Volatility forecasting for low-volatility investing

Christian Conrad*[1] and Onno Kleen†[2,3]

[1]Heidelberg University
[2]Erasmus University Rotterdam
[3]Tinbergen Institute

July 10, 2022

Abstract

Low-volatility investing is typically implemented by sorting stocks based on simple risk measures; for example, the empirical standard deviation of last year's daily returns. In contrast, we understand identifying next-month's ranking of volatilities as a forecasting problem aimed at the ex-post optimal sorting. We show that time-series models based on intraday data outperform simple risk measures in anticipating the cross-sectional ranking in real time. The corresponding portfolios are more similar to the ex-ante infeasible optimal portfolio in multiple dimensions. Moreover, the increased signal in our improved volatility sorts survives portfolio weight smoothing for mitigating transaction costs.

Keywords: Factor investing, low-volatility anomaly, volatility forecasts, forecast evaluation

*Alfred-Weber-Institute, Heidelberg University, Bergheimer Str. 58, 69115 Heidelberg, Germany. Email: christian.conrad@awi.uni-heidelberg.de.

†Corresponding author: Econometric Institute, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, Netherlands. Email: kleen@ese.eur.nl.

1 Introduction

In the financial industry, low-risk strategies have become increasingly popular during recent years. Examples for those strategies are: betting against beta (Frazzini and Pedersen, 2014), low-volatility portfolios (Blitz and van Vliet, 2007), minimum variance portfolios (Clarke et al., 2006), and volatility-managed portfolios (Moreira and Muir, 2017). The low-volatility anomaly lead to the creation of the S&P 500 Low Volatility Index which measures the performance of the 100 least volatile stocks in the S&P 500.[1] The weights in the S&P 500 Low Volatility Index are proportional to the inverse of a stock's volatility measured as the standard deviation of daily returns in the preceding year.

In this paper, we focus on the implementation of low-volatility portfolios; that is, portfolios invested in low-volatility stocks. In financial practice, stocks are usually sorted according to some simple metric of a stock's total or idiosyncratic volatility. One example is the empirical standard deviation of monthly or daily returns over a certain period (e.g., the previous year, the previous 6 months, the previous month).[2] The corresponding low-volatility portfolio simply consists of, say, the 20% stocks with the lowest volatility. The portfolio is re-balanced on a monthly basis.

Clearly, from an ex-ante perspective it is not clear which proxy for stock volatility is best suited for stock selection. Therefore, we think of targeting the *optimal* low-volatility portfolio as a forecasting problem. We first introduce the *oracle* low-volatility portfolio which we define as the portfolio that an investor would choose with hindsight. Following the literature on estimating volatility from high-frequency intraday return data, we measure the monthly volatility ex-post by realized variances (Andersen et al., 2003). Using data for all common stocks in the United States (US) from the Center of Research in Security Prices (CRSP) during the period 1993–2019, we document the low-volatility anomaly from an ex-post perspective: low-volatility stocks have higher returns than high-volatility stocks.

We then investigate the question whether state-of-the-art volatility models are useful for anticipating the correct composition of the oracle portfolio in real time. That is, in each month we estimate various volatility models and use them to forecast the next month's volatility of each stock. We then form low-volatility portfolios based on the sorting of stocks according to the forecasted volatilities.

During recent years there has been substantial progress in the development of volatil-

[1] https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-low-volatility-indices.pdf

[2] Bali et al. (2016) provide an overview of the various metrics that are commonly used.

ity models. We employ a wide range of time series models. First, we use simple RiskMetrics models and various generalized autoregressive conditional heteroskedasticity (GARCH)-type models. In those models the conditional variance is treated as a latent process and daily (or monthly) returns are used for estimating volatilities. Second, we use heterogeneous autoregression (HAR)- and mixed-frequency data sampling (MIDAS)-type models. Here, the realized variances are modeled directly as a function of past realized variances. In addition, we consider forecast combinations; that is, we combine the forecasts from various volatility models according to measures of past forecast performance. We refer to those forecast combinations as "loss-based forecasts." We also use the measures that are commonly used for the volatility sorting of stocks as forecasting "models." For example, we consider the rolling window sample variance of daily returns based on the previous twelve months as the forecast for next month's volatility. We refer to those models/forecasts as benchmark models/forecasts. We then compare the forecast performance of the volatility models with the forecast performance of those benchmark models.

For the evaluation of the forecast performance we take two alternative perspectives. The first one is common in the financial econometrics literature (see, e.g., Ghysels et al., 2019): For each stock we evaluate the forecast performance of each model and check which model performs best and how the volatility models compare with the benchmark models. Unsurprisingly, the volatility forecasts of state-of-the-art volatility models outperform the simple benchmark model when measuring forecast accuracy by standard loss functions. For example, for 29% of the stocks the best performing model (according to the squared error loss) is a HAR-type specification that employs pooled estimation. In general, the HAR models dominate GARCH-type models and the benchmark models are dominated by essentially all other models. However, identifying the "optimal" volatility model for each stock is only possible ex-post and not in real-time because of potential time-variation in model performance and the small sample period. Even though the main focus of our paper is to improve low-volatility factor investing, it is noteworthy that our forecast evaluation is one of the most comprehensive in terms of forecasting volatility in a large cross section.

Alternatively, in each month we use just one volatility model to forecast the volatilities of all stocks. Based on the cross-sectional forecast performance of each model, we select the optimal model on a period-by-period basis. This is our second perspective which is feasible in real time. Again, we find that GARCH- and HAR-type models dominate the benchmark models. The loss-based forecasts lead to further improvements in forecast performance.

Next, we investigate whether the model-based volatility forecasts allow us to construct low-volatility portfolios that are "closer" to the oracle portfolio than the portfolios that are based on the benchmark forecasts (henceforth called "benchmark portfolios"). In that respect, it is important to note that it is not necessary to perfectly forecast each stocks' volatility in order to perfectly mimic the oracle portfolio. For example, if a model generates volatility forecasts which overestimate the volatility of each stock by 10%, the implied ordering of the stocks will still be fully correct. In addition, the empirical evidence in previous studies suggests that the relation between risk and return is rather flat for low- and medium-volatility stocks and then decreasing for high-volatility stocks (Blitz et al., 2019). Hence, misclassifying stocks may not be that costly as long as we avoid to include high-volatility stocks in the portfolio. Our results suggest that portfolios which employ loss-based forecasts (henceforth loss-based portfolios) mimic the true oracle portfolio more closely than the benchmark portfolios. We measure "closeness" by the "oracle overlap;" that is, the time-series average of the share of stocks that a particular low-volatility portfolio has in common with the oracle portfolio. At maximum we reach an oracle overlap of 66% which is more than 20 percentage points above the oracle overlap of the best benchmark portfolio. In that sense, the low-volatility portfolios that are based on state-of-the-art volatility models clearly improve upon the benchmark low-volatility portfolios.

However, the model-based weights lead to turnover twice as high as the leading benchmark portfolio which is based on 12 months of daily returns. As a consequence, we employ partial weight adjustments in our final portfolio allocation. The returns implied by these smoothed model-based portfolio weights are significantly outperforming the benchmark portfolios by an around 1pp higher annualized return and 10pp higher Sharpe ratio even after proportional transaction costs of 25bps. That returns are only significantly higher after partially adjusting the weights is not a bug but a feature. In practice, fund managers often mitigate transaction costs by such partial weight adjustment.

The rest of the article is structured as follows. In Section 2 we review the previous literature and present empirical evidence for the low-volatility anomaly. Section 3 presents the volatility models and Section 4 the data. We then evaluate the forecast performance of the volatility models in Section 5. A comparison of the various low-volatility portfolios is provided in Section 6. Finally, Section 7 concludes.

2 The Low-Volatility Anomaly

2.1 Related literature

Since the 1970s, numerous empirical studies have shown that the risk-return relationship is either flat or even negative which is in contrast to the prediction of the CAPM. The anomaly holds irrespectively whether risk is defined as beta (Black et al., 1972; Haugen and Heins, 1972, 1975), total volatility (Haugen and Heins, 1972, 1975) or idiosyncratic volatility (Ang et al., 2006, 2009). This is due to the fact that high-beta stocks are typically high-volatility stocks and total volatility is highly correlated with idiosyncratic volatility (Baker et al., 2011; Blitz et al., 2019).

Both rational and behavioral explanations have been proposed. One rational explanation is that investors face leverage constraints (Black, 1972); for example, regarding short-selling. Frazzini and Pedersen (2014) propose a model that incorporates such leverage constraints. Another rational explanation by Blitz and van Vliet (2007) argues that portfolio managers are typically subject to relative performance objectives which might render low-volatility stocks unattractive. A behavioral explanation is the possible preference of some investors for lottery-like payoffs examined by Barberis and Huang (2008); Bali et al. (2011). Asness et al. (2020) find evidence that support both the leverage and the lottery hypothesis.

In contrast to the studies above, we examine the low-volatility anomaly from a forecasting perspective by employing time-series models that are widely documented to perform better than trailing volatility. Ghysels et al. (2005) derive variance forecasts for the market based on MIDAS models to provide evidence for a positive risk-return relationship. In a similar manner, Fu (2009) uses the exponential generalized autoregressive conditional heteroskedasticity model by Nelson (1991) to forecast idiosyncratic volatilities which he finds to be positively correlated with returns—contradicting Ang et al. (2006, 2009). The fact that total volatility predicts returns is also exploitable by machine-learning techniques as shown by Gu et al. (2020). In this regard, Ghysels et al. (2005), Fu (2009), and Gu et al. (2020) demonstrate the usefulness of time-series models for portfolio sorting but their analyses are restricted to using daily return data and, as a consequence, inferior to forecasting models based on realized variances.

