all individual volatilities. Yet, where the CCC estimator assumes a constant correlation matrix over time, the DCC estimator allows for time-varying correlations.

To model the volatilities of returns, a GARCH(1,1) model without drift is often used in practice:

$$r_{i,t} = \sigma_{i,t}\epsilon_{i,t}, \quad \epsilon_{i,t} \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$$
(1)

$$\sigma_{i,t+1}^2 = \omega_i + a_i r_{i,t}^2 + b_i \sigma_{i,t}^2 \quad \text{for } i = 1, ..., N, \ t = 1, ..., T,$$
(2)

where $r_{i,t}$ is the return of asset *i* at time *t*, and $\sigma_{i,t}^2 = \operatorname{var}(r_{i,t}|I_{t-1})$ is its variance conditional on the full information set known at time t-1. The parameters of the model are restricted by $\omega_i > 0$, and $a_i, b_i \ge 0$. Also, for covariance stationarity, $a_i + b_i < 1$ is required.

However, in an application to financial stock returns of IBM,Hansen and Lunde (2005) show that GARCH can be improved upon by allowing for a leverage effect. This effect was noted by Black (1976), and it exploits the finding that positive return shocks have less effect on the volatility of a stock than similar-sized negative shocks. Multiple models exist that take this effect into account, but the best-known are GJR-GARCH (Glosten et al. (1993)), APARCH (Ding et al. (1993)) and EGARCH (Nelson and Cao (1992)). All three allow for a different effect of positive and negative lagged residuals. Because of its relatively good performance (e.g., see Brailsford and Faff (1996), Peters (2001)) and the fact that it preserves the interpretability of GARCH, this research also focuses on the GJR-GARCH model. This corresponds to the proposed Model 2 of Glosten et al. (1993), but assuming no drift in the underlying returns:

$$\sigma_{i,t+1}^2 = \omega_i + (a_i + \gamma_i I[r_{i,t} < 0])r_{i,t}^2 + b_i \sigma_{i,t}^2, \quad \text{for } i = 1, ..., N, \ t = 1, ..., T,$$
(3)

where $I[\cdot]$ represents an indicator function. The parameters are restricted by $\omega_i, a_i > 0, a_i + \frac{1}{2}\gamma_i > 0$ and $b_i \ge 0$. Now, for covariance stationarity, $a_i + \frac{1}{2}\gamma_i + b_i < 1$ is required. Note that $\gamma_i = 0$ yields GARCH, indicating that GJR-GARCH is a generalization of it, which can be very useful in testing it against GARCH. Presence of a leverage effect, as described above, would imply $\gamma_i > 0$, causing volatilities to respond heavier to negative stock innovations.

Either using GARCH or GJR-GARCH does not matter for the DCC model, since it models correlations. The cross-sectional dynamics are thus modelled separately from the individual stock returns. Given the conditional variances over time, a diagonal matrix $D_t := \text{Diag}(\sigma_{1,t}, \dots, \sigma_{N,t})$ is constructed for each time t, only containing the conditional standard deviations on the diagonal. They are also used to construct $\mathbf{s}_t := \left(\frac{r_{1,t}}{\sigma_{1,t}}, \dots, \frac{r_{N,t}}{\sigma_{N,t}}\right)'$, vectors of devolatized returns, which are used to construct correlation matrix innovation terms.

This research focuses on the DCC model as described in Engle et al. (2019), governing the evolution of the correlation matrix by the correlation targeting idea of Engle (2009):

$$Q_{t+1} = (1 - \alpha - \beta)C + \alpha \mathbf{s}_t \mathbf{s}'_t + \beta Q_t, \tag{4}$$

where C is the unconditional correlation matrix of the returns, and thus the covariance matrix of \mathbf{s}_t . Q_t is a pseudo-correlation matrix, which has diagonal elements slightly different from 1. Yet, a correctly defined correlation matrix can be obtained by:

$$R_t := \text{Diag}(Q_t)^{-1/2} Q_t \text{Diag}(Q_t)^{-1/2}.$$
(5)

Using the diagonal matrix D_t and correlation matrix R_t , the covariance matrix becomes:

$$\Sigma_t := D_t R_t D_t. \tag{6}$$

Note that the resulting covariance matrix Σ_t is positive definite (PD). Q_t is PD because it is a weighted sum of C, which is PD, and positive semi-definite weighted sum of $\mathbf{s}_t \mathbf{s}'_t$. Therefore, R_t is too, while D_t is a strictly positive diagonal matrix, and thus also PD. In total, Σ_t eventually is PD because D_t and R_t are too. This is important, because it ensures that Σ_t is invertible, and its inverse is needed when determining GMV portfolio weights.

To estimate the model, Engle et al. (2019) use a so-called composite likelihood method, proposed by Pakel et al. (2020). Instead of considering the full likelihood, they combine likelihoods of certain pairs of assets to ease computation. They have also shown consistency of the estimators in the DCC model. Engle et al. (2019) use the proposed 2MSCLE method, which maximizes likelihoods of a limited number of pairs, in this case, all contiguous pairs. This changes the number of calculations to be made from $O(N^2)$ for all pairs to O(N) for only contiguous pairs, where N is the asset space. Pakel et al. (2020) also show that their performances are approximately equal for a large N, wherefore the 2MSCLE approach is considered in this research too.

Altogether, fitting GARCH or GJR-GARCH, devolatized returns \mathbf{s}_t and volatility forecasts

 $\hat{\sigma}_{i,T+1}$ are obtained first. Afterwards, unconditional correlation matrix $C = \operatorname{cov}(\mathbf{s}_t)$ is estimated. Lastly, the DCC model is fitted by maximizing the composite likelihood to eventually obtain covariance matrix forecast $\hat{\Sigma}_{T+1}$.

4.2 Nonlinear shrinkage

To understand how nonlinear shrinkage works, an explanation of linear shrinkage for covariance matrices is useful. It was proposed by Ledoit and Wolf (2004), and based on the shrinkage estimation of Stein (1956) and James and Stein (1961) for the mean vector. The idea is that a weighted sum of the sample covariance matrix and a certain shrinkage target is taken as estimator:

$$\Sigma^* = \gamma^* \Gamma + (1 - \gamma^*) S,\tag{7}$$

where Γ denotes the target matrix, S the sample covariance matrix, and γ^* is called the shrinkage intensity. The target matrix can be set equal to a scaled identity matrix as in Ledoit and Wolf (2004), or to a factor model-implied covariance matrix, for example. However, the analytical effort comes in with determining the optimal shrinkage intensity γ^* . Although each type of target matrix requires a different way of computation, the goal is always to minimize the MSE function of the difference between the estimated and real covariance matrix: $\hat{\Sigma} - \Sigma$. For matrices, this function is the Frobenius loss function (scaled by $\frac{1}{N}$ for convenience): $||A||_F^2 = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N A_{ij}^2$. The case of a scaled identity matrix as target ($\Gamma = \mu I$) is obtained when only allowing for an estimator that is a convex linear combination of the identity and sample covariance matrix. This gives the following minimization problem:

$$\min_{\rho_1,\rho_2} \mathbb{E}\left[\left\| \widehat{\Sigma} - \Sigma \right\|_F^2 \right]$$
(8)

s.t.
$$\widehat{\Sigma} = \rho_1 I + \rho_2 S,$$
 (9)

where filling in the optimal ρ_1 and ρ_2 turns out to yield

$$\Sigma^* := \frac{\beta^2}{\delta^2} \mu I + \frac{\alpha^2}{\delta^2} S,\tag{10}$$

with α^2 , β^2 , δ^2 and μ being functions of Σ and S, whose values can be estimated using the data. Eventually, the optimal covariance matrix estimator is obtained by:

$$\widehat{\Sigma}^* = \gamma^* \widehat{\mu} I + (1 - \gamma^*) S. \tag{11}$$

Ledoit and Wolf (2004) note that $\hat{\Sigma}^*$ can also be expressed using the spectral decomposition:

$$\widehat{\Sigma}^* = U\Delta^* U', \text{ where } \Delta^* := \text{Diag}(\delta_1^*, ..., \delta_N^*), \\ \delta_i^* := \gamma^* \mu + (1 - \gamma)\lambda_i.$$
(12)

Here, λ_i represents the *i*-th sample eigenvalue, while U represents the matrix of eigenvectors. This implies that linear shrinkage towards μI is the same as shrinking the eigenvalues of Σ towards μ , where μ turns out to equal the grand mean of these eigenvalues.

Nonlinear shrinkage, as proposed by Ledoit and Wolf (2012), works similar, in a way that it also shrinks the eigenvalues. However, it does not use a common shrinkage intensity γ , but an individual intensity for each sample eigenvalue. This is equivalent to adjusting the sample eigenvalue upward or downward, as long as the optimal eigenvalues δ_i^* stay positive, ensuring the positive definiteness of $\hat{\Sigma}^*$. Using findings from a discipline called random matrix theory, Ledoit and Wolf (2012) find an oracle shrinkage function $\phi(\lambda_i)$, which transforms a sample eigenvalue λ_i into an optimal estimator of the true eigenvalue. Although Ledoit and Wolf (2017b) come up with the QuEST-function, which estimates ϕ , Ledoit and Wolf (2020) derive an analytical formula. This makes nonlinear shrinkage much faster and straightforward than the indirect approach needed with the QuEST-function. For a detailed description of the analytical nonlinear shrinkage, the reader is referred to Ledoit and Wolf (2020), or Section 4.2 of Ledoit and Wolf (2019) for a summary.

4.3 DCC-NL

Instead of just using $C = cov(\mathbf{s}_t)$, Engle et al. (2019) use nonlinear shrinkage of Ledoit and Wolf (2012) to estimate the unconditional correlation matrix, yielding the DCC-NL model. To show why this estimate is important in determining conditional correlation matrices, notice that rewriting Equation (4) gives:

$$Q_{t+1} = (1 - \alpha - \beta)C\sum_{k=0}^{t} \beta^k + \alpha \sum_{k=0}^{t} \beta^k \mathbf{s}_k \mathbf{s}'_k + \beta^t C$$
(13)

$$\rightarrow \frac{1 - \alpha - \beta}{1 - \beta} C + \alpha \sum_{k=0}^{\infty} \beta^k \mathbf{s}_k \mathbf{s}'_k \text{ as } t \rightarrow \infty$$
(14)

where $Q_0 = C$ is assumed for the sake of simplicity. One could consider this limit to be a weighted sum of the unconditional correlation and devolatized covariance matrix innovations, with weights adding up to $\frac{1-\alpha-\beta}{1-\beta} + \frac{\alpha}{1-\beta} = 1$. In an application to daily returns of the 500 most liquid stocks in the CRSP database from 2005 through 2009, Engle et al. (2019) find estimates of $\hat{\alpha} = 0.0490$ and $\hat{\beta} = 0.9297$ in their DCC-NL model. This indicates that a forecast of the next day's correlation matrix has a weight of $\frac{1-\hat{\alpha}-\hat{\beta}}{1-\hat{\beta}} = 0.303$ on C. Therefore, the estimate of the unconditional correlation matrix is important for the forecasts, especially when correlations are less persistent.