The literature on intraday data for variance-based portfolio sorting follows the simple trailing volatility approach. The study by Boudt et al. (2015) may be considered to be closest to ours. They use a S&P 500 real-time constituents data set to overcome the survivorship bias in De Pooter et al. (2008); Hautsch et al. (2015). In Boudt et al. (2015), the authors come to the conclusion that there is no (statistically significant) benefit in returns from using intraday data but portfolio returns are less volatile. In contrast to our study, they do not use volatility models and have a short sample from 2007–2012. However, already Haugen and Heins (1975) note that high-volatility stocks are primarily outperformed by low-volatility stocks at longer investment periods which they attribute to superior performance during bear markets. Liu (2009) concludes that at a monthly investment horizon there is no benefit from intraday data if an investor has access to at least 12 months of daily data. Similarly, Amaya et al. (2015) find no significant predictive power of lagged realized variances on weekly stock returns. Another branch of the literature examines volatility timing for aggregated portfolio returns (Moreira and Muir, 2017, 2019; Cederburg et al., 2020).

2.2 A new perspective on the anomaly

In this section, we take a new perspective on the low-volatility anomaly by creating and evaluating the performance of an ex-ante infeasible "oracle portfolio." The usual approach in the literature on the low-volatility anomaly is as follows: At the end of each month m, all stocks are ranked according to a proxy of their volatility. Volatility is often measured as the square-root of the sum of squared daily returns over the previous month, the previous six months or the previous year (Bali et al., 2016).[3] Based on the ranking for month m, equal-weighted decile portfolios for month $m+1$ are constructed. Then, according to the low-volatility anomaly the portfolio of stocks in the first decile has higher average and risk-adjusted returns than the portfolio of stocks in the fifth decile (see, e.g., Blitz and van Vliet, 2007).

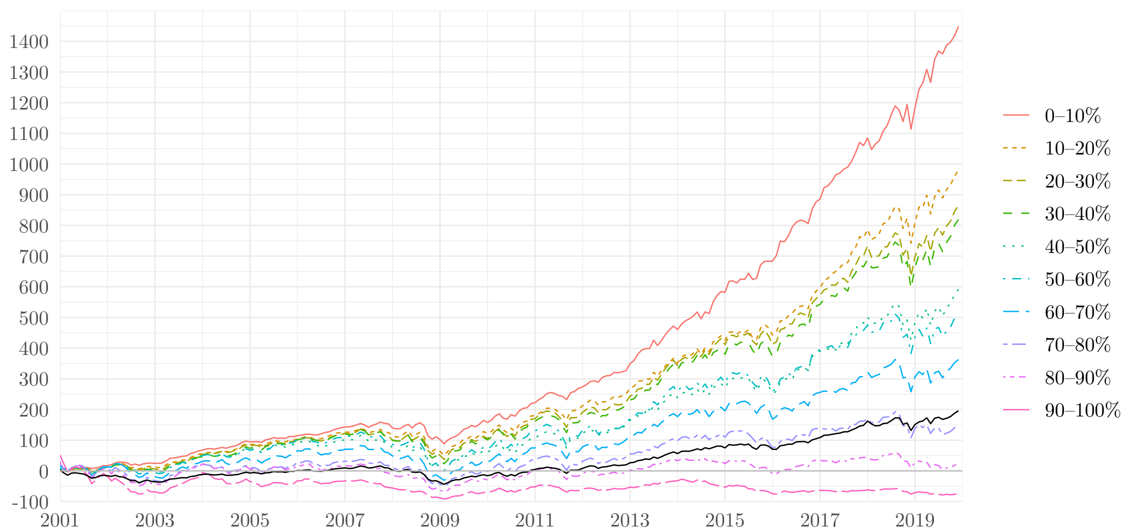
We now take an ex-post perspective by asking the following question: What would have been the "true" decile portfolios, that is, the portfolios that are formed based on the ex-post volatilities? Because stock volatilities are latent, even ex-post the correct ranking of stocks is not absolutely certain. We rely on the literature on estimating stock volatility from high-frequency intraday data and, hence, base the ex-post oracle portfolios on realized variances: At the end of each month $m+1$, we compute the monthly realized volatility of each stock as the square-root of the sum of daily realized variances based on 5-minute intraday data.[4] We then consider the ex-ante infeasible decile portfolios that are formed at the end of month m according to the realized volatility from the end of month $m+1$. Although these decile portfolios cannot be constructed in real

[3]Because "total volatility" and idiosyncratic volatility are typically highly correlated, Bali et al. (2016) argue that portfolio sortings on one or the other measure of volatility usually lead to the same results. Using a measure based on data from the last month was suggested in Ang et al. (2006, 2009).

[4]For details see Section 4.

time, they tell us how an investor would have behaved with hindsight. Figure 1 shows the performance of the decile portfolios during the 2001 to 2019 period. The first decile portfolio clearly outperforms all other decile portfolios. The weakest performance can be observed for the 10th decile portfolio. Thus, the preliminary results for our oracle portfolio confirm the low-volatility anomaly from an ex-post perspective. In the following, we will refer to the first decile portfolio as "the" oracle portfolio. As an alternative oracle portfolio, we considered a portfolio that is based on a monthly volatility measure which uses squared daily returns only. However, the oracle portfolio based on intraday realized variances clearly outperforms the portfolio based on squared daily returns in terms of average excess returns and Sharpe ratio. It also has considerably lower turnover.

Figure 1: Discrete returns of equal-weighted decile oracle portfolios.



Notes: Monthly discrete excess returns of the decile oracle portfolios in between 2001:M1 to 2019:M12. The volatility sorts are based on a large cross section of stocks in the CRSP sample. For more info about the data, see Section 4. As a benchmark, excess returns of the CRSP value-weighted index are depicted in black.

In Table 1, we compare the performance of the infeasible oracle portfolio with the performance of feasible low-volatility portfolios based on the volatility of the previous 4 years (4y-RV^d), the previous year (12m-RV^d), the previous six months (6m-RV^d) and the previous month (1m-RV^d). We will refer to those three portfolios as (feasible) benchmark portfolios. The equal-weighted portfolios in Panel A of Table 1 show that the infeasible oracle portfolio has a higher return, a lower volatility and, hence, a higher Sharpe ratio than the 4 benchmark portfolios. Note that the oracle as well as the benchmark portfolios clearly beat the market portfolio in all three dimensions. While the oracle portfolio achieves a risk reduction of 33%, the benchmark portfolios reach a risk reduction of almost 30%.

Table 1: Summary of rolling-average-based portfolios.

	Ret	Std	SR	ARVol	OO	OD	TO
Panel A: Equal-weighted portfolios							
Oracle	14.96	9.45	1.58	16.80	—	—	57.80
4y-RV^d	9.17	10.65	0.86	21.89	49.68	0.18	8.36
12m-RV^d	9.24	10.16	0.91	21.31	50.89	0.17	16.51
6m-RV^d	9.44	10.19	0.93	21.26	51.34	0.17	27.58
1m-RV^d	9.59	10.62	0.90	22.28	47.37	0.18	106.89
Panel B: Volatility-weighted portfolios							
Oracle	14.92	8.62	1.73	15.68	—	—	61.88
4y-RV^d	9.10	10.54	0.86	21.65	43.42	0.21	7.82
12m-RV^d	9.24	10.03	0.92	21.06	44.62	0.21	15.82
6m-RV^d	9.21	9.91	0.93	20.92	46.05	0.20	27.96
1m-RV^d	8.72	9.88	0.88	21.30	46.21	0.19	108.01
Panel C: Smoothed volatility-weighted portfolios							
4y-RV^d	9.08	11.50	0.79	24.96	40.49	0.21	4.94
12m-RV^d	9.27	11.24	0.82	24.58	42.42	0.20	5.58
6m-RV^d	9.18	11.11	0.83	24.49	42.93	0.19	6.16
1m-RV^d	9.26	11.60	0.80	25.78	41.01	0.18	8.37
Market	7.63	15.42	0.50	—	—	—	—

Notes: We report arithmetic means of discrete excess returns (Ret), their standard deviation, and the corresponding Sharpe ratio (SR). ARVol is the square-root of the time-averaged "average realized variance" which is defined to be the cross-sectional average RV inside the corresponding low-volatility portfolio, $ARVol = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{N} w_{i,m} RV_{i,m}}$. Annualized scale. The true positive (TP) ratio is the average share of ex-post oracle stocks that are included in the respective portfolio. Oracle overlap (OO) is the share of equally allocated weight relative to the equal-weighted oracle portfolio (Panel A) or relative to the volatility-weighted portfolio (Panel B and C). Note that the OO in Panel C is relative to the non-smoothed oracle portfolio in Panel B. OD is the average RMSE of the allocated weights relative to the oracle portfolio weights. For the definition of turnover (TO) see Subsection 6.3. OO, OD, and TO are reported in percentages. The evaluation period 2001:M1–2019:M12. The market return is given by the CRSP value-weighted portfolio return.

At first sight, there seem to be no major differences in the performance of the 4 benchmark portfolios. However, differences become apparent when considering additional characteristics of the portfolios. First, we compute the average realized volatility (ARVol) of each portfolio. That is, in each month m we compute the cross-sectional average of the realized variance of the stocks in the portfolio and then average over time. ARVol is the square-root of this quantity. By construction, the oracle portfolio has the lowest ARVol. ARVol is the highest for 1m-RV^d which suggests that the 1m-RV^d

portfolio has the severest classification errors. This is confirmed when computing the "oracle overlap" (OO): In each month m we calculate the overlap in weights between the benchmark portfolios and the oracle portfolio. We average the corresponding share over time. On average, only 47.37% of the stocks in the 1m-RV^d portfolio are also part of the oracle portfolio. This number increases to almost 51% for the 6m-RV^d and 12m-RV^d portfolios. The oracle distance (OD) measures the average root-mean square error (RMSE) of the allocated weights relative to the oracle portfolio weights. With respect to OD, all portfolios perform roughly the same. However, where we see a huge difference across portfolios is the turnover which is up to 106% for 1m-RV^d. We clearly observe a smoothing effect for volatility proxies based on larger trailing windows like 4y-RV^d. This implies that after transaction costs the 12m-RV^d portfolio clearly dominates the 1m-RV^d portfolio (see Section 6.4). In the following, we will refer to the 12m-RV^d portfolio as the "benchmark portfolio." We choose the 12-month window as the main benchmark because this window length is also applied in the S&P 500 Low Volatility Index. In Panel B of Table 1, we report figures for portfolios that are inversely proportional to the forecasted volatility, which is in line with the S&P 500 Low Volatility Index weights. In Panel C of Table 1, we report figures with partial weight adjustment; for more information, see Eq. (8). Although the TO of the oracle portfolio is comparably high, we will show that even after (reasonable) transaction costs it generates higher returns than any of the benchmark portfolios.

Obviously, an investor would be interested in replicating the oracle portfolio as closely as possible. We denote the realized variance of stock i, $i = 1, \ldots, n$, in month $m+1$ by $RV_{i,m+1}$. The oracle portfolio is based on the ascending ordering of the monthly realized variances of all n stocks: $RV_{1,m+1} \leq RV_{2,m+1} \leq \ldots \leq RV_{n,m+1}$. Hence, we can think of the task of replicating the oracle portfolio as a forecasting problem. We forecast the realized variances of the n stocks based on information up to the end of month m and form a portfolio based on the ranking that is implied by the forecasted variances $\widehat{RV}_{i,m+1|m}$, $i = 1, \ldots, n$. We will address the forecasting problem in three steps:

1. We first estimate various volatility models for each stock and evaluate the forecast performance of each model. This allows us to answer the following questions: Do state-of-the-art volatility models provide better forecasts of the cross-sectional stock volatility than the simple benchmark models? Is there a single volatility model (or a few volatility models) that outperform(s) the others? Because the benchmark models are not designed to accurately forecast volatility but rather to "identify" stocks that qualify for the low-volatility portfolio, we expect that the answer to the first question will be "yes." As most of the literature on volatil-

ity forecasting focuses on daily forecasts, the one-month horizon that is needed in our setting will shed some new light on the potential advantages of models that directly model the realized variances over models that treat the conditional variance as latent when forecasting volatility over longer horizons.

2. We will evaluate whether the forecasts from the volatility models do translate into a "more accurate" ranking of stock volatilities than the forecasts from the benchmark models. We will measure the accuracy by the oracle overlap. That is, we evaluate whether the decision to include a stock in the low-volatility portfolio is correct. Note that the oracle overlap can be high, even if the ranking that is implied by the volatility forecasts is far from perfect. However, a perfect ranking would imply a 100% oracle overlap.

3. Do the portfolios with the highest oracle overlap generate the highest returns? We will see that the answer to this question crucially depends on portfolio turnover and transaction costs.

3 Models

We consider a wide range of models which represent the state of the art in volatility modeling. The models can be broadly classified as either RiskMetrics, GARCH, HAR or MIDAS. While in the GARCH and RiskMetrics approach volatility is treated as a latent variable, the HAR and MIDAS specifications model realized variances directly. In the following, we briefly introduce the various model specifications. A more detailed description of the different models can be found in Appendix A.

RiskMetrics (RM): We use four variants of the RiskMetrics model. Two variants employ monthly realized variances based on squared daily returns while the other two employ weighted averages of squared daily returns directly. The RiskMetrics models use either six or twelve months of past return data. Note that the RiskMetrics models can be considered as restricted GARCH models with fixed ARCH/GARCH parameters and a constant equal to zero.

GARCH: Besides the simple GJR-GARCH of Glosten et al. (1993), we employ a "Panel GARCH" model which uses variance targeting for each stock and restricts the ARCH/GARCH coefficients to be the same across stocks. We also use the Factor GARCH model of Engle et al. (1990) and combine it with the GARCH-MIDAS of Engle et al. (2013). As explanatory variables in the long-term component, we use the VIX, housing starts and the term spread. Those variables have been shown to be powerful predictors of longer term volatility (Conrad and Loch, 2015; Conrad and Kleen, 2020). Correspondingly, these models are denoted as Factor GARCH-VIX, Factor GARCH-ΔHous, and Factor GARCH-TS. We also consider two types of multiplicative error (MEM) models (Engle and Gallo, 2006).

HAR: We consider the original HAR specification as suggested by Corsi (2009) as well as seven extensions. In the original HAR model the realized variance is a linear function of the lagged daily, weekly, and monthly realized variances. Among the extensions are specifications that model the realized variance of stock i as depending on stock i's lagged realized variances but also on a HAR-type forecast for the S&P 500, or the VIX index. We also use the "Panel HAR" model of Bollerslev et al. (2018). The tradeoff of pooled vs. individual estimation in HAR models is also assessed by Kleen and Tetereva (2022) which employ panel HAR models in a local linear forest. Their approach of endogenous clustering via decision trees leads to significant increases in forecast performance relative to the Panel HAR.

MIDAS: This type of volatility model has been proposed in Ghysels et al. (2004, 2005, 2006). The realized variance is modeled as a weighted average of lagged daily realized variances. The weights are parsimoniously parameterized via a flexible parametric weighting scheme. The HAR model of Corsi (2009) is nested when imposing certain constraints on the weights.

We estimate all models on a rolling window of four years with a minimum number of 600 observations.[5] Forecasts are computed for month $m = 1, \ldots, M$.

Ghysels et al. (2019) study the performance of iterated versus direct multi-step-ahead forecasting for GARCH, HAR and MIDAS models. Following their recommendations, we directly forecast the average 22-day realized variance for all HAR-type models. Similarly, we construct direct forecasts for the MIDAS models. The GARCH and MEM models are estimated using daily data and then iterated volatility forecasts are computed.

4 Data

Monthly portfolio returns are calculated from monthly total returns taken from the Center of Research in Security Prices (CRSP). We adjust for CRSP delisting returns such that we have a survivorship bias free data set (Shumway, 1997; Bali et al., 2016).

[5]The only exceptions are three variants of Factor GARCH-MIDAS models which employ housing starts or term spread data beginning in 1987 and the VIX and S&P 500 returns beginning in 1990 in order to identify the long-term component.

Similar to Bollerslev et al. (2019) and Bollerslev et al. (2022), we merge daily CRSP data with NYSE TAQ intraday data. Open and close prices per day are from taken from the daily CRSP data files. All intraday transaction data is obtained from NYSE TAQ. This trade data is cleaned according to Barndorff-Nielsen and Shephard (2002) and we include only trades from the exchange that is referenced in the daily CRSP data. Merging the two data sets is carried out via the WRDS linking tables. In each month, we only include stocks with a price larger than five dollars and a market capitalization of at least 1 million dollar.

Some of our models rely on intraday market data. For this, one-minute intraday data for the S&P 500 is downloaded from Tick Data.[6] Daily values for the VIX are obtained from the Cboe website.[7]. We estimate all time series models and evaluate our forecasts on the daily/intraday data set. Observations start in January 1993 and end in December 2019. For the intraday realized variance estimates, we include prices during market hours from 9:30 to 16:00 and calculate 5-minute log-returns. The first 5-minute return of each day is an open-to-close return and all others are close-to-close ones. We use 5-minute returns for two reasons: First, because this frequency is most commonly used, it makes our analysis comparable to others. Second, it has been shown to be a fairly robust choice as a trade-off between using high-frequency data and obstructing micro-structure noise related estimation errors (Liu et al., 2015). We rescale the intraday-based realized variance to the daily close-to-close period as discussed in Hansen and Lunde (2006). At day t and for stock i we will denote this combined measure by $RV_{i,t}$. The average monthly realized variance, $RV_{i,m}$, of stock i is defined as the average $RV_{i,t}$ over all days t in month m. Alternatively, squared daily (close-to-close) returns are often used as a simple but less accurate measure of volatility. We will denote this noisy proxy by $RV_{i,t}^{d}$.

Discrete excess market returns $R_{mkt,t}$ and the corresponding risk-free rates $R_{rf,t}$ are obtained from Kenneth R. French's data library.[8] For further factor analyses, we use the Fama-French(-Carhart) four- and five-factor portfolio returns; that is, daily average returns of SMB (Small Minus Big), HML (High Minus Low), MOM (Momentum), RMW (Robust Minus Weak) and CMA (Conservative Minus Aggresive) portfolios (Fama and French, 1993; Carhart, 1997; Fama and French, 2015). These are also obtained from Kenneth R. French's data library website.