Although the sample correlation matrix is often used as an estimator for C, it starts lacking for large stock universes compared to the sample size (high N/T). Therefore, Engle et al. (2019) nonlinearly shrink it, as described in Section 4.2. Yet, they still use the QuEST-function, while this research implements the analytical nonlinear shrinkage formula of Ledoit and Wolf (2020).

4.4 Regularized returns

To improve the DCC-NL model even further, De Nard et al. (2020) propose a way in which data of a higher frequency can be exploited. They specifically focus on the use of OHLC price data, which compresses much intraday data into four simple, yet powerful observations. First, they intend to improve upon the very noisy squared returns as a volatility proxy in the GARCH model, as proposed by Molnár (2016). However, doing this only improves the conditional variance forecast, while correlations might also be improved using OHLC data. Therefore, De Nard et al. (2020) introduce so-called regularized returns, which can be used for both purposes. The procedure boils down to changing the actual returns to a function of the volatility proxy, with the sign of the return remaining intact. In this way, correlation estimates can be made that are based on intraday data too.

To understand the definition of the regularized return, first note that the actual return of an asset at a certain time can be written as:

$$r_{i,t} = \operatorname{sign}(r_{i,t}) \sqrt{r_{i,t}^2}.$$
(15)

The first considered change regards the volatility proxy $r_{i,t}^2$, which is turned into an improved proxy $\hat{v}_{i,t}$, to obtain a naive version of the regularized return:

$$\tilde{r}_{i,t}^{\text{naive}} := \operatorname{sign}(r_{i,t}) \sqrt{\hat{v}_{i,t}}.$$
(16)

However, De Nard et al. (2020) also show that changing the sign-function to a 'scaled' hyperbolic tangent function yields better results in practice:

$$\tilde{r}_{i,t} := \operatorname{stanh}(r_{i,t},\kappa)\sqrt{\hat{v}_{i,t}} = \frac{e^{\kappa r_{i,t}} - 1}{e^{\kappa r_{i,t}} + 1}\sqrt{\hat{v}_{i,t}}.$$
(17)

The stanh-function ensures that very small differences of $r_{i,t}$ from 0 are made less extreme compared to the sign-function. For returns in decimals, De Nard et al. (2020) $\kappa = 10,000$ is approximately optimal in practice, which only shrinks the sign-function for returns within a range of 5 bps from 0. This is somewhat equal to the average transaction cost for highly liquid U.S. stocks nowadays.

Given the regularized returns, the DCC-NL model in Section 4.3 can be estimated using devolatized returns $\mathbf{s}_t := \left(\frac{\tilde{r}_{1,t}}{\sigma_{1,t}}, \cdots, \frac{\tilde{r}_{N,t}}{\sigma_{N,t}}\right)'$. From now on, this OHLC-based model will be referred to as DCC-NL-OHLC.

4.5 Volatility proxies

One question remains, which is how to make use of the OHLC data to obtain the best possible volatility proxy $\hat{v}_{i,t}$ in terms of efficiency and unbiasedness. Because it sometimes occurs that OHLC observations are missing for a specific date-asset combination, different proxies are used in various cases. However, missing values of OHLC prices in the considered dataset occur when either only the open price is missing, or all prices are missing. The latter usually happens around mergers, acquisitions, and S&P 500 constituent transitions.

This research mostly uses $\hat{v}_{i,t}^{COHLC}$, but turns to $\hat{v}_{i,t}^{CHLC}$ when the open price is missing. When all values are missing, the squared return is used, which logically equals zero, since missing returns are replaced by zeros. To explain the proxies mentioned above, $\hat{v}_{i,t}^{CC}$, $\hat{v}_{i,t}^{COC}$, and $\hat{v}_{i,t}^{HL}$ are first introduced.

All volatility proxies described are based on the strict assumption that, every stock price follows a continuous-time geometric Brownian motion with variance σ^2 on a certain full day. This means that the variance is assumed to be constant during a day, irrespective of whether the market is closed or open. Also, by the Brownian motion property of independent increments, the log-return of a stock between times $0 \le t_0 < t_1 \le 1$ within that day, follows a normal distribution:

$$\log(1 + r_{t_0, t_1}) = \log(P_{t_1}) - \log(P_{t_0}) \sim N(\mu, (t_1 - t_0)\sigma^2).$$
(18)

The benchmark proxy is only based on the close prices, as normally used in GARCH. However, instead of using the squared simple return, compounded returns are considered because of the Brownian motion-assumption. Note that this makes little difference numerically, since daily returns are close to zero. Eventually, the estimator is:

$$\hat{v}_{i,t}^{CC} := [\log(1+r_{i,t})]^2.$$
(19)

In their research to exploiting OHLC data, Garman and Klass (1980) come up with different ideas. Yet, they are all partly routed in decomposing the variance into two independent parts:

$$\log(1+r_{i,t}) = \log\left(\frac{o_{i,t}}{\tilde{c}_{i,t-1}}\right) + \log\left(\frac{c_{i,t}}{o_{i,t}}\right),\tag{20}$$

where $o_{i,t}$ and $c_{i,t}$ represent the open and close price of asset *i* at time *t*, respectively. $\tilde{c}_{i,t-1}$ is the close price of the previous day, but adjusted for, e.g., dividend pay-outs or stock splits:

$$\tilde{c}_{i,t-1} := \frac{c_{i,t-1}}{1+r_{i,t-1}}.$$
(21)

Note that this can be done because $r_{i,t-1}$ was already adjusted for all types of corporate action.

The first improving proxy of Garman and Klass (1980) is based on the decomposition in Equation

(20), and only uses the open and close prices:

$$\hat{v}_{i,t}^{COC} := \frac{1}{2f} \left[\log \left(\frac{o_{i,t}}{\tilde{c}_{i,t-1}} \right) \right]^2 + \frac{1}{2(1-f)} \left[\log \left(\frac{c_{i,t}}{o_{i,t}} \right) \right]^2, \tag{22}$$

where f represents the proportion of variance realized when the market is closed. Because of the much higher trading volume during the day, this proportion does not equal the physical time proportion of closed markets. In other words, the assumed to be constant variance σ^2 in Equation (18) is different for closed and open markets in practice. However, setting f to the realized variance proportion ensures that σ^2 can still be assumed to be constant. Estimation of f is described in Section 4.5.1, but one can assume it is known for now.

Garman and Klass (1980) show that $\hat{v}_{i,t}^{COC}$ is twice as efficient as $\hat{v}_{i,t}^{CC}$, since $\operatorname{Var}(\hat{v}_{i,t}^{COC}) = \sigma^4$, while $\operatorname{Var}(\hat{v}_{i,t}^{CC}) = 2\sigma^4$. One caveat of the proxy is the fact that it assumes zero drift, meaning $\mu = 0$ in Equation (18). Although it is plausible for daily returns, volatilities of stocks in a bull or bear market might be overestimated.

In a study to the connection between high-low price ranges and an asset's volatility, Parkinson (1980) finds a proxy that only uses the high and low price:

$$\hat{v}_{i,t}^{HL} := \frac{1}{4\log(2)} \left[\log\left(\frac{h_{i,t}}{l_{i,t}}\right) \right]^2, \tag{23}$$

where $h_{i,t}$ and $l_{i,t}$ represent the high and low price of asset *i* in day *t*, respectively. This formula is derived using random-walk mathematics, and it is based on a distribution function of range $\log\left(\frac{h_{i,t}}{l_{i,t}}\right)$, found by Feller (1951). Parkinson (1980) derives that $\mathbb{E}\left(\left[\log\left(\frac{h_{i,t}}{l_{i,t}}\right)\right]^2\right) = 4\log(2)\sigma^2$ considering highs and lows of an entire day. Garman and Klass (1980) show that $\operatorname{Var}(\hat{v}_{i,t}^{HL}) \approx 0.385\sigma^4$, which means $\hat{v}_{i,t}^{HL}$ is approximately 5.2 times more efficient than $\hat{v}_{i,t}^{CC}$. The proxy and this theoretical efficiency, however, are based on some arguable assumptions: a zero drift, as well as the absence of overnight price jumps are assumed. This could potentially cause over- and underestimation, respectively. Yet, using it as a proxy for the variance realized only during open markets, only the zero-drift assumption needs to be made. And just as with other proxies based on questionable assumptions, the estimator still tends to improve upon $\hat{v}_{i,t}^{CC}$ (Chou et al. (2010)).

However, Yang and Zhang (2000) show that a better variance estimator is suggested by Rogers and Satchell (1991), because they also consider the close (and open) price. Furthermore, it has the property that it remains unbiased under non-zero drift, unlike $\hat{v}_{i,t}^{HL}$. It is defined as follows:

$$\hat{v}_{i,t}^{OHLC} := \log\left(\frac{h_{i,t}}{o_{i,t}}\right) \log\left(\frac{h_{i,t}}{c_{i,t}}\right) + \log\left(\frac{o_{i,t}}{l_{i,t}}\right) \log\left(\frac{c_{i,t}}{l_{i,t}}\right) \tag{24}$$

$$\hat{v}_{i,t}^{CHLC} := \log\left(\frac{h_{i,t}}{\tilde{c}_{i,t-1}}\right) \log\left(\frac{h_{i,t}}{c_{i,t}}\right) + \log\left(\frac{\tilde{c}_{i,t-1}}{l_{i,t}}\right) \log\left(\frac{c_{i,t}}{l_{i,t}}\right).$$
(25)

Rogers and Satchell (1991) show that $\operatorname{Var}(\hat{v}_{i,t}^{OHLC}) \approx 0.331\sigma^4$, wherefore it is even slightly more efficient than $\hat{v}_{i,t}^{HL}$. Also, notice that open prices *can* be used in proxy $\hat{v}_{i,t}^{OHLC}$. But assuming there are no overnight jumps, Rogers, Satchell, and Yoon (1994) set $o_{i,t} = \tilde{c}_{i,t-1}$ to obtain $\hat{v}_{i,t}^{CHLC}$, which can be used in absence of the open price. However, there is a possibility that the previous day close price is even higher or lower than the current day high or low, respectively. Knowing the open price must be in range $[l_{i,t}, h_{i,t}]$, the assumed to be open price would be theoretically impossible, and could cause a negative volatility proxy. To overcome this problem, I set $\tilde{c}_{i,t-1} = l_{i,t}$ when $\tilde{c}_{i,t-1} < l_{i,t}$, and $\tilde{c}_{i,t-1} = h_{i,t}$ when $\tilde{c}_{i,t-1} > h_{i,t}$.