SMB_t is the return on a diversified portfolio of small stocks minus the return on a

[6] https://www.tickdata.com

[7] http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data

[8] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

diversified portfolio of big stocks, HML_t is the difference between the returns on diversified portfolios of high and low book-to-market ratio stocks The two additional factors in the five-factor model can be understood as measures of profitability and investment. Hence, RMW_t is the calculated as the difference between returns on diversified portfolios of stocks with robust and weak profitability, and CMA_t is calculated as the difference between returns on diversified portfolios of low and high investment stocks, which Fama and French call conservative and aggressive. Because of collinearity, Fama and French (2015) refrain from including the momentum effect in their five-factor model and we follow their approach.

Last, real-time housing starts data is downloaded from ALFRED[9] and term-spread data from the New York Federal Reserve website.[10]

In the investment period from 2001–2019, our final data set includes on average 1820 stocks per month to invest in.

5 Forecast evaluation and model selection

In a first step, we evaluate the volatility forecasts from the different models. In the following subsection, we introduce four loss functions and then provide empirical results from an ex-post and a real-time perspective.

5.1 Loss functions

Following Patton (2011), we evaluate the volatility forecasts using robust loss functions. Suppose we are interested in evaluating the conditional variance forecast $\widehat{RV}_{m+1|m}$ against the true but unobservable conditional variance σ^2_{m+1} using the loss function $L(\sigma^2_{m+1}, \widehat{RV}_{m+1|m})$.[11] Then, the loss function is called robust if the expected loss ranking of two competing forecasts is preserved when replacing σ^2_{m+1} by a conditionally unbiased proxy. In the empirical application, we use the monthly realized variances RV_{m+1} as proxies for the unobservable σ^2_{m+1}. We will employ two popular loss functions which are robust: the squared error (SE) loss,

$$L(\sigma^2_{m+1}, \widehat{RV}_{m+1|m}) = (\sigma^2_{m+1} - \widehat{RV}_{m+1|m})^2,$$

[9] https://alfred.stlouisfed.org/series?seid=HOUST

[10] https://www.newyorkfed.org/research/capital_markets/ycfaq.html#/

[11] In this subsection, for simplicity in the notation we drop the index i.

and the QLIKE loss,

$$L(\sigma^2_{m+1}, \widehat{RV}_{m+1|m}) = \sigma^2_{m+1}/\widehat{RV}_{m+1|m} - \log(\sigma^2_{m+1}/\widehat{RV}_{m+1|m}) - 1.$$

The differences between the two evaluation criteria can be summarized as follows: While the SE is a symmetric loss function, the QLIKE is asymmetric and penalizes underestimation more heavily than overestimation. The QLIKE is less affected by extreme observations and requires weaker moment conditions when performing Diebold-Mariano-type tests of equal predictive ability (Patton, 2006).

5.2 Time series perspective

First, we evaluate the forecast performance of the various volatility models from an ex-post perspective. For each stock i, we consider the out-of-sample volatility forecasts $\widehat{RV}^j_{i,m|m-1}$, $m = 1, \ldots, M$, stemming from model j. For each loss function and with hindsight, we can measure the average loss of model j for stock i across time as

$$L^j_i = \frac{1}{M} \sum_{m=1}^{M} L^j(RV_{i,m}, \widehat{RV}^j_{i,m|m-1}). \tag{1}$$

We denote the stock specific loss of the benchmark forecast (12m-RV^d) by L^B_i. As a measure for the forecast accuracy of a particular model j relative to the benchmark, we consider the following statistic

$$LR^j = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{L^j_i/L^B_i<1}, \tag{2}$$

where $\mathbf{1}_{L^j_i/L^B_i<1}$ is an indicator function which equals one if $L^j_i/L^B_i < 1$ and zero else. Hence, LR^j_i reports the share of stocks for which model j outperforms the benchmark.

The left-hand-side of Table 2 shows LR^j for the two loss functions. Independently of the loss function, almost all models beat the benchmark for more than 50% of the stocks. In particular, we find that HAR-type models perform very well relative to the benchmark. For example, the SE loss of the Panel-HAR-LR model is lower than the loss of the benchmark (12m-RV^d) for 94% of the stocks. Among the GARCH-type models, the Factor GARCH with the VIX as an external covariate does best according to the SE. Interestingly, the RiskMetrics model based on 12 months of monthly returns (RM monthly, 12 months) performs very well for both loss functions.

In order to compare the various models not only to the benchmark but also with each other, we also report the share of stocks for which a particular model j performed best (denoted by $Rk_i^j \leq 1$) or was among the best four models (denoted by $Rk_i^j \leq 4$) as measured by L_i^j. Our previous findings are confirmed: According to the SE loss, the Panel-HAR-LR model has the lowest loss for 29% of the stocks. Other models that perform well are the Panel HAR without longrun component, the HAR-VIX, and the MIDAS models. When considering the top-4 models and according to the SE, the Panel-HAR-LR model is included in this set for 65% of stocks. Interestingly, the three benchmark models are almost never among the top-4. Note that the ranking of models is relatively robust across loss functions. In summary, HAR-type models clearly dominate when forecast performance is evaluated for each stock separately from an ex-post perspective; that is, based on the time-series of out-of-sample forecast errors for each stock. MIDAS models perform almost equally well. Our finding is largely in line with Ghysels et al. (2019).

5.3 Cross-sectional perspective

The model and stock specific losses L_i^j are available ex-post only. Hence, we cannot use them in the real-time portfolio selection process.[12] Instead, we will rely on cross-sectional forecast losses. That is, for each loss function and model j, we define the cross-sectional average loss in month m as:[13]

$$L_m^j = \frac{1}{n}\sum_{i=1}^{n} L^j(RV_{i,m}, \widehat{RV}^j_{i,m|m-1}). \tag{3}$$

We denote the loss of the benchmark model by L_m^B. The losses L_m^j and L_m^B can be used in real-time for the selection of models. We compute the statistic

$$LR_m^j = \frac{1}{M}\sum_{m=1}^{M} \mathbf{1}_{L_m^j/L_m^B<1}. \tag{4}$$

where $\mathbf{1}_{L_m^j/L_m^B<1}$ equals one if $L_m^j/L_m^B < 1$ and zero else. Hence, LR_m^j reports the share of months during which model j outperforms the benchmark. As the right-hand side of Table 2 shows, most model-based forecasts still beat the benchmark forecast from

[12]Of course, it is possible to compute the losses L_i^j for rolling/expanding windows of out-of-sample forecasts and select models based on this. However, given the monthly frequency of the forecasts and our sample period, model selection will be difficult due to the small sample that is used to compute a rolling/expanding window version of L_i^j.

[13]For simplicity in the notation, we assume that the number of stocks, n, in the cross-section is fixed. However, in the empirical analysis, we are assessing an unbalanced panel of stocks.

a cross-sectional perspective. However, forecast improvements are less dintinctive with LR^j_m often being around 50% for the GARCH-type models. In general, the HAR-type models are still dominant. In addition, we now report for how many months a specific model j is ranked top (denoted by $Rk^j_m \leq 1$) or among the top-4 models (denoted by $Rk^j_m \leq 4$) in terms of L^j_m. Independent of the loss function, the Panel-HAR-LR is most often the best model. That is, from a cross-sectional perspective we still find the the HAR class to be superior in forecasting 1-month ahead.

Table 2: Forecast comparison of model performance.