Replacing the latter part of $\hat{v}_{i,t}^{COC}$ by $\hat{v}_{i,t}^{HL}$, Garman and Klass (1980) find a range-based estimator that also considers the open and close price:

$$\hat{v}_{i,t}^{COHL} := \frac{w}{f} \left[\log \left(\frac{o_{i,t}}{\tilde{c}_{i,t-1}} \right) \right]^2 + \frac{1-w}{1-f} \hat{v}_{i,t}^{HL},$$
(26)

with $\operatorname{Var}(\hat{v}_{i,t}^{COHL}) \approx 0.323\sigma^4$. However, note that it does *not* involve the close price at time *t*. Also, as argued before, the estimator of Parkinson (1980) is theoretically suboptimal to the estimator of Rogers and Satchell (1991). Therefore, a logical improvement over this proxy would be to replace the $\hat{v}_{i,t}^{HL}$ part by $\hat{v}_{i,t}^{OHLC}$, forming:

$$\hat{v}_{i,t}^{COHLC} := \frac{w}{f} \left[\log \left(\frac{o_{i,t}}{\tilde{c}_{i,t-1}} \right) \right]^2 + \frac{1-w}{1-f} \hat{v}_{i,t}^{OHLC},$$
(27)

For $\hat{v}_{i,t}^{COHL}$, Garman and Klass (1980) argue that setting w = 0.17 would be theoretically optimal. However, the value of w for which $\hat{v}_{i,t}^{COHLC}$ has the lowest variance is w = 0.14 (see Appendix A). This yields $\operatorname{Var}(\hat{v}_{i,t}^{COHLC}) \approx 0.284\sigma^4$, theoretically improving the efficiency of $\hat{v}_{i,t}^{COHL}$ by 12.1%, and of $\hat{v}_{i,t}^{CC}$ by no less than 85.8% (7.0 times more efficient). Furthermore, by using $\hat{v}_{i,t}^{OHLC}$ instead of $\hat{v}_{i,t}^{HL}$ for the second part, the assumption of a zero drift is let go.

Although Garman and Klass (1980) propose even more extensive estimators, Molnár (2016) shows that their theoretically possible improvements over $\hat{v}_{i,t}^{COHL}$ are hard to be realized in practice. This is probably mainly due to the heavier exploitation of some assumptions. Yet, $\hat{v}_{i,t}^{COHLC}$ is not based on the assumptions of zero drift or the absence of overnight jumps, and it considers more information, wherefore it could be a practical improvement over $\hat{v}_{i,t}^{COHL}$ as well.

4.5.1 Estimating overnight variance proportion

Based on historical data, Yang and Zhang (2000) argue that f = 0.25 typically yields a good fit for U.S. stocks when using daily data. However, this is a generalization of their finding that it depends on the stock and that it can differ somewhere between 0.18 and 0.30. Therefore, this research considers an asset-dependent f_i , which can be determined ex ante by estimating proportions of

return variances realized overnight. To determine the optimal values, Yang and Zhang (2000) use the fact that, theoretically,

$$\left(\frac{1}{f_i} - 1\right) \frac{V_i^c}{V_i^o} \sim F(T - 1, T - 1),$$
 (28)

where T denotes the number of observations used, and

$$V_i^o := \operatorname{Var}_i \left[\log \left(\frac{o_{i,t}}{\tilde{c}_{i,t-1}} \right) \right] \quad \text{and} \quad V_i^c := \operatorname{Var}_i \left[\log \left(\frac{c_{i,t}}{o_{i,t}} \right) \right], \tag{29}$$

with $\operatorname{Var}_i[\cdot]$ representing the variance for asset *i* over period 1, ..., *T*. Although Yang and Zhang (2000) do not go into further detail, a logical way to find the optimal f_i is by using the expected value of an F-distributed random variable:

$$\mathbb{E}\left[\left(\frac{1}{f_i}-1\right)\frac{V_i^c}{V_i^o}\right] = \frac{T-1}{T-3} \Leftrightarrow f_i = \left[1+\frac{T-1}{T-3} \cdot \left[\mathbb{E}\left(\frac{V_i^c}{V_i^o}\right)\right]^{-1}\right]^{-1}.$$
(30)

Next, one could estimate f_i by estimating the expected open-close variance ratio with $\frac{V_i^c}{\hat{V}_i^o}$:

$$\hat{f}_{i} = \left[1 + \frac{T-1}{T-3} \cdot \left(\frac{\hat{V}_{i}^{o}}{\hat{V}_{i}^{c}}\right)\right]^{-1} = \frac{\hat{V}_{i}^{c}}{\hat{V}_{i}^{c} + \frac{T-1}{T-3}\hat{V}_{i}^{o}}, \text{ where } \hat{f}_{i} \to \frac{\hat{V}_{i}^{c}}{\hat{V}_{i}^{c} + \hat{V}_{i}^{o}} \text{ as } T \to \infty.$$
(31)

Therefore, using a large enough sample size, one could estimate f_i by the fraction of the sample variance during closed markets over the total variance. Notice that this matches the definition of f given before.

4.6 Covariance matrix forecasts

Because this research considers portfolios held for longer than one period, a certain approach different from using one-step-ahead forecasts is necessary. Not much research has been conducted in which different multi-step-ahead forecasts are compared, especially not for multivariate covariance or correlation forecasts. However, there are studies that have compared the best-known types of forecasts for longer horizons. They concern scaled, iterated, and direct forecasts.

A scaled (myopic) forecast is a naive type of forecast, since it only predicts one step ahead and assumes that it is also a good prediction for the subsequent periods. It is then just scaled by the number of periods in the forecast horizon. Despite its naivety, it is often used in practice because of its simplicity and sufficient accuracy. Iterated forecasts are the theoretically correct predictions to use, in the sense of being unbiased. A model-based prediction is made of every step in the considered forecast horizon, after which they are accumulated to one period's prediction. Direct forecasts are somewhat different from iterated and scaled forecasts since they require a different model. For example, if a four-step-ahead direct forecast is preferred, one needs to model the considered time series against information known four steps before, instead of just one, as usual.

In an application to variances of multiple asset classes, Ghysels et al. (2019) study these different multi-step-ahead forecasts made using GARCH-type, but also Realized Volatility (RV) and Mixed Data-Sampling (MIDAS) models. They compare the forecast accuracy of direct, iterated and scaled myopic variance forecasts. For four different stock market indices and five different forecast horizons, they find various types of models to yield the best fit. The best in-sample GARCH forecasts for the S&P 500, with a horizon of 22 days, are given by the iterated model. This model also is most often optimal among different asset classes and horizons. Out-of-sample however, the scaled GARCH yields the most efficient forecasts, and the iterated GARCH is never optimal for any asset class and horizon. Also considering the RV and MIDAS models, they still find that the scaled GARCH is optimal for the S&P 500 and a 22-day forecast horizon.

Yet, one could doubt whether their findings also hold for more volatile returns than those of a stock index, which is very diverse. Another interesting question is whether scaled forecasts perform best in a multivariate setting too, considering covariance matrices that are forecasted multiple periods ahead. This research only aims to improve the variance forecasts and keep the correlation forecasts simple. This choice is grounded in the decomposition of a covariance matrix forecast:

$$\widehat{\Sigma}_{t+1:t+h} = \widehat{D}_{t+1:t+h} \widehat{R}_{t+1:t+h} \widehat{D}_{t+1:t+h}.$$
(32)

To make sure that $\widehat{\Sigma}_{t+1:t+h}$ stays PD, and thus invertible, one can change the volatilities (the diagonal of $\widehat{D}_{t+1:t+h}$) freely. Yet, $\widehat{R}_{t+1:t+h}$ is tightly constrained, since it must still be PD too. Therefore, I focus on improving the variance instead of correlation forecasts in this research.

The correlation matrix forecast is set equal to the simple myopic forecast, but because it concerns correlations instead of variances, there is no need to scale it. Therefore, the forecast becomes $\hat{R}_{t+1:t+h} = \hat{R}_{t+1}$. Now, each entry (i, j) of $\hat{\Sigma}_{t+1:t+h}$ is described by

$$\sqrt{\widehat{\operatorname{Var}}(r_{i,t+1:t+h})}\sqrt{\widehat{\operatorname{Var}}(r_{j,t+1:t+h})}\hat{R}_{i,j,t+1}.$$
(33)

For both variance parts, using a scaled myopic forecast would mean

$$\widehat{\operatorname{Var}}(r_{i,t+1:t+h}) = h\widehat{\sigma}_{i,t+1}^2, \tag{34}$$

while an iterated forecast is

$$\widehat{\operatorname{Var}}(r_{i,t+1:t+h}) = \sum_{k=1}^{h} \widehat{\sigma}_{i,t+k}^2.$$
(35)

4.7 VFS

To understand the possible need for some type of shrinkage on GARCH variance forecasts, it is useful to rewrite their formulas. Note that myopic GARCH variance forecasts consist of previous innovation terms and the unconditional variance:

$$\sigma_{t+1}^2 = (1 - a - b)\bar{\sigma}^2 \sum_{k=0}^t b^k + a \sum_{k=0}^t b^k r_k^2$$
(36)

$$\rightarrow \frac{1-a-b}{1-b}\bar{\sigma}^2 + a\sum_{k=0}^{\infty} b^k r_k^2 \text{ as } t \rightarrow \infty,$$
(37)

where, similar as in Equation (13) for the correlation matrix, the weights of the right part add up to $\frac{a}{1-b}$.

The same holds for iterated h-day-ahead forecasts:

$$\hat{\sigma}_{t+h}^2 = (1-a-b)\bar{\sigma}^2 + a \mathbb{E}[r_{t+h-1}^2|I_t] + b\hat{\sigma}_{t+h-1}^2$$
(38)

$$= (1 - a - b)\bar{\sigma}^2 + (a + b)\hat{\sigma}_{t+h-1}^2$$
(39)

$$= (1+a+b)(1-a-b)\bar{\sigma}^2 + (a+b)^2\hat{\sigma}_{t+h-2}^2$$
(40)

$$= \dots = (1 - a - b)\bar{\sigma}^2 \sum_{k=0}^{n-2} (a+b)^k + (a+b)^{h-1}\sigma_{t+1}^2$$
(41)

$$= (1 - (a+b)^{h-1})\bar{\sigma}^2 + (a+b)^{h-1}\sigma_{t+1}^2, \tag{42}$$

where one could see that $\hat{\sigma}_{t+h}^2 \to \bar{\sigma}^2$ if $h \to \infty$. Note that a single iterated forecast is a weighted combination of the unconditional and myopic conditional forecast, thus also of the unconditional variance and previous innovation terms. Therefore, accumulating them to an iterated forecast also yields this linear combination, but with their weights summed up. In the end, as well scaled myopic as iterated variance forecasts are just a weighted sum of the unconditional and conditional variance (or historical innovation terms).

The research of Ghysels et al. (2019) has shown the important practical difference between in-sample and out-of-sample forecasting (see Section 4.6). As shown before, the iterated multi-stepahead forecasts (optimally in-sample) are just weighted sums of the unconditional and conditional variance. However, the weights are determined under the assumption that the estimated model parameters are correctly specified. Yet, the longer the horizon, the more uncertainty is involved around whether the model-implied weights are optimal. One could argue that out-of-sample iterated forecasts of a certain asset's variance, should be shrunk towards the scaled forecast when they have shown to be relatively inaccurate. This follows the empirical findings of Ghysels et al. (2019) that iterated forecasts are preferred when models are correctly specified (in-sample optimal), and scaled forecasts are preferred when these specifications are doubtful (out-of-sample optimal). Therefore, this research proposes VFS, in which an optimal linear combination of the myopic and iterated variance forecast is determined, to minimize out-of-sample forecasting errors for each asset. This means that, instead of using either a scaled myopic or iterated forecast for every asset, cross-sectionally different weights are assigned to both:

$$\widehat{\operatorname{Var}}(r_{i,t+1:t+21}) = \delta_i \hat{\sigma}_{IT,i,t+1:t+21}^2 + 21(1-\delta_i)\sigma_{i,t+1}^2.$$
(43)

where $0 \le \delta_i \le 1$, and $\sigma_{i,t+1}^2$ is the conditional one-day-ahead forecast of asset *i*'s variance made at time *t*. Iterated forecast $\hat{\sigma}_{IT,i,t+1:t+21}^2$ can be calculated summing up Equation (42) for h = 1, ..., 21. An advantage of this approach is that it is empirically useful even if it would not improve forecasts. It namely ensures that conclusions are drawn ex ante about using either myopic or iterated forecasts.

To determine the optimal δ_i at each point in time for every asset, this research uses a forecast uncertainty measure as input, referred to as FUM_i. An intuitive measure would be based on past out-of-sample performances, but one might also use a theoretically estimated uncertainty. Either way, it can be fitted to the following framework. Although different functions can be argued as useful, I use the standard normal cumulative distribution function in a probit-type manner:

$$\delta_i = \Phi(\eta_0 + \eta_1 \text{FUM}_i)), \tag{44}$$

where η_0 and η_1 are asset-independent parameters to be estimated.

One might estimate them by minimizing the squared differences between variance forecasts and their 'realizations'. Yet, one could argue that the sum of squared errors should be made relative to the levels of RVs, such that all assets are treated equally. Their forecasts are then seen as equally accurate when they are a certain percentage away from the actual volatility, not a certain absolute value. Therefore, this research considers the relative sum of squared errors:

$$\min_{\eta} \sum_{i=1}^{N} \sum_{m} \frac{\left(\widehat{\operatorname{Var}}(r_{i,21m+1:21(m+1)};\eta) - \operatorname{RV}_{i,m}\right)^{2}}{\sum_{m} \operatorname{RV}_{i,m}^{2}}, \quad \operatorname{RV}_{i,m} = \sum_{k=1}^{21} v_{i,21m+k}, \tag{45}$$

where M is the number of months used to solve the problem. This research considers an expanding window for the first months, but when at least five years of forecast errors are available, a moving window of five years is used (M = 60).

Since variance realizations need to be estimated, the best obtainable daily volatility proxies $v_{i,t}$, based on the OHLC price data, are used for the monthly $RV_{i,m}$. The problem can be solved numerically, and to decrease the risk of obtaining a local instead of global minimum, multiple starting values for η are considered. The values for which the objective function is minimized are then chosen as starting value for the minimization problem.

One could argue that the volatility proxies might not be a very accurate estimate of the actual variance. However, in a research to multiple loss functions for volatility forecasts, Patton (2011) show that the MSE loss function is robust for any frequency of data used to determine the proxies.

4.7.1 Predictive accuracy as uncertainty measure

The first type of measure to use for VFS that might come to mind is an out-of-sample accuracy measure like the Root Mean Squared Prediction Error (RMSPE). With such a measure, a direct estimate of the uncertainty of specific volatility forecasts for each asset can be obtained and compared with another forecast. However, as argued in Section 4.7, a relative measure might be preferred over an MSE-like measure. Therefore, this research considers the Root Mean Squared Relative Prediction Error (RMSRPE) of monthly forecasts:

$$\text{RMSRPE}_{i} = \sqrt{\frac{1}{M} \sum_{m} \frac{\left(\hat{\sigma}_{i,m}^{2} - \text{RV}_{i,m}\right)^{2}}{\sum_{m} \text{RV}_{i,m}^{2}}}, \quad \text{RV}_{i,m} = \sum_{k=1}^{21} v_{i,21m+k},$$
(46)

where $\hat{\sigma}_{i,m}^2$ is the variance prediction of asset *i* in month *m*, and the range for *m* is set to the last five years if possible (M = 60). If there is not enough information, an expanding window is used. The same holds for estimating the optimal values of η in Equation (44).

As forecast uncertainty measure for Equation (43), it could be interesting to use the difference between the RMSRPEs of the myopic and iterated forecast:

$$\Delta RMSRPE = RMSRPE_{MY} - RMSRPE_{IT}.$$
(47)

The idea behind it is the expectation that a historical outperformance of one forecast over the other would imply the same outperformance soon. Therefore, one would expect $\eta_1 > 0$ in Equation (44), such that more weight is given to the iterated forecast if its recent performance were relatively better. From now on, VFS with the RMSRPE differences as uncertainty measures is referred to as VFS1.

An easier approach than VFS, using the differences in RMSRPEs, is taking the iterated forecast when Δ RMSRPE > 0, and the scaled forecast when Δ RMSRPE < 0. Although it does not involve VFS, it is implied when setting η such that $|\eta_0| < \infty$ and $\eta_1 \rightarrow \infty$. Therefore, this 'picking forecast' can be seen as a restricted version of VFS, and it can be used as a benchmark for the unrestricted version. Note that it assumes momentum in the optimality of a forecast, while VFS also allows for reversal or neither two. Yet, the most important difference is that VFS considers linear combinations, and the picking forecast only makes binary choices.

4.7.2 Kurtosis as uncertainty measure

As mentioned in Section 4.7.1, the difference in historical predictive accuracies might be used as an indicator for variance forecast uncertainties. However, another useful indicator might be an asset's kurtosis estimate.

Kurtoses tend to differ cross-sectionally: some stock returns come close to a normal distribution, while others are far more fat-tailed. Therefore, Student t-distributions with cross-sectionally differing degrees of freedom could be considered, implying different kurtoses among stocks. Alberg et al. (2008) estimate the daily Tel Aviv 25 and 100 index volatilities with asymmetric GARCH models. They show that one-month-ahead forecasts of an EGARCH model with skewed Student t-distribution are 22 and 25% more accurate than GARCH forecasts, respectively (in terms of MSPE). Yet, for the DCC-NL(-OHLC) model, not only good volatility forecasts need to be made. Also, a good fit of the variances is necessary to devolatize returns, which are used to estimate the unconditional correlation matrix. Since the different fat tails of stocks are easy to capture with t-distributions, not only the volatility forecasts, but also their fits might be improved with it. This already indicates that a t-distribution could be of great importance in this research.

Moreover, an important attribute of the t-distribution for this research is the fact that kurtosis estimates can be obtained for each asset. One can namely interpret it as some uncertainty measure of an asset that cannot be captured in the variance. Elaborating on that idea: a portfolio like GMV wants to obtain minimum variance and therefore puts most weight on assets with the lowest estimated volatilities. However, GMV can also be seen as some minimum risk portfolio, where one should not confuse variance with risk. Although an asset could have a relatively low variance, it might still have a high kurtosis, increasing the risk on very extreme returns, which are unwanted in a minimum risk portfolio. Therefore, using kurtosis estimates of assets as uncertainty measures in the VFS framework could turn out to be useful.

The only change compared with the GARCH model, introduced in Section 4.1, is the distribution assumed to be followed by the error terms. Therefore, the estimation procedure stays the same, although the likelihood function changes, and an extra parameter, the degrees of freedom ν , needs to be estimated. Since a GARCH model is fitted to each stock, an asset-dependent ν_i is estimated. To obtain an estimate for the kurtosis of a stock's return, κ_i , the following property is used:

$$\kappa_i = 3 \frac{\nu_i - 2}{\nu_i - 4}, \text{ for } \nu_i > 4,$$
(48)

while $\kappa_i = \infty$ for $2 < \nu_i \leq 4$. The forecast uncertainty measure can then be set to the estimated excess kurtosis: FUM_i = $\hat{\kappa}_i - 3$. From now on, VFS with the kurtosis as uncertainty measure is referred to as VFS2. When $\eta_1 > 0$ in Equation (44), a lower weight on the scaled forecast is ensured for higher kurtoses. This is intuitive, since the scaled forecast can be expected to be relatively inaccurate with higher tail risks.

4.8 Performance measures

To compare the accuracy of the covariance matrix forecasts, different measures are used. First, two loss functions are considered, which focus on the differences between the one-month-ahead forecast and the realized covariance matrix of that month. However, the matrix is unobservable, and the sample covariance matrix would be a bad estimator for only T = 21 days and $N \approx 500$ assets. Therefore, nonlinear shrinkage is applied as described in Section 4.2, using the devolatized regularized returns as described in Section 4.4. In this way, the best possible estimate is obtained, using the highest frequency of data considered in this research.

The first considered loss function is the well-known Frobenius loss, summing up the squared errors of every element in the covariance matrix:

$$\mathcal{L}_F(\Sigma,\widehat{\Sigma}) = \left\|\widehat{\Sigma} - \Sigma\right\|_F^2 = \sum_{i=1}^N \sum_{j=1}^N (\widehat{\Sigma}_{i,j} - \Sigma_{i,j})^2.$$
(49)

However, because this research focuses on the impact of the forecast on the GMV portfolio, another loss function is considered. Proposed by Engle et al. (2019), it is called the minimum-variance loss function (MV), representing the out-of-sample variance of the forecast-based GMV portfolio. One can therefore consider it a direct GMV portfolio accuracy measure. It is defined as follows:

$$\mathcal{L}_{MV}(\Sigma,\widehat{\Sigma}) = \frac{\operatorname{Tr}(\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1})/N}{(\operatorname{Tr}(\widehat{\Sigma}^{-1})/N)^2} - \frac{1}{\operatorname{Tr}(\widehat{\Sigma}^{-1})/N},$$
(50)

where $Tr(\cdot)$ denotes the trace of a matrix. For convenience, the function is scaled by 1000 in the results.

Although the loss functions are more direct measures of the covariance matrix forecasts' accuracy, one might argue that the proxy for the realization, Σ , is still too uncertain. Therefore, this research also compares the actual out-of-sample volatilities of the GMV portfolios, which are based on the covariance matrix forecast. These are expected to be lower for more accurate covariance matrix forecasts, since GMV portfolios are targeted to have the lowest out-of-sample variance. As described in Section 3, one can also expect the GMV portfolios to yield relatively high (risk-adjusted) returns, due to the low-volatility anomaly. For this reason, the average return and Sharpe ratio are also important in comparing results.

Lastly, portfolio weight statistics should indicate the extremity of the GMV portfolios. These are the averages of the turnover, Sum of Squared/Absolute Portfolio Weights (SSPW/SAPW), and the minimum and maximum weight. While the latter four measures are self-evident, the turnover is defined as follows:

$$TO = \sum_{t=1}^{T} \sum_{i=1}^{N} |w_{i,t} - w_{i,t}^{+}|, \qquad (51)$$

where w_{it}^+ represents the weight at the end of month t before rebalancing.