	Time series perspective						Cross-sectional perspective					
	SE			QLIKE			SE			QLIKE		
Model	$L^j_i < L^B_i$	$Rk^j_i = 1$	$Rk^j_i \leq 4$	$L^j_i < L^B_i$	$Rk^j_i = 1$	$Rk^j_i \leq 4$	$LR^j_m < 1$	$Rk^j_m = 1$	$Rk^j_m \leq 4$	$LR^j_m < 1$	$Rk^j_m = 1$	$Rk^j_m \leq 4$
Panel A: Model-based forecasts												
12m-RV^d	—	0.01	0.03	—	0.00	0.02	—	0.00	0.03	—	0.00	0.00
4y-RV^d	0.26	0.00	0.01	0.11	0.00	0.01	0.28	0.00	0.02	0.21	0.01	0.03
6m-RV^d	0.52	0.01	0.03	0.73	0.00	0.02	0.35	0.00	0.02	0.54	0.01	0.02
1m-RV^d	0.25	0.00	0.01	0.37	0.00	0.01	0.18	0.00	0.00	0.27	0.00	0.01
RM monthly, 12 months	0.85	0.01	0.04	0.89	0.01	0.02	0.68	0.00	0.04	0.74	0.00	0.01
RM monthly, 6 months	0.55	0.01	0.03	0.74	0.00	0.03	0.36	0.00	0.03	0.56	0.00	0.03
RM daily, 12 months	0.56	0.01	0.04	0.74	0.01	0.06	0.32	0.00	0.02	0.60	0.00	0.03
RM daily, 6 months	0.54	0.00	0.03	0.72	0.01	0.05	0.31	0.00	0.01	0.58	0.00	0.04
GJR-GARCH	0.63	0.02	0.06	0.59	0.02	0.05	0.46	0.01	0.03	0.40	0.00	0.01
Panel GJR-GARCH	0.57	0.03	0.08	0.58	0.01	0.03	0.36	0.00	0.01	0.48	0.00	0.03
Factor GARCH	0.71	0.04	0.13	0.64	0.01	0.04	0.48	0.00	0.00	0.41	0.01	0.03
Factor GARCH-VIX	0.77	0.03	0.13	0.71	0.02	0.06	0.51	0.00	0.00	0.50	0.00	0.00
Factor GARCH-ΔHous	0.76	0.01	0.08	0.71	0.00	0.04	0.51	0.00	0.01	0.49	0.00	0.00
Factor GARCH-TS	0.76	0.01	0.07	0.70	0.00	0.04	0.53	0.00	0.01	0.49	0.00	0.00
MEM	0.61	0.04	0.11	0.09	0.02	0.04	0.69	0.07	0.20	0.15	0.00	0.03
Panel MEM	0.62	0.02	0.08	0.08	0.01	0.02	0.65	0.11	0.23	0.30	0.14	0.15
HAR	0.88	0.02	0.30	0.93	0.03	0.26	0.88	0.02	0.23	0.95	0.03	0.36
HAR-LR	0.81	0.04	0.18	0.89	0.08	0.33	0.77	0.04	0.17	0.89	0.10	0.37
HAR-SPX	0.85	0.03	0.16	0.92	0.02	0.17	0.87	0.04	0.27	0.93	0.01	0.21
HAR-SPX-LR	0.78	0.03	0.14	0.87	0.04	0.28	0.75	0.06	0.28	0.86	0.05	0.27
HAR-VIX	0.88	0.07	0.29	0.93	0.12	0.41	0.87	0.08	0.39	0.95	0.07	0.33
HAR-VIX-LR	0.81	0.03	0.16	0.86	0.10	0.36	0.75	0.07	0.33	0.87	0.03	0.24
Panel HAR	0.91	0.12	0.55	0.94	0.03	0.32	0.94	0.09	0.35	0.95	0.07	0.37
Panel HAR-LR	0.94	0.29	0.65	0.96	0.30	0.65	0.97	0.21	0.59	0.97	0.26	0.55
MIDAS	0.91	0.06	0.35	0.95	0.08	0.41	0.92	0.07	0.33	0.96	0.09	0.52
Panel MIDAS	0.88	0.07	0.27	0.92	0.07	0.28	0.87	0.11	0.38	0.89	0.10	0.35
Panel B: Combined forecasts												
$\eta = 0$	0.95	—	—	0.96	—	—	0.92	—	—	0.96	—	—
$\eta = 1/2$ SE	0.95	—	—	0.97	—	—	0.93	—	—	0.95	—	—
QLIKE	0.94	—	—	0.97	—	—	0.91	—	—	0.96	—	—
$\eta = 1$ SE	0.95	—	—	0.97	—	—	0.94	—	—	0.95	—	—
QLIKE	0.94	—	—	0.97	—	—	0.92	—	—	0.96	—	—
$\eta = \infty$ SE	0.94	—	—	0.96	—	—	0.97	—	—	0.97	—	—
QLIKE	0.94	—	—	0.96	—	—	0.96	—	—	0.96	—	—

Notes: LR^j_i reports the share of losses L^j_i to be smaller than L^B_i; this is, the proportion of stocks for which the loss of the respective model j is smaller than the one of the 12m-RV^d benchmark model. $Rk^j_i \leq 1$ and $Rk^j_i \leq 4$ report the proportion of the model being the best or among the four best-performing models as measured by L^j_i. LR^j_m reports the proportion of months in which the cross-sectional loss L^j_m of model j is lower than the one of the 12m-RV^d benchmark forecast. $Rk^j_m \leq 1$ and $Rk^j_m \leq 4$ report the proportion of the model being the best or among the four best-performing models as measured by L^j_m. In Panel B we report results for combined forecasts with $\delta = 0.95$. The evaluation period is 2001:M1–2019:M12.

Thus, the differences between the time series perspective and the cross-sectional perspective could mean that the forecast performance of the different models varies over time. The latter could be the case if one model is particularly suited for a high-volatility environment while another one performs best in a low-volatility environment (Conrad and Kleen, 2020). In the following, we consider forecast combinations as a means to safeguard against such time-varying model performance. The idea to combine the forecasts from different models to achieve diversification gains was popularized by Bates and Granger (1969). For further discussions see, for example, Timmermann (2006).

We follow the approach described in Caldeira et al. (2017) for combining the forecasts of the various models. First, for each model j we determine the cross-sectional forecast performance at month m as

$$\bar{L}_m^j = \frac{1}{m} \sum_{k=0}^{m-1} \delta^k L_{m-k}^j, \tag{5}$$

with $\delta \in (0, 1]$. When δ approaches zero, we exclusively rely on the loss ratio in month m. In the other extreme, when $\delta = 1$, the forecast performance is measured by the simple average of the loss ratios over the previous m months. For $0 < \delta < 1$ all loss ratios are taken into account but the weights are declining from the most recent to the most distant observation in time. Throughout the main analysis we will set $\delta = 0.95$.

The combined forecast for the volatility of stock i, $i = 1, \dots, n$, in period $m + 1$ is given by

$$\widehat{RV}_{i,m+1|m}^{cf} = \sum_{j=1}^{J} \lambda_{j,m} \widehat{RV}_{i,m+1|m}^{j}, \tag{6}$$

where the weights are given by

$$\lambda_{j,m} = \frac{(\bar{L}_m^j)^{-\eta}}{\sum_{j=1}^{J} (\bar{L}_m^j)^{-\eta}}. \tag{7}$$

with $\eta \geq 0$. For $\eta = 0$, we attach equal weights, $\lambda_{j,m} = 1/J$, to each model. For $\eta = \infty$ a weight of one is attached to the model for which the loss in Equation (5) is the lowest and all other models receive a weight of zero. When $\eta = 1$, the weights are inverse proportional to the loss of the respective model. Note that $\eta = 1/2$ in combination with the SE means that the weights are chosen according to the root mean squared error.

Panel B of Table 2 presents the forecast performance of the combined forecasts. There is a remarkable finding which holds independently of the choice of the loss function and the choice of η: For almost all combined forecasts the statistic LR_m^{cf} is considerably higher than for most individual model-based forecasts. That is, the dominance of the combined forecasts over the benchmark forecast is much stronger than for the individual models. For $\eta = \infty$ the advantage is less striking because in each month now all weight is attached to one specific model which reduces the potential diversification gains. Thus, from a pure forecasting perspective it clearly pays off to consider the combined forecast. The finding that the loss function which is used to combine the individual model-based forecasts does not seem to matter much is highly interesting. The theoretical arguments that can be made in favor or against certain loss functions appear not being relevant in our setting. Even a simple average ($\eta = 0$) of the forecasts appears to do the job.

6 Comparison of Low-Volatility Portfolios

6.1 Portfolio construction

We illustrate the construction of the low-volatility portfolios for volatility forecasts based on model j. Assume that the volatility forecasts $\widehat{RV}^j_{i,m|m-1}$ for the n stocks in month m are already in ascending order; that is, $\widehat{RV}^j_{1,m|m-1} \leq \widehat{RV}^j_{2,m|m-1} \leq \ldots \leq \widehat{RV}^j_{n,m|m-1}$. Based on this ordering of the forecasts, the 10% stocks with the lowest volatility are included in the portfolio for month m. Those stocks receive weights inverse proportional to their forecasted volatility—in line with the construction of the S&P 500 Low Volatility Index. We denote the individual weight of stock i w.r.t. forecasts from model j by

$$w^j_{i,m} \propto \frac{1}{\sqrt{\widehat{RV}^j_{i,m|m-1}}}.$$

The weights are standardized such that the sum of all weights is equal to 1. All remaining stocks that are not in the lowest volatility decile receive a weight of zero.

Mitigating transaction costs can be achieved by smoothing the weights. There are several approaches in the literature that aim at optimal smoothing (e.g., Bertsimas and Lo, 1998; Engle and Ferstenberg, 2007; Obizhaeva and Wang, 2013) but they are typically hard to implement in practice. As a consequence, we follow the straightforward partial adjustment strategy used in Gârleanu and Pedersen (2013) and Bollerslev et al. (2018, 2022); that is, given the forecasted weight $w^F_{i,t}$, we employ

$$\tilde{w}^j_m = \nu \; \tilde{w}^j_{m-1} + (1 - \nu) \; w^j_m \tag{8}$$

with partial-adjustment rate ν. We do not optimize the smoothing parameter ν but employ the default value of $\nu = 0.95$ from Bollerslev et al. (2022). Note that partial adjustment leads to investments which assign positive weights to stocks outside of the lowest volatility decile.

When constructing low-volatility portfolios, the decision whether a particular stock is included in the portfolio or not solely depends on the ascending ordering of the forecasted volatilities of all stocks. Hence, for correctly mimicking the oracle portfolio it is not necessary to perfectly forecast volatility. All that matters is an accurate ranking of the stocks' volatilities. However, a perfect forecast leads to an accurate ranking. Hence, we conjecture that volatility models which provide more accurate forecasts should also deliver a more accurate ranking of the volatilities.

6.2 ARVol and oracle overlap

For each low-volatility portfolio the column denoted ARVol in Table 3 shows the time-series mean of the average cross-sectional volatility in each month. The left side of Table 3 corresponds to the volatility-forecast-implied weights and the right side of Table 3 corresponds to the weights with partial adjustment (see Eq. (8)). Recall that the oracle portfolio has an ARVol of 16.80%. Among the model-based portfolios, the HAR-based portfolios achieve the lowest ARVol. The best-performing model is the Panel-HAR-LR with an ARVol of 18.06%. All loss-based portfolios (i.e., the portfolios which are based on the combined forecasts) achieve ARVols that are in the range of the best-performing models. The ARVol of the 12m-RV^d benchmark is 21.06%. When testing for significant differences in the time series of ARVol between the benchmark portfolio and the different model-based portfolio, we observe that most models not only imply lower ARVol but that this difference is also statistically significant with p-values below 0.01.