Besides comparing performances of the covariance matrix forecasts, measures to compare individual volatility forecasts are considered. In this way, we can directly find the impact of VFS, and compare scaled with iterated forecasts. The performances are measured by means of the average RMSRPE, as described earlier in Equation (46), but also the QLIKE loss function. Next to the MSE, it is shown to be robust by Patton (2011), although it is less sensitive to outliers. Just as with the RMSRPE, a volatility proxy is used as realization, yielding the feasible QLIKE as used by Ghysels et al. (2019):

$$\text{QLIKE}_{i} = \sum_{m=1}^{M} \log(\hat{\sigma}_{i,m}^{2}) + \frac{\text{RV}_{i,m}}{\hat{\sigma}_{i,m}^{2}},$$
(52)

where M is the total number of months considered in the out-of-sample results.

5 Results

In this section, first the estimated parameters are investigated. Next, the predictive accuracies of the GARCH models, with and without OHLC proxies, and various volatility forecast types are compared. A special focus is put on the estimated parameters and forecasting accuracies of VFS1. Subsequently, out-of-sample performances of GMV portfolios based on the different models are discussed. Lastly, forecast and portfolio performances of VFS2, but also the other methods with a t-distribution for GARCH are evaluated.

5.1 In-sample fits

Figure 5 shows the (median) estimated parameters of the DCC-NL model over time, using GARCH to model the volatilities. One can see in Figure 5a that the exposure to innovations in correlations, α , somewhat evenly increases over time. However, a sudden decrease of β is visible around November 2016. Therefore, one could argue that the persistence of the correlations somewhat decreased over the entire sample period.

A similar pattern occurs for the median GARCH parameters in Figure 5b. Yet, the value of \bar{a} (\bar{b}) is somewhat higher than α (lower than β) over time, as can also be seen in Equations (53) and (54). This indicates that correlations have higher persistence than variances, which therefore tend to deviate less over time. In addition, one can see the same decrease for \bar{b} as for β around November 2016. Observing the median RV in Figure 6, one can see the start of a period with relatively low volatilities at approximately November 2011. Since all models are evaluated with an estimation window of five years, the decreases in β and \bar{b} around November 2016 seem to be caused by the absence of highly volatile periods. Nevertheless, no detailed explanation for this phenomenon can be given, wherefore it is left to further research.



Figure 5: Estimated DCC parameters α and β , and cross-sectional median GARCH parameters \bar{a} and \bar{b} of the DCC-NL model with GARCH over time, evaluated in January 2000 — December 2020 using the previous five years of data at each point in time.



Figure 6: Median RV over January 2000 — December 2020

The median GARCH and DCC-NL model, with parameters averaged over time:

$$\sigma_{t+1}^2 = 4.10e - 5 + 0.095r_t^2 + 0.819\sigma_t^2 \quad \text{for} \ t = 1, ..., T,$$
(53)

$$Q_{t+1} = 0.045C + 0.020\mathbf{s}_t \mathbf{s}_t' + 0.936Q_t.$$
⁽⁵⁴⁾

In Figure 7, the cross-sectional distributions of the estimated innovation parameters of the GJR-GARCH model are shown. First, one can see that the average temporal mean $\bar{a} = 0.052$, while $\bar{\gamma} = 0.090$ is almost twice as high. This indicates that a negative return $r_{i,t}$ leads to an average $0.142r_{i,t}^2$ increase of the median stock's variance, while a positive return only increases it by $0.052r_{i,t}^2$. One can also see in Figure 7b that approximately 95% of the stocks have a positive $\hat{\gamma}_i$. Therefore, the GJR-GARCH seems to be important in capturing the much-skewed news impact curve in-sample.



Figure 7: Histograms of all stocks' temporal average estimated GJR-GARCH parameters \hat{a}_i and $\hat{\gamma}_i$, with the red solid line representing the mean, and the dashed lines the 5th and 95th percentiles.

Figure 8 shows the average, and 5th and 95th percentile of the estimated proportion of variance in closed markets over time. First, note that the average increased from 0.21 in January 2000 to 0.36 in December 2020. This indicates that overnight trading has gained more popularity compared with intraday trading over the sample period. It is probable that this is caused by the increasingly common after-hours trading, but also by increased globalization. Since demand for U.S. stocks by foreign investors has risen², and due to different time zones, increased stock market activity overnight could have been expected. Second, one can see that there are large differences in the proportion between various companies. In January 2000, around the 5th percentile, only 12% of stock variance is realized during closed markets, while this is no less than 33% for the 95th percentile. In December 2020, these percentages have risen to 23% and 47%, respectively.

The findings emphasize the importance of using time- and asset-dependent overnight trading proportions f in calculating volatility proxies, as described in Section 4.5.



Figure 8: Average estimated proportion of variance realized during closed markets (f_i) over time, with the dashed lines representing the 5th and 95th cross-sectional percentiles.

²https://www.investopedia.com/the-biggest-u-s-stock-buyer-in-q1-was-foreign-investors-5069449

5.2 Volatility forecasts

As mentioned in Section 4.8, volatility forecasts are evaluated using RMSRPE and QLIKE. Because this research considers many stocks, not only the average measure over the stocks is considered, but also the median and 95th percentile. In this way, one can determine whether a certain method is optimal for the whole cross-section, or conversely, suboptimal because of a few outliers.

Table 2 shows the out-of-sample measure realizations for GARCH and GJR-GARCH models, with and without OHLC, for the different forecast types: the unconditional, scaled, iterated, VFS1 and picking forecast (referring to the methods in Section 4.7.1). Although differences are small overall, one can see that the iterated forecast is highlighted most often, for all four models. When it is not bold, note that it is often still close to the optimal value. Comparing the RMSRPEs of the models with and without OHLC, higher averages can be found for the former. Yet, this only seems to be caused by outliers, since the medians are lower, and 95th percentiles are higher. Therefore, most stocks' variance forecasts seem to be improved when using OHLC data, although differences are small. For GJR-GARCH, we also see this overall improvement in the average.

		wi	thout OHL	$^{\prime}\mathrm{C}$			W	vith OHLC		
GARCH	Uncond.	Scaled	Iterated	VFS1	Pick	Uncond.	Scaled	Iterated	VFS1	Pick
Av. RMSRPE	0.112	0.110	0.104	0.107	0.108	0.113	0.118	0.108	0.114	0.115
Med.	0.082	0.074	0.070	0.071	0.071	0.078	0.072	0.067	0.066	0.068
95%	0.248	0.264	0.236	0.264	0.264	0.244	0.298	0.243	0.293	0.293
Av. QLIKE	-3.189	-3.530	-3.530	-3.532	-3.530	-3.010	-3.153	-3.270	-3.282	-3.153
Med.	-3.385	-3.644	-3.635	-3.651	-3.648	-3.351	-3.608	-3.588	-3.606	-3.609
95%	-1.433	-2.044	-2.005	-2.028	-2.045	-1.414	-2.059	-2.090	-2.058	-2.058
GJR-GARCH	Uncond.	Scaled	Iterated	VFS1	Pick	Uncond.	Scaled	Iterated	VFS1	Pick
Av. RMSRPE	0.112	0.112	0.109	0.111	0.111	0.113	0.111	0.106	0.108	0.109
Med.	0.082	0.074	0.073	0.073	0.073	0.078	0.069	0.067	0.066	0.066
95%	0.248	0.260	0.240	0.256	0.257	0.244	0.246	0.234	0.244	0.244
Av. QLIKE	-3.189	-3.520	-3.525	-3.521	-3.523	-3.010	-3.164	-3.271	-3.238	-3.164
Med.	-3.385	-3.631	-3.629	-3.632	-3.633	-3.351	-3.601	-3.588	-3.595	-3.595
95%	-1.433	-2.054	-2.091	-2.051	-2.054	-1.414	-2.098	-2.130	-2.077	-2.077

Table 2: Temporal average, median, and 95th percentile of RMSRPEs and QLIKEs in January 2000 — December 2020 based on GARCH or GJR-GARCH models, with or without the use of OHLC-based proxies. Unconditional forecasts are the unconditional variance. Scaled and iterated forecasts are as mentioned in Section 4.6. In each panel, the measure realization corresponding to the optimal forecast type, in terms of that measure, is made bold.

Yet, the conclusions change when considering the QLIKE measure. Almost all QLIKEs clearly deteriorate when moving from squared returns to OHLC-based proxies, for as well GARCH as GJR-GARCH. Looking at the forecast types, we also see that VFS1 is optimal using GARCH, while the iterated forecast remains optimal for GJR-GARCH.

Another thing one might notice, is the overall underperformance of the unconditional 'forecast' compared with all other forecast types, for almost every model-measure combination. This indicates that modelling the time dependence pays off out-of-sample.

5.3 VFS1

To evaluate VFS1, a comparison is made with its restricted version, the picking forecast. Figure 9 shows the average $\hat{\delta}_i$ over time for both forecasts using DCC-NL-OHLC with GARCH. Recall that the picking forecast 'picks' either $\hat{\delta}_i = 0$ or $\hat{\delta}_i = 1$. Therefore, the average at a specific time represents the proportion of stocks for which an iterated forecast has outperformed the scaled forecast in the previous five years. Since this proportion is beneath 0.5 most of the time, scaled forecasts are preferred for most of the stocks over the sample period. Yet, solely using scaled forecasts for all stocks does not seem to pay off, compared with only using iterated forecasts (see upper right panel in Table 2). This indicates that the choice of a forecast type should be made individually for each asset. Furthermore, one can see in Figure 9 that the proportion fluctuates over time. Therefore, one can conclude that not either scaled or iterated volatility forecasts should be preferred for all S&P 500 stocks, and that the individual choice should be made dependent of time.

Comparing the average $\hat{\delta}_i$ of the picking forecasts with those of VFS1, a striking finding is the sudden tendency towards the iterated forecasts. Half of the time it moves like the picking forecast, but the other half it is above 0.6, and even equal to 1 for the first five years. This is possibly caused by some outliers affecting η such that all $\hat{\delta}_i$ are suddenly set to 1, which is the downside of only using two parameters. However, the advantage of using VFS instead of the picking forecast is the possibility for linear combinations of both the iterated and scaled forecast.

Observing the estimated optimal values of η over time, one period is found for which $\eta_1 < 0$, in



Figure 9: Cross-sectional average of $\hat{\delta}_i$ over time for VFS1 and Pick, for the DCC-NL-OHLC model with GARCH.

which a recent outperformance of iterated over scaled forecasts was overall followed by the opposite. This period concerns 2009 — 2014, and η is estimated by using the last five years of data. Therefore, one can conclude that the sudden change of direction is caused by big outliers during the very volatile market at the end of 2008 (see Figure 6). One can also see in Figure 9 that this period exactly matches the period that the average $\hat{\delta}_i$ largely differs between VFS1 and the picking forecast.