The column denoted OO in Table 3 shows the oracle overlap of the low-volatility portfolios that are either based on the volatility forecasts of a single model (Panel A) or the combined forecasts (Panel B). The benchmark portfolio has an average overlap of 44.62%. The ex-post comparison with the oracle portfolio in month $m + 1$ shows that on average the decision was correct for 64.32% of the weights. Panel A shows that the GARCH-based portfolios generally do not improve much upon the benchmark portfolio with the exception of the two MEM specifications. The improved forecast performance of the combined forecasts leads to improvements in the oracle overlap of up to 24pp relative to the benchmark portfolio.

6.3 Portfolio turnover and transaction costs

As we are interested in measuring the actual performance of our low-volatility portfolios, we need to take into account the accruing transaction costs when implementing

Table 3: Portfolio characteristics.

		Volatility-weighted				Smoothed volatility-weighted			
		ARVol	OO	OD	TO	ARVol	OO	OD	TO
Panel A: Model-based portfolios									
12m-RV^d		21.06	44.62	0.21	15.82	24.58	42.42	0.20	5.58
4y-RV^d		21.65***	43.42	0.21	7.82	24.96***	40.49	0.21	4.94
6m-RV^d		20.92***	46.05	0.20	27.96	24.49	42.93	0.19	6.16
1m-RV^d		21.30***	46.21	0.19	108.01	25.78***	41.01	0.18	8.37
RM monthly, 12 months		21.00***	44.74	0.21	16.00	24.50***	42.54	0.19	5.57
RM monthly, 6 months		20.89***	46.16	0.20	28.00	24.48	43.00	0.19	6.16
RM daily, 12 months		20.89**	46.16	0.20	48.21	24.61**	42.86	0.19	6.50
RM daily, 6 months		20.85***	46.44	0.20	49.90	24.62**	42.83	0.19	6.63
GJR-GARCH		20.87***	45.73	0.20	46.14	24.60	42.11	0.19	5.85
Panel GJR-GARCH		20.98	45.17	0.20	37.32	24.54	41.96	0.20	5.31
Factor GARCH		21.18	44.27	0.21	30.81	24.61	41.47	0.20	5.38
Factor GARCH-VIX		21.42***	42.95	0.21	33.81	25.03***	40.97	0.20	5.47
Factor GARCH-ΔHous		21.38***	43.04	0.21	34.46	24.92***	41.05	0.20	5.46
Factor GARCH-TS		21.40***	43.01	0.21	34.17	24.90***	41.05	0.20	5.46
MEM		19.89***	53.71	0.19	36.90	22.78***	49.08	0.18	6.31
Panel MEM		18.51***	62.28	0.17	40.36	21.44***	54.89	0.17	5.83
HAR		18.44***	64.89	0.16	32.24	21.09***	56.97	0.17	5.69
HAR-LR		21.74	63.84	0.17	36.40	22.30***	56.67	0.17	6.12
HAR-SPX		19.25***	63.48	0.17	31.42	21.40***	56.30	0.17	5.70
HAR-SPX-LR		22.48	61.83	0.17	37.82	22.60***	55.76	0.17	6.19
HAR-VIX		18.94***	63.46	0.17	29.96	21.56***	55.96	0.17	5.67
HAR-VIX-LR		22.72	61.63	0.17	36.93	22.77***	55.34	0.17	6.16
Panel HAR		18.12***	65.51	0.16	30.65	20.96***	56.82	0.17	5.59
Panel HAR-LR		18.06***	66.29	0.16	26.23	20.89***	57.22	0.17	5.61
MIDAS		18.22***	65.21	0.16	30.59	20.84***	57.22	0.17	5.71
Panel MIDAS		18.34***	64.63	0.16	30.88	21.07***	56.47	0.17	5.63
Panel B: Loss-based portfolios									
$\eta = 0$		19.33***	54.75	0.19	27.39	22.80***	49.28	0.19	5.20
$\eta = 1/2$	SE	19.28***	55.34	0.19	27.08	22.68***	49.81	0.18	5.20
	QLIKE	19.24***	55.76	0.18	27.39	22.64***	50.12	0.18	5.22
$\eta = 1$	SE	19.22***	56.11	0.18	26.80	22.53***	50.40	0.18	5.20
	QLIKE	19.09***	57.09	0.18	27.14	22.39***	51.11	0.18	5.21
$\eta = \infty$	SE	18.06***	66.29	0.16	26.23	20.89***	57.22	0.17	5.61
	QLIKE	18.07***	66.34	0.16	28.61	20.89***	57.40	0.17	5.67

Notes: Summary measures of the model-based and loss-based portfolios are reported. ARVol is the square-root of the time-averaged "average realized variance" which is defined to be the cross-sectional average RV inside the corresponding low-volatility portfolio, $ARVol = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{N} w_{i,m} RV_{i,m}}$. Annualized scale. The true positive (TP) ratio is the average share of ex-post oracle stocks that are included in the respective portfolio. Statistically significant different ARVol relative to the 12m-RV^d benchmark model at the 10%, 5%, and 1% are indicated by *, **, and *** respectively. The test statistics are based on Newey-West covariance estimates with 3 lags. Oracle overlap (OO) is the share of equally allocated weight relative to the volatility-weighted oracle portfolio. OD is the average RMSE of the allocated weights relative to the oracle portfolio weights. For the definition of turnover (TO) see Subsection 6.3. OO, OD, and TO are reported in percentages. The evaluation period 2001:M1–2019:M12.

the strategy. We compute the turnover and the respective transaction costs following the recent literature on portfolio-allocation based on high-frequency-based measures of realized (co-)variation (Bandi et al., 2008; De Pooter et al., 2008; DeMiguel et al., 2009;

Hautsch et al., 2015; Nolte and Xu, 2015). Recall that $w^j_{i,m}$ is the weight assigned for stock i by model j at the very end of month $m-1$. Before the next rebalancing at the end of period m, due to price movements, the weight of stock i changes to $w^j_{i,m}\frac{1+R_{i,m}/100}{1+(w^j_m)'R_m/100}$ where $w^j_m = (w^j_{1,m}, \ldots, w^j_{n,m})'$ and $R_m = (R_{1,m}, \ldots, R_{n,m})'$. Based on the volatility forecasts for month $m+1$, the new desired weights are $w^j_{i,m+1}$. Hence, the turnover due to portfolio rebalancing at the end of month m is given by

$$TO^j_m = \sum_{i=1}^{N} \left| w^j_{i,m+1} - w^j_{i,m}\frac{1+R_{i,m}/100}{1+(w^j_m)'R_m/100} \right|. \tag{9}$$

The quantity TO_m can be interpreted as the proportion of wealth reallocated at the end of month m. In column TO of Table 3, we report the average turnover for each portfolio—again both for the forecast-implied and partially-adjusted portfolios. The turnover of the benchmark portfolio without smoothing is 15.82% which means that per dollar invested the average transaction volume per month is 15.82 cents. Relying on model-based forecasts increases the turnover for most models to be in the 25%–40% range. The highest turnover is observed for the naive 1m-RV^d forecast. The loss-based portfolios have a turnover of around 27%. Adjusting the weights only partially, reduces the turnover substantially. Almost all turnover numbers are in the range of only 5% which is lower than any turnover before smoothing.

6.4 Portfolio returns

Assuming transaction costs to be proportional to the portfolio turnover TO_m, we follow DeMiguel et al. (2009) and compute monthly portfolio excess returns as

$$R^j_{p,m} = \frac{W^j_m}{W^j_{m-1}} - 1 - R_{rf,m}, \tag{10}$$

where W^j_m is the wealth of the model/loss-based portfolio which can be obtained as

$$W^j_m = W^j_{m-1} \cdot (1 + w'_m R_m) \cdot (1 - c \cdot TO_m). \tag{11}$$

We assume that c is ranging from 0 to 25bps which is a realistic range of recent cost estimates for trading large US stocks (Novy-Marx and Velikov, 2016).

Table 4 shows the annualized returns of each portfolio for $c = 0$ and $c = 25$bps. When there are no transaction costs, the benchmark portfolio earns an annualized return of 9.24%. For the model-based and loss-based portfolios, we report the annualized return and test whether there is a significant difference between the return of the respective model-/loss-based portfolios and the benchmark. We indicate significant differences at the 10%, 5%, and 1% level based on Newey-West standard error by *, **, and ***. Although the returns of most of the model-/loss-based portfolios are somewhat higher than the return of the benchmark, we do not find evidence for a significant difference besides for the HAR-VIX, HAR-VIX-LR, and Panel HAR—three of the best-performing models in Subsection 5.3. Hence, even without transaction costs the superior performance of model-/loss-based volatility forecasts does not necessarily translate into higher returns. The same holds for the standard deviation of returns and the Sharpe ratios. We employ the Sharpe ratio test introduced by Ledoit and Wolf (2008) and find no significant higher Sharper ratios relative to the benchmark if we don't apply some sort of weight smoothing.

The picture changes severely if we consider partial weight adjustment. The returns of all HAR-type models are consistently 1pp higher than the average return of the benchmark. The difference is also statistically significant at the 5% or 1% significance level. We also observe Sharpe ratios that are around 10pp higher than the Sharpe ratio of the benchmark. With the reduced turnover due to partial readjustment it holds that the results with and without transaction costs are well-aligned.