Although the two implied $\hat{\delta}_i$ tend to differ, one cannot see clearly different forecast performances in Table 2. All measures tend to be very close to each other, despite the forecasts' theoretical differences. One explanation could be the somewhat larger, but still small difference between scaled and iterated forecast performances. Since VFS1 and the picking forecast are weighted combinations of those, an even smaller difference should not be unexpected. Based on the predictive measures and ease of computation and understanding, the picking forecast therefore might be preferred.

Furthermore, a slight underperformance over the iterated forecast can be found for both, while their measures are never worse than the scaled forecast. Yet, the VFS1 and picking forecast do not contain any look-ahead bias concerning the choice for either scaled or iterated forecasts. For this reason, the methods remain important for use in practice.

5.4 Portfolio performances

This section focuses on implications of the covariance matrix for GMV portfolios, instead of the volatility forecasts. Table 3 shows the estimated matrix losses and the performances of the GMV portfolios for multiple combinations of methods. First, for a direct comparison of the covariance matrix accuracies, the Frobenius and MV losses are used. As well their averages as medians are considered because of large outlier impact on the averages. Second, for an indirect but more conventional approach, the estimated volatilities of the portfolios are compared. Third, for portfolio purposes the average returns and Sharpe ratios are given. Lastly, some portfolio weight statistics are provided, indicating the (rebalancing) extremity of the positions.

The first striking result coming from Table 3 is the overall improvement of all methods over the 1/N portfolio. Each combination of methods approximately halves the portfolio volatility, while average annualized return improvements range between 1.0-5.5% in absolute terms, which is 14-75% higher, relatively. The risk- and return-wise improvements therefore almost triple the Sharpe ratio from 0.340 for the 1/N portfolio, to an average of no less than 0.920 for the GMV portfolios.

In Appendix B, the results of significance tests on the means and Sharpe ratios of the portfolios are given.³ Performing a *t*-test on the difference between 1/N and GMV mean returns, no significant improvements for any of the latter are found in Table 6. Yet, using the robust Sharpe ratio test of Ledoit and Wolf (2008), significant improvements *can* be found for the Sharpe ratio of all GMV portfolios over 1/N. This emphasizes the low-volatility anomaly as introduced in Section 3.2.

³From now on, significance levels of 5% are considered for every test.

	Pick	3.460	1.137	12.177	2.603	10.591	10.586	0.999	2.567	0.149	2.863	-0.017	0.225	Pick	3.439	1.139	11.407	2.590	10.646	10.397	0.977	2.590	0.153	2.849	-0.017	0.230	
C)	VFS1	3.488	1.186	10.511	2.690	10.768	11.037	1.025	2.428	0.147	2.860	-0.017	0.217	VFS1	3.464	1.146	11.094	2.621	10.787	10.949	1.015	2.453	0.144	2.846	-0.017	0.216	
-NL-OHL	Iterated	3.503	1.199	12.079	2.693	10.958	11.746	1.072	2.135	0.142	2.931	-0.019	0.203	Iterated	3.662	1.173	11.558	2.639	10.952	11.649	1.064	2.164	0.149	2.899	-0.018	0.212	
DCC	Scaled	3.506	1.140	11.411	2.536	10.186	10.543	1.035	2.722	0.151	2.846	-0.017	0.229	Scaled	3.465	1.119	10.710	2.532	10.238	10.400	1.016	2.748	0.154	2.829	-0.017	0.232	
	Uncond.	3.493	1.273	9.031	2.760	12.853	12.785	0.995	0.506	0.050	3.385	-0.023	0.059	Uncond.	3.488	1.278	9.036	2.745	12.881	12.900	1.001	0.503	0.050	3.379	-0.024	0.059	
	Pick	2.105	1.129	4.004	2.537	11.911	8.925	0.749	3.105	0.177	3.744	-0.028	0.224	Pick	2.135	1.123	4.031	2.521	11.849	9.242	0.780	3.121	0.174	3.669	-0.026	0.224	
	VFS1	2.104	1.130	3.965	2.574	11.730	9.500	0.810	3.083	0.178	3.769	-0.029	0.222	VFS1	2.159	1.134	4.043	2.553	11.953	9.760	0.817	3.059	0.174	3.704	-0.027	0.221	
DCC-NL	Iterated	2.101	1.204	3.932	2.622	11.642	9.155	0.786	2.832	0.182	3.850	-0.030	0.219	Iterated	2.141	1.183	4.007	2.611	11.870	9.746	0.821	2.904	0.179	3.781	-0.028	0.220	
Π	Scaled	2.174	1.124	3.884	2.480	11.492	8.384	0.730	3.267	0.177	3.702	-0.028	0.228	Scaled	2.178	1.095	3.960	2.485	11.650	8.869	0.761	3.280	0.172	3.627	-0.026	0.224	
	Uncond.	2.252	1.651	4.440	2.609	12.627	12.350	0.978	0.708	0.085	4.287	-0.037	0.081	Uncond.	2.221	1.632	4.470	2.619	12.734	12.445	0.977	0.704	0.084	4.265	-0.037	0.081	
	1/N					21.630	7.360	0.340	0.067	0.002	1.000	0.002	0.002	1/N					21.630	7.360	0.340	-2.0000	0.002	1.000	0.002	0.002	
	GARCH	Av. Frob. loss	Med. Frob. loss	Av. MV loss	Med. MV loss	Ann. Vol. $\%$	Ann. Mean $\%$	Ann. Sharpe	Av. Turnover	Av. SSPW	Av. SAPW	Av. Min. Weight	Av. Max. Weight	GJR-GARCH	Av. Frob. loss	Med. Frob. loss	Av. MV loss	Med. MV loss	Ann. Vol. %	Ann. Mean $\%$	Ann. Sharpe	Av. Turnover	Av. SSPW	Av. SAPW	Av. Min. Weight	Av. Max. Weight	

blios in January 2000 — December 2020 based on DCC-NL(-OHLC) models with either	ts of the $1/N$ portfolio. Unconditional forecasts are made using the unconditional variances	are as mentioned in Section 4.6. Performance measures are separated from portfolio weight	easure realization corresponding to the optimal forecast type, in terms of that measure, is	
e results of GMV portfolios in January $2000 - $ December 20	CH, compared with results of the $1/N$ portfolio. Unconditional	d and iterated forecasts are as mentioned in Section 4.6. Perfor	es. In each panel, the measure realization corresponding to th	
Table 3: Out-of-sampl	GARCH or GJR-GAR	and correlations. Scale	statistics by dashed lin	made bold.

A second important finding is the overall outperformance in terms of loss values and portfolio volatility for the scaled forecasts, despite the earlier found suboptimality of individual volatility forecasts. Since the correlation matrix that is used is the same for each forecast type, given which DCC and GARCH model is used, this performance should be caused by the volatility forecasts. This can probably be explained by the scaled forecasts overall being optimal for stocks with larger weights in the GMV portfolios. One can namely see low values for the MV loss of scaled forecasts too.

The third and seemingly most counter-intuitive result is the big increase of the average loss values, but decrease in portfolio volatility, when using DCC-NL-OHLC instead of DCC-NL. Yet, observing the losses over time for DCC-NL-OHLC, a high Frobenius loss of no less than 234 can be found for each forecast type around May 2010. Also, high but stable Frobenius losses of between 4.0-5.5 are realized for exactly five years onward, with a sudden decrease to values around 1.0 afterwards. However, this big outlier and its impact on future values is not observed for DCC-NL. The problem causing these high values is the combination of using high and low prices for volatility proxies, and one of the most volatile moments in the history of stock markets; the 2010 flash crash on May 6, 2010. This crash caused the S&P 500 to decrease by 9% in approximately half an hour, after which it rebounded just as quickly. Because the DCC-NL-OHLC method uses the COHLC-proxy, the low point in lots of stock prices during that day have increased predicted volatilities, also via increased unconditional variances. Observing the medians, no big differences can be found, although they still are not lower for DCC-NL-OHLC. Altogether, the findings indicate no improvement in the covariance matrix forecasts when using OHLC data.

Another observation one can make from Table 3, is the surprisingly good portfolio performance using the unconditional forecast. Recall that it uses the unconditional (but nonlinearly shrunk) correlation matrix and volatilities without exploiting any time dependence. It looks slightly suboptimal in terms of portfolio volatility and losses, but it has a relatively high average return. Therefore, it still yields competitive Sharpe ratios, which are even highest for DCC-NL. However, because dynamic forecasts with DCC-NL-OHLC have higher average returns and lower volatilities, they often obtain even higher Sharpe ratios.

Again, performing significance tests on the mean returns and Sharpe ratios, conclusions can be drawn on forecast type differences and whether the use of OHLC data pays off. Comparing the portfolio returns of the unconditional with the dynamic forecasts, not a single significantly different mean or Sharpe ratio can be found (see Table 6 of Appendix B). Therefore, one cannot conclude that the unconditional forecast outperforms the dynamic forecasts (or vice versa). Also, almost no significantly different performances can be found between all dynamic forecasts. Only for DCC-NL with GARCH, a significantly higher Sharpe ratio and mean return is found for VFS1, compared with the scaled forecast. Finally, testing the DCC-NL portfolios against their equivalents of DCC-NL-OHLC for the same forecast type, again no significantly different mean returns can be found. Yet, the scaled and iterated forecast *do* yield significantly higher Sharpe ratios, and *p*-values for the VFS1 and picking forecast are also relatively low. Therefore, one can conclude that the use of OHLC price data in dynamic forecast-based GMV portfolios pays off in terms of Sharpe ratios.

Furthermore, the table shows that taking into account the leverage effect, by using GJR-GARCH, does not improve the covariance matrix forecasts. Neither the losses nor the portfolio volatilities are lower when adding the GJR term to the model. So, the in-sample estimates showed the presence of a leverage effect in Section 5.1. Yet, one can conclude that GJR-GARCH has no added value in predicting either only volatilities or covariance matrices of S&P 500 constituents. Therefore, from now on, only the methods using GARCH are considered.

Lastly, looking at the portfolio weight statistics, relatively mediocre turnovers can be observed for the unconditional forecasts, as could be expected. This indicates that the positions are adjusted much smoother compared with the more dynamic forecasts. One can also observe higher turnovers for scaled than for iterated forecasts, which can be explained by the higher dependence of iterated forecasts on the unconditional variance. Another interesting finding is the overall decrease in magnitude of the statistics when using the OHLC-based volatility proxies instead of squared returns. However, high leverages can be found in all methods, since the method-average SAPW is always larger than 2.8. In that case, negative and positive weights sum up to -0.9 and 1.9, respectively, indicating a leverage of $\frac{0.9}{1.9} = 0.47$. For the unconditional forecast with DCC-NL and GARCH, an average SAPW of 4.287, and thus a leverage of no less than $\frac{1.644}{2.644} = 0.62$ occurs. Yet, the unconditional forecasts yield much more diverse portfolios than the dynamic forecasts, arising from the much lower average SSPWs.