6.5 Utility

We follow Fleming et al. (2001, 2003) and evaluate the various portfolios in a utility-based framework. This allows us to judge whether the differences between the benchmark and the model/loss-based portfolios are of economic significance. Using a quadratic utility function with risk-aversion parameter γ, the monthly utility generated by a portfolio based on model j is given by

$$U_\gamma(R^j_{p,m}) = (1 + R^j_{p,m}/100) - \frac{\gamma}{2(1+\gamma)} \left(1 + R^j_{p,m}/100\right)^2.$$

We are now interested in comparing this utility with the utility from the oracle portfolio. Denote the return of the oracle portfolio by $R^o_{p,m}$. We can compute the maximum fee Δ^j_γ that an investor would be willing to pay in order to switch from portfolio j to the oracle portfolio by solving

$$\sum_{m=1}^{M} U_\gamma(R^j_{p,m}) = \sum_{m=1}^{M} U_\gamma(R^o_{p,m} - \Delta^j_\gamma). \tag{12}$$

The smaller Δ_γ^j the closer the model j based portfolio mimics the utility of the oracle portfolio. We report the fee Δ_γ^j in Table 4 in annualized percentage points for $\gamma = 4$. Without smoothing, the model-/loss-based portfolios outperform the benchmark portfolio only before transaction costs in utility terms. We observe the lowest fees for the HAR-VIX-LR portfolio. Again, partial weight adjustment is the key to end up having fees for the HAR-type models that are 1pp lower than the fees for the benchmark model. This is equal to a reduction in fees of around 20%.

6.6 Sector concentration

We now examine whether our low-volatility investing strategies may generate high exposure to a small set of industries. In Figure 2 we depict histograms of the average sector concentration by primary SIC codes. We report numbers for the low-volatility oracle, the 12m-RV^d benchmark, and the SE- and QLIKE-based portfolios for the leading case with $\eta = \infty, \delta = 0.95$. For brevity, the latter two are considered to be representative for our model-based strategies. We use realtime SIC codes from the CRSP files in order to allow companies to be reassigned to a new sector. One example is S&P Global Inc., formerly McGraw-Hill Companies, for which industry classification changes from "Printing and Publishing," which is part of the "Manufacturing"-sector, to "Security and Commodity Brokers" in "Finance, Insurance, and Real Estate" after the acquisition of financial service providers like SNL Financial in April 2015 and divestures like the sale of McGraw-Hill Education in 2013.

In Figure 2, we see that 35% of the weight in the oracle portfolio belongs to stocks classified as "Manufacturing."[14] The second largest industry is "Transportation, Communications, Electric, Gas, and Sanitary service" (27%), followed by "Finance, Insurance, and Real Estate" (18%), "Services" (9.1%), "Trade" (9%), and "Mining and Construction" (0.8%). The other three histograms of our low-volatility portfolios show that the higher returns do not come at the cost of overexposure to one particular sector relative to the oracle portfolio. The largest absolute difference to the composition of the oracle portfolio can be seen in the weight on the "Transportation, Communications, Electric, Gas and Sanitary service"-sector which increases by around 7 percentage points for all feasible low-volatility portfolios. Noteworthy is also the low weight for the sector "Mining and Construction." The weight for this sector is a good example of the difference between our low-volatility portfolios and minimum-variance portfolios: The mining companies are high-risk stocks but they exhibit only low correlation with

[14]SIC code 2 and 3, see https://www.osha.gov/pls/imis/sic_manual.html.

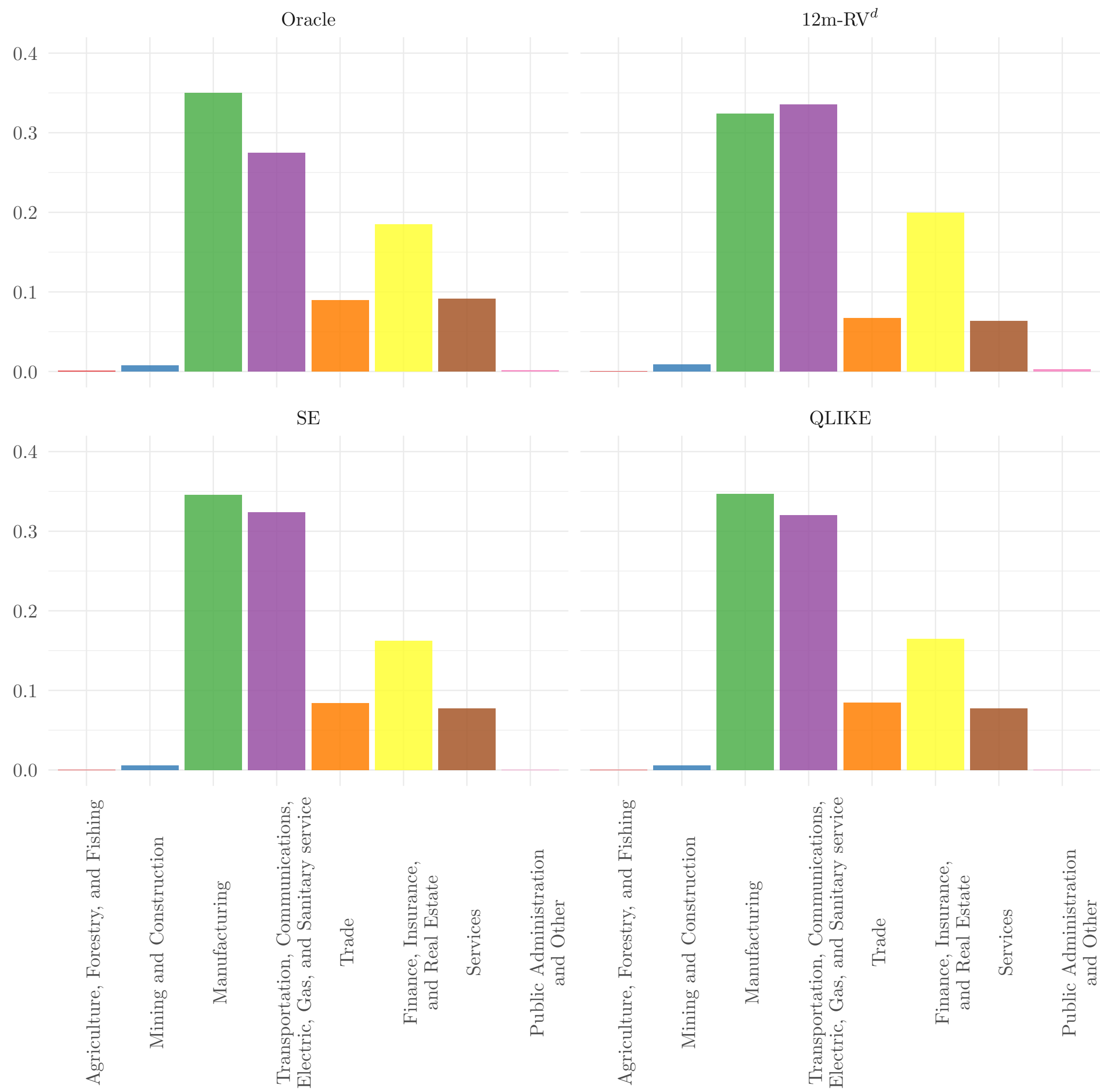
Table 4: Returns of low-volatility portfolios depending on transaction costs.