Figure 10: Cumulative returns of the GMV portfolios in January 2000 — December 2020 based on DCC-NL-OHLC models with GARCH compared with the 1/N portfolio. The shaded areas represent high-volatility regimes from a bivariate Markov Switching model fitted to the monthly 1/N portfolio returns.

In Figure 10, the cumulative returns of the different forecast types are compared for the DCC-NL-OHLC model with GARCH. To measure the overall cumulative return performance of the forecasts, they are also compared with the 1/N portfolio. First, one can see a big difference between the 1/N and GMV portfolios. As shown before in Table 3, the former is more volatile, and the eventual cumulative return ($R_{\text{cum},t}$) is more than four times as low as for the average GMV portfolio (170.31% compared with 720.48%). Especially the relatively high and stable returns of the GMV portfolios during high-volatility regimes are impressive, indicating robust performances in different periods.

Another remarkable observation is, again, the good performance of the unconditional forecast. However, now becomes clear that most of the outperformance in (cumulative) returns comes from its stable returns during the 2007-2009 global financial crisis. One can see that this has not been the case for the 2020 recession, indicating a possibly coincidental superior performance previously. Over the whole sample period, none of the forecast types seems to be superior to one another.

5.5 VFS2

In this section, VFS based on the kurtosis is evaluated. Because it uses a *t*-distribution instead of a normal distribution for GARCH, the results for the other forecast types changed.

Table 4 shows the performances of the volatility forecasts using the *t*-distributed GARCH model with OHLC proxies. One can see that VFS2 has a lower average RMSRPE and QLIKE than the VFS1 and picking forecast, but it is also caused mostly by the higher percentiles. The medians namely are approximately equal, although the 95th percentile shows relatively lower values. Furthermore, while differences are small, an optimal average QLIKE measure is observed for VFS2 compared with all other forecast types. Both findings indicate that the estimated model-implied kurtosis is useful in combining iterated with scaled forecasts. A further look into the estimated η values also shows that $\hat{\eta}_1 > 0$ most of the time, implying that the variance of stocks with higher kurtoses is better forecasted using iterated instead of scaled forecasts. This is intuitive since higher kurtoses tend to go paired with more forecast uncertainty, and scaled forecasts are expected to be less accurate wit high tail risks.

Next to that, a comparison of the overall measures with those of the GARCH model with OHLC proxies and *normally* distributed returns in Table 2 can be made. One can see that most forecasts yield slightly lower RMSRPEs, while QLIKEs remain similar. This indicates that the use of a *t*-distribution might yield better GARCH forecasts.

In Table 5, the GMV portfolio performances for all forecast types are given. Comparing them with the upper-right panel of Table 3, no overall improvement of the portfolio volatility or any loss value is found. Different than for the DCC-NL-OHLC model with normal GARCH, one can now also see a relatively low average Frobenius loss value for the unconditional forecast. Moreover, although the VFS2 portfolio volatility is highest, VFS2 does yield the highest mean and Sharpe ratio. Yet, these differences are not significant, as can be seen in Table 7 of Appendix B.

	Uncond.	Scaled	Iterated	VFS1	VFS2	Pick
Av. RMSRPE	0.113	0.110	0.094	0.105	0.097	0.105
Med.	0.078	0.070	0.065	0.064	0.064	0.066
95%	0.244	0.260	0.219	0.259	0.227	0.256
Av. QLIKE	-3.010	-3.162	-3.296	-3.167	-3.305	-3.161
Med.	-3.351	-3.614	-3.619	-3.623	-3.624	-3.613
95%	-1.414	-2.050	-2.088	-2.043	-2.070	-2.028

Table 4: Average, median, and 95th percentile of RMSRPEs and QLIKEs for GMV portfolios in January 2000 — December 2020 based on the GARCH model with OHLC-based proxies and t-distributed returns. Unconditional forecasts are the unconditional variance. Scaled and iterated forecasts are as mentioned in Section 4.6. In each panel, the measure realization corresponding to the optimal forecast type, in terms of that measure, is made bold.

Furthermore, contrary to the optimistic forecast performances in Table 4, worse portfolio measures are found for VFS2 compared with VFS1 and Pick. Both losses and the portfolio volatility are also higher than for most other forecast types. Again, however, Table 7 of Appendix B shows no significant differences between the means and Sharpe ratios of the forecast types. As argued before, the VFS and picking forecasts remain important because they leave out any look-ahead bias. Therefore, one can conclude that the picking forecast is the most prominent for use in practice. It namely is easier to calculate, interpret, and it does not show a (significantly) weaker performance than the VFS forecasts.

	1/N	Uncond.	Scaled	Iterated	VFS1	VFS2	Pick
Av. Frob. loss		3.497	3.903	3.807	3.810	3.816	3.798
Med. Frob. loss		1.260	0.769	0.877	0.820	0.860	0.779
Av. MV loss		9.074	17.390	10.693	13.478	18.063	13.575
Med. MV loss		2.837	3.428	3.367	3.425	3.626	3.332
Ann. Vol. %	21.630	12.901	10.380	10.546	10.846	11.213	10.712
Ann. Mean $\%$	7.360	12.697	8.316	9.968	10.049	8.935	9.557
Ann. Sharpe	0.340	0.984	0.801	0.945	0.927	0.797	0.892
Av. Turnover	0.067	0.497	3.076	2.711	2.790	2.888	2.853
Av. SSPW	0.002	0.051	0.238	0.189	0.199	0.257	0.199
Av. SAPW	1.000	3.380	3.009	3.271	3.085	3.168	3.026
Av. Min. Weight	0.002	-0.026	-0.041	-0.030	-0.030	-0.050	-0.028
Av. Max. Weight	0.002	0.058	0.303	0.255	0.270	0.299	0.279

Table 5: Out-of-sample results of GMV portfolios in January 2000 — December 2020 based on the DCC-NL-OHLC model with GARCH, compared with results of the 1/N portfolio. Unconditional forecasts are made using the unconditional variances and correlations. Scaled and iterated forecasts are as mentioned in Section 4.6. Performance measures are separated from portfolio weight statistics by dashed lines. In each panel, the measure realization corresponding to the optimal forecast type, in terms of that measure, is made bold.

6 Conclusion & Discussion

In this research, I evaluate multiple one-month-ahead (co)variance forecasts and their implied GMV portfolios for the S&P 500 constituents in the period January 2000 — December 2020. Covariance matrix predictions are made using the recently-proposed DCC-NL model, and its improved version that exploits OHLC price data. For the latter model, I derive the daily volatility COHLC-proxy, that theoretically improves the efficiency of one of the most efficient proxies before by 12.1%. Furthermore, I propose VFS, which combines iterated and scaled myopic (GJR-)GARCH model forecasts. It is investigated whether the theoretically optimal iterated forecast, the often practically optimal scaled myopic forecast, or a combination should be used.

In the empirical results, all GMV portfolios yield Sharpe ratios around 0.920, which are significantly higher than the Sharpe ratio of 0.340 for the 1/N portfolio. Also, GJR-GARCH seems to be of added value in-sample, since almost all stock volatilities tend to be affected more by negative than positive returns. However, no improvement can be found in the out-of-sample forecasts, comparing them with regular GARCH forecasts. Furthermore, for OHLC-based volatility proxies, in-sample estimates of the variance proportion realized overnight should be determined assetting pattern of relative trading activity during closed markets is found over time. I also show that using OHLC-based volatility proxies decreases portfolio volatility and increases Sharpe ratios significantly. Yet, covariance matrix losses and volatility forecast accuracies do not improve on average. A possible explanation is the large impact of outliers in high-low ranges, boosting the proxy.

The VFS and picking forecasts do not show better results than either iterated or scaled forecasts. However, their estimates show that the optimal choice between iterated or scaled forecasts differs cross-sectionally and over time. They also are practically better than choosing either iterated or scaled forecasts for the whole dataset, since it prevents look-ahead biases. The picking forecast is recommended over VFS because of its easy interpretation and calculation, while it still yields insignificantly different results. Furthermore, no significant performance differences can be found between the VFS forecasts based on the kurtosis and those based on the historical predictive accuracy. Lastly, one can conclude that there is no need to use a t-distribution instead of a normal distribution for the GARCH models.

Although many new insights can be gained from this research, it also contains some limitations. First, since volatility and covariance predictions concern unobservable values, accuracy measures are never precisely correct. Given the estimated realizations that are used, conclusions could possibly change when more accurate estimations are made. Therefore, this research considers the most accurate estimates given the dataset, which are based on OHLC price data. Although the use of higher-frequency intraday price data could (slightly) improve the realization estimates, it would be an unnecessarily large computational burden given the high dimensionality of the dataset. Second, only a forecast horizon and portfolio holding period of one month is considered in this research. Therefore, none of the conclusions made are expected to be universal for other horizons. For example, one might expect the GJR-GARCH model to have more impact on the accuracy of one-day-ahead forecasts, since it focuses on a short-term leverage effect.

Next to using higher-frequency data and considering other forecast horizons, further research could also focus on dealing with outliers. Although outliers tend to have a role in any quantitative research area, it might be larger in dynamic (co)variance forecasting. First, each outlier can negatively affect upcoming GARCH forecasts when persistence is high. Second, as shown for the losses in the empirical results, volatility outliers can be so extreme, that performance measures are shifted, and therefore becoming unrepresentative for the whole sample. To prevent outliers from having such a big impact, further research could consider weakening extreme volatility proxies. This will decrease their effect on the model fit, upcoming forecasts, and performance measures that consider these proxies as realization estimates. Lastly, extending on the VFS research, a closer look could be given into the performance differences between scaled and iterated correlation matrix forecasts, thus in a multivariate setting.

References

- Alberg, D., Shalit, H., & Yosef, R. (2008). Estimating stock market volatility using asymmetric GARCH models. Applied Financial Economics, 18(15), 1201–1208.
- Andersen, T. G., Bollerslev, T., & Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance*, 6(5), 457–477.
- Black, F. (1976). Studies of stock market volatility changes. 1976 Proceedings of the American Statistical Association Business and Economic Statistics Section.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, 72(3), 498–505.
- Brailsford, T. J., & Faff, R. W. (1996). An evaluation of volatility forecasting techniques. Journal of Banking & Finance, 20(3), 419–438.
- Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: The Conditional Autoregressive Range (CARR) model. *Journal of Money, Credit and Banking*, 561–582.
- Chou, R. Y., Chou, H., & Liu, N. (2010). Range volatility models and their applications in finance. Handbook of quantitative finance and risk management (pp. 1273–1281). Springer.
- De Nard, G., Engle, R. F., Ledoit, O., & Wolf, M. (2020). Large dynamic covariance matrices: Enhancements based on intraday data. University of Zurich, Department of Economics, Working Paper 356.

- 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.
- Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1), 83–106.
- Engle, R. (2002). Dynamic Conditional Correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics, 20(3), 339–350.
- Engle, R. (2009). Anticipating correlations: A new paradigm for risk management. Princeton University Press.
- Engle, R. F., Ledoit, O., & Wolf, M. (2019). Large dynamic covariance matrices. Journal of Business & Economic Statistics, 37(2), 363–375.
- Feller, W. (1951). The asymptotic distribution of the range of sums of independent random variables. The Annals of Mathematical Statistics, 22(3), 427–432.
- Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. Journal of Business, 53(1), 67–78.
- Ghysels, E., Plazzi, A., Valkanov, R., Rubia, A., & Dossani, A. (2019). Direct versus iterated multiperiod volatility forecasts. Annual Review of Financial Economics, 11(1), 173–195.
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5), 1779–1801.
- Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? Journal of Applied Econometrics, 20(7), 873–889.
- Hautsch, N., Kyj, L. M., & Malec, P. (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, 30(2), 263–290.
- Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. The Journal of Finance, 58(4), 1651–1683.
- James, W., & Stein, C. (1961). Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 4(1), 361–380.
- Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2), 365–411.
- Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal* of Empirical Finance, 15(5), 850–859.
- Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. The Annals of Statistics, 40(2), 1024–1060.
- Ledoit, O., & Wolf, M. (2017a). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *The Review of Financial Studies*, 30(12), 4349–4388.

- Ledoit, O., & Wolf, M. (2017b). Numerical implementation of the QuEST function. Computational Statistics & Data Analysis, 115(1), 199–223.
- Ledoit, O., & Wolf, M. (2019). The power of (non-)linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*.
- Ledoit, O., & Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. Annals of Statistics, 48(5), 3043–3065.
- Molnár, P. (2016). High-low range in GARCH models of stock return volatility. Applied Economics, 48(51), 4977–4991.
- Nelson, D. B., & Cao, C. Q. (1992). Inequality constraints in the univariate GARCH model. Journal of Business & Economic Statistics, 10(2), 229–235.
- Pakel, C., Shephard, N., Sheppard, K., & Engle, R. F. (2020). Fitting vast dimensional time-varying covariance models. Journal of Business & Economic Statistics, 39(3), 1–17.
- Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. Journal of Business, 53(1), 61–65.
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256.
- Peters, J.-P. (2001). Estimating and forecasting volatility of stock indices using asymmetric GARCH models and (skewed) Student-t densities. *Preprint, University of Liege, Belgium*.
- Plyakha, Y., Uppal, R., & Vilkov, G. (2012). Why does an equal-weighted portfolio outperform value- and price-weighted portfolios? Available at SSRN 2724535.
- Rogers, L. C. G., & Satchell, S. E. (1991). Estimating variance from high, low and closing prices. The Annals of Applied Probability, 1(4), 504–512.
- Rogers, L. C., Satchell, S. E., & Yoon, Y. (1994). Estimating the volatility of stock prices: A comparison of methods that use high and low prices. *Applied Financial Economics*, 4(3), 241–247.
- Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 3(1), 197–206.
- Yang, D., & Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business*, 73(3), 477–492.

A Optimal weight COHLC-proxy

To find the optimal w in $\hat{v}_{i,t}^{COHLC}$, as described in Equation (27), the variance of the proxy is minimized. To derive the variance, some assumptions are made, also as explained in Section 4.5. First, prices are assumed to follow a continuous-time geometric Brownian motion with zero drift. Second, volatility is assumed to be constant during a day, whether the market is open or closed. Therefore, prices are assumed to fluctuate overnight too. Yet, as indicated in Section 4.5, adjusting the proxy using f ensures that variance differences between closed and open markets can still be exploited.

The variance of the COHLC-proxy equals:

$$\operatorname{Var}(\hat{v}_{i,t}^{COHLC}) = \frac{w^2}{f^2} \operatorname{Var}\left(\left[\log\left(\frac{o_{i,t}}{\tilde{c}_{i,t-1}}\right) \right]^2 \right) + \frac{(1-w)^2}{(1-f)^2} \operatorname{Var}(\hat{v}_{i,t}^{OHLC}),$$
(55)

where the covariance part equals zero, because the two combined proxies consider non-overlapping periods (closed and open market times), which are independent for Brownian motions. The first variance part can be easily found using the independent increments property of the Brownian motion:

$$\log(o_{i,t}) - \log(\tilde{c}_{i,t-1}) \sim N(0, \sigma^2 f),$$
(56)

since f is the time between close and open. Also using the fact that $\operatorname{Var}(X^2) = 2a^2$ when $X \sim N(0, a)$, yields $\operatorname{Var}\left(\left[\log\left(\frac{o_{i,t}}{\tilde{c}_{i,t-1}}\right)\right]^2\right) = 2\sigma^4 f^2$. Rogers and Satchell (1991) already found $\operatorname{Var}(\hat{v}_{i,t}^{OHLC}) \approx 0.331\sigma^4$ considering open to close being a full day. Adjusting for the fact that σ^2 is from close to close now, $\operatorname{Var}(\hat{v}_{i,t}^{OHLC}) = 0.331\sigma^4(1-f)^2$. Therefore,

$$\operatorname{Var}(\hat{v}_{i,t}^{COHLC}) = 2w^2 \sigma^4 + 0.331(1-w)^2 \sigma^4.$$
(57)

Minimizing this function, one can find the optimal $w^* = 0.14$ giving $\operatorname{Var}^*(\hat{v}_{i,t}^{COHLC}) = 0.284\sigma^4$. Compared with the COHL-proxy, which has variance $\operatorname{Var}^*(\hat{v}_{i,t}^{COHL}) = 0.323\sigma^4$, this means a theoretical efficiency gain of 12.1%.

B Portfolio performance *p*-values

GARCH		DCC-NL-OHLC									
Mean	Uncond.	Scaled	Iterated	VFS1	Pick	Uncond.	Scaled	Iterated	VFS1	Pick	
vs $1/N$	0.154	0.440	0.332	0.281	0.355	0.090	0.185	0.098	0.137	0.176	
vs Uncond.		0.065	0.126	0.195	0.118		0.301	0.642	0.421	0.300	
vs Scaled			0.217	0.029	0.338			0.078	0.408	0.995	
vs Iterated				0.530	0.665				0.231	0.065	
vs VFS1					0.110					0.296	
vs DCC-NL						0.745	0.105	0.060	0.258	0.207	
Sharpe	Uncond.	Scaled	Iterated	VFS1	Pick	Uncond.	Scaled	Iterated	VFS1	Pick	
vs $1/N$	0.010	0.015	0.008	0.004	0.011	0.003	0.000	0.000	0.000	0.000	
vs Uncond.		0.164	0.285	0.364	0.202		0.652	0.486	0.727	0.843	
vs Scaled			0.269	0.032	0.655			0.603	0.742	0.325	
vs Iterated				0.611	0.429				0.399	0.213	
vs VFS1					0.047					0.532	
vs DCC-NL						0.869	0.020	0.021	0.107	0.058	
GJR-GARCH		Ι	DCC-NL			DCC-NL-OHLC					
Mean	Uncond.	Scaled	Iterated	VFS1	Pick	Uncond.	Scaled	Iterated	VFS1	Pick	
vs 1/N	0.147	0.363	0.255	0.247	0.308	0.084	0.196	0.104	0.144	0.192	
vs Uncond.		0.106	0.216	0.233	0.155		0.241	0.563	0.355	0.230	
vs Scaled			0.154	0.099	0.458			0.053	0.340	0.911	
vs Iterated				0.995	0.369				0.184	0.030	
vs VFS1					0.219					0.219	
vs DCC-NL						0.723	0.243	0.174	0.374	0.376	
Sharpe	Uncond.	Scaled	Iterated	VFS1	Pick	Uncond.	Scaled	Iterated	VFS1	Pick	
vs $1/N$	0.010	0.008	0.003	0.003	0.005	0.002	0.000	0.000	0.000	0.000	
vs Uncond.		0.252	0.404	0.410	0.312		0.779	0.560	0.808	1.000	
vs Scaled			0.233	0.204	0.630			0.456	0.895	0.302	
vs Iterated				0.897	0.370				0.318	0.103	
vs VFS1					0.316					0.353	
vs DCC-NL						0.792	0.039	0.055	0.118	0.117	

Table 6: *p*-values for significance tests on the means and Sharpe ratios of the portfolios from Section 5.4. *t*-tests are used to test means against each other, and robust Sharpe ratio tests of Ledoit and Wolf (2008) are used for the Sharpe ratios (pre-whitened HAC *p*-values are given). The null hypotheses are $H_0: \mu_i = \mu_j$ and $H_0: SR_i = SR_j$, and the alternatives are $H_a: \mu_i \neq \mu_j$ and $H_a: SR_i \neq SR_j$. All tests are against the mean or Sharpe ratio of the forecast type indicated in the row, with the GARCH- and DCC-type model of the corresponding panel. Only for row 'vs DCC-NL', tests are against the mean or Sharpe ratio of the DCC-NL model, given the forecast type and GARCH-type model. For each test, the *p*-value is shaded if it is lower than a 5% significance level.

Mean	Uncond.	Scaled	Iterated	VFS1	VFS2	Pick
vs 1/N vs Uncond. vs Scaled vs Iterated vs VFS1 vs VFS2	0.094	$0.475 \\ 0.050$	$0.256 \\ 0.181 \\ 0.085$	$\begin{array}{c} 0.240 \\ 0.229 \\ 0.130 \\ 0.956 \end{array}$	$\begin{array}{c} 0.396 \\ 0.094 \\ 0.355 \\ 0.268 \\ 0.331 \end{array}$	$\begin{array}{c} 0.292 \\ 0.149 \\ 0.264 \\ 0.606 \\ 0.199 \\ 0.571 \end{array}$
Sharpe	Uncond.	Scaled	Iterated	VFS1	VFS2	Pick
vs 1/N vs Uncond. vs Scaled vs Iterated vs VFS1 vs VFS2	0.003	0.014 0.372	$\begin{array}{c} 0.001 \\ 0.932 \\ 0.112 \end{array}$	$\begin{array}{c} 0.001 \\ 0.834 \\ 0.241 \\ 0.783 \end{array}$	$\begin{array}{c} 0.014 \\ 0.353 \\ 0.816 \\ 0.098 \\ 0.216 \end{array}$	$\begin{array}{c} 0.002 \\ 0.686 \\ 0.371 \\ 0.490 \\ 0.320 \\ 0.354 \end{array}$

Table 7: *p*-values for significance tests on the means and Sharpe ratios of the portfolios from Section 5.5, thus for DCC-NL-OHLC with *t*-distributed GARCH. *t*-tests are used to test means against each other, and robust Sharpe ratio tests of Ledoit and Wolf (2008) are used for the Sharpe ratios (prewhitened HAC *p*-values are given). The null hypotheses are $H_0: \mu_i = \mu_j$ and $H_0: SR_i = SR_j$, and the alternatives are $H_a: \mu_i \neq \mu_j$ and $H_a: SR_i \neq SR_j$. All tests are against the mean or Sharpe ratio of the forecast type indicated in the row. For each test, the *p*-value is shaded if it is lower than a 5% significance level.