	Volatility-weighted								Smoothed volatility-weighted							
	0 bps				25 bps				0 bps				25 bps			
	Ret	Std	SR	Δ_4	Ret	Std	SR	Δ_4	Ret	Std	SR	Δ_4	Ret	Std	SR	Δ_4
Oracle	14.92	8.62	1.73	—	13.04	8.61	1.51	—	—	—	—	—	—	—	—	—
Panel A: Model-based portfolios																
12m-RV^d	9.24	10.03	0.92	6.24	8.76	10.03	0.87	4.84	9.27	11.24	0.82	6.74	9.10	11.24	0.81	5.03
4y-RV^d	9.10	10.54	0.86	6.59	8.87	10.55	0.84	4.95	9.08	11.50	0.79	7.06	8.93	11.50	0.78	5.33
6m-RV^d	9.21	9.91	0.93	6.21	8.37	9.92	0.84	5.18	9.18	11.11	0.83	6.78	8.99	11.11	0.81	5.08
1m-RV^d	8.72	9.88	0.88	6.69	5.46***	9.88	0.55***	8.07	9.26	11.60	0.80	6.93	9.00	11.60	0.78	5.30
RM monthly, 12 months	9.24	10.02	0.92	6.23	8.75	10.03	0.87	4.84	9.22	11.20	0.82	6.78	9.05	11.20	0.81	5.06
RM monthly, 6 months	9.27	9.91	0.94	6.15	8.43	9.92	0.85	5.12	9.20	11.11	0.83	6.75	9.02	11.11	0.81	5.05
RM daily, 12 months	9.49	10.05	0.94	5.99	8.04*	10.08	0.80	5.58	9.19	11.19	0.82	6.80	8.99	11.19	0.80	5.11
RM daily, 6 months	9.46	9.97	0.95	5.99	7.95*	9.99	0.80	5.63	9.11	11.20	0.81	6.88	8.91	11.20	0.80	5.20
GJR-GARCH	8.59*	10.22	0.84	6.96	7.20***	10.25	0.70***	6.49	9.20	11.27	0.82	6.82	9.03	11.27	0.80	5.12
Panel GJR-GARCH	8.72	10.27	0.85*	6.86	7.59***	10.30	0.74***	6.12	9.14	11.24	0.81	6.87	8.98	11.24	0.80	5.15
Factor GARCH	8.92	10.43	0.85	6.73	7.99**	10.44	0.76**	5.78	9.28	11.35	0.82	6.78	9.12	11.35	0.80	5.06
Factor GARCH-VIX	9.03	10.67	0.85*	6.72	8.01**	10.68	0.75***	5.86	9.24	11.59	0.80	6.94	9.08	11.59	0.78	5.23
Factor GARCH-ΔHous	9.19	10.59	0.87	6.53	8.15*	10.60	0.77***	5.69	9.25	11.54	0.80	6.91	9.09	11.54	0.79	5.19
Factor GARCH-TS	9.13	10.62	0.86*	6.60	8.10*	10.64	0.76***	5.75	9.26	11.55	0.80	6.91	9.09	11.56	0.79	5.19
MEM	10.02	11.53	0.87	6.14	8.90	11.55	0.77	5.38	10.40**	11.90	0.87	5.93	10.21**	11.90	0.86	4.24
Panel MEM	9.75	10.75	0.91	6.04	8.53	10.76	0.79	5.38	9.99**	11.15	0.90*	5.98	9.81*	11.15	0.88*	4.28
HAR	9.73	10.67	0.91	6.02	8.76	10.70	0.82	5.12	10.14**	11.08	0.92**	5.80	9.97**	11.09	0.90**	4.09
HAR-LR	10.31	11.13	0.93	5.65	9.22	11.16	0.83	4.88	10.25**	11.43	0.90*	5.85	10.07**	11.43	0.88*	4.16
HAR-SPX	9.89	10.76	0.92	5.90	8.95	10.79	0.83	4.98	10.21**	11.16	0.91**	5.77	10.03**	11.16	0.90**	4.06
HAR-SPX-LR	10.50	11.04	0.95	5.42	9.36	11.06	0.85	4.69	10.28**	11.46	0.90	5.84	10.09**	11.46	0.88	4.15
HAR-VIX	10.33**	10.62	0.97	5.40	9.42	10.64	0.89	4.44	10.33***	11.12	0.93**	5.63	10.16***	11.12	0.91**	3.92
HAR-VIX-LR	10.87**	11.10	0.98	5.08	9.75	11.11	0.88	4.32	10.28***	11.47	0.90	5.85	10.09**	11.47	0.88	4.15
Panel HAR	9.83	10.66	0.92	5.92	8.91	10.67	0.83	4.97	10.17**	11.04	0.92**	5.75	10.00**	11.04	0.91**	4.04
Panel HAR-LR	10.03*	10.58	0.95	5.68	9.24	10.59	0.87	4.60	10.19**	11.00	0.93**	5.71	10.03**	11.00	0.91**	4.00
MIDAS	9.89	10.69	0.93	5.87	8.97	10.72	0.84	4.92	10.21**	11.03	0.93**	5.70	10.04**	11.03	0.91**	3.99
Panel MIDAS	9.81	10.64	0.92	5.93	8.88	10.66	0.83	4.99	10.10**	11.10	0.91**	5.85	9.93**	11.10	0.89**	4.14
Panel B: Loss-based portfolios																
$\eta = 0$	9.44	10.13	0.93	6.07	8.62	10.15	0.85	5.03	9.36	10.91	0.86	6.50	9.21	10.91	0.84	4.77
$\eta = 1/2$ SE	9.40	10.16	0.92	6.13	8.58	10.18	0.84	5.08	9.39	10.91	0.86	6.47	9.23	10.91	0.85	4.75
QLIKE	9.44	10.13	0.93*	6.08	8.62	10.15	0.85	5.03	9.40	10.90	0.86	6.45	9.25	10.91	0.85	4.73
$\eta = 1$ SE	9.33	10.18	0.92	6.21	8.52	10.20	0.84	5.14	9.46	10.89	0.87*	6.39	9.30	10.89	0.85*	4.67
QLIKE	9.45	10.12	0.93*	6.06	8.63	10.13	0.85	5.00	9.52	10.85	0.88**	6.32	9.36	10.85	0.86**	4.59
$\eta = \infty$ SE	10.03	10.58	0.95*	5.68	9.24	10.59	0.87	4.60	10.19**	11.00	0.93**	5.71	10.03**	11.00	0.91**	4.00
QLIKE	10.05	10.55	0.95*	5.65	9.18	10.57	0.87	4.64	10.19**	11.03	0.92**	5.72	10.02**	11.03	0.91**	4.01

Notes: Average annualized excess mean return (Ret), annualized standard deviation (Std), and Sharpe Ratio (SR). Δ_γ is the annualized fee in percent an investor would be willing to pay for switching to the infeasible oracle portfolio; see Equation (12). We perform two-sided tests of equal returns using Newey-West standard errors with 3 lags against the benchmark model 12m-RV^d. Sharpe ratio test according to Ledoit and Wolf (2008). Statistical significance at the 10%, 5%, and 1% are indicated by *, **, and *** respectively. The evaluation period is 2001:M1–2019:M12.

other stocks (Blitz et al., 2019). Hence, in a minimum-variance portfolio it may be sensible to include such high-risk but low-correlation stocks in order to minimize the overall portfolio risk. However, given the histograms we can conclude that both the benchmark and our loss-based strategies do not generate excess returns on the downside of excessive sector exposure relative to other low-volatility strategies.

Figure 2: Sector concentration of smoothed oracle and feasible low-volatility portfolios.



Notes: Sector concentration by realtime Standard Industrial Classification (SIC). We report the time-average share inside each SIC class for the infeasible oracle portfolio along the corresponding numbers for three exemplary ex-ante feasible portfolios: the smoothed benchmark 12m-RV^d portfolio and the smoothed SE-based and QLIKE-based portfolios with $\eta = \infty$ and $\delta = 0.95$. Industries are classified by the first number of the SIC code as follows: "Agriculture, Forestry and Fishing" (0), "Mining and Construction" (1), "Manufacturing" (2 and 3), "Transportation, Communications, Electric, Gas and Sanitary service" (4), "Trade" (5), "Finance, Insurance, and Real Estate" (6), "Services" (7 and 8), "Public Administration and Other" (9). The evaluation period is 2001:M1–2019:M12.

6.7 Factor analysis

In Table 5, we evaluate the trading strategies of the 12m-RV^d benchmark and our smoothed loss-based portfolios for the exemplary case of $\eta = \infty$ and $\delta = 0.95$ by means of Fama-French regressions. First, we observe a statistically significant CAPM-excess return for both the benchmark portfolio and the loss-based portfolios which is around half in size relative to the total portfolio return. In the Fama-French-Carhart (FFC) four-factor model (Fama and French, 1993; Carhart, 1997), we observe a strongly significant positive coefficient for the HML portfolio returns which is in line with the observation that low volatility is correlated with value. Similarly, momentum also helps to partially explain the superior performance of the low-volatility strategies. In general, the average FFC-excess returns of our strategies are only slightly below the one for the CAPM. Turning to the Fama-French five-factor model (Fama and French, 2015), we see that the excess returns are not as good captured by value but its exposure to highly profitable but conservative investment stocks. The Fama-French five-factor model implies a reduction in monthly excess returns by around one-third. However, with values in range of 2.3–3.1% the annualized excess returns are still statistically significant with t-statistics larger than 3.4 in case of our loss-based portfolios. Using daily return of the entire CRSP cross-section but a longer evaluation period, Fama and French (2016) report similar results for total (and idiosyncratic) volatility portfolios.

Table 5: Low-volatility portfolio returns and factor loadings.

	CAPM		FFC					FF5					
	α	β_{MKT}	α	β_{MKT}	β_{SMB}	β_{HML}	β_{MOM}	α	β_{MKT}	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}
12m-RV^d	4.90	0.64	4.21	0.68	0.02	0.28	0.09	2.32	0.73	0.06	0.11	0.29	0.26
	(3.37)	(14.02)	(3.61)	(25.48)	(0.52)	(6.58)	(3.08)	(2.17)	(24.93)	(1.33)	(1.66)	(4.29)	(2.91)
SE	5.87	0.63	4.98	0.67	0.09	0.23	0.11	3.10	0.71	0.13	0.06	0.31	0.26
	(4.63)	(15.91)	(4.84)	(30.02)	(2.12)	(6.34)	(4.12)	(3.46)	(31.38)	(3.11)	(1.07)	(4.60)	(3.10)
QLIKE	5.84	0.64	4.95	0.68	0.09	0.23	0.11	3.08	0.72	0.13	0.07	0.31	0.25
	(4.64)	(15.94)	(4.86)	(30.42)	(2.26)	(6.46)	(4.13)	(3.47)	(31.74)	(3.28)	(1.15)	(4.63)	(3.07)

Notes: Regressions of the low-volatility portfolio returns in the leading case with $\eta = \infty$ and $\delta = 0.95$ on different factor models. As factors we consider the excess market return *MKT*, the size factor *SMB*, the value factor *HML* in conjunction with the momentum factor *MOM*, or the profitability factor *RMW* and the investment factor *CMA*. Regarding the factor loading coefficients, we report t-test statistics based on Newey-West covariance estimates in parentheses. Excess returns are reported on an annualized scale. The evaluation period is 2001:M1–2019:M12.

7 Conclusion

We examine the effect of employing intraday data and corresponding volatility models on long-only low-volatility investments. The portfolio choice problem at hand is to identify the bottom decile of stocks with the lowest volatility in the United States stock market. In general, the anomaly is exploited by sorting based on last year's volatility which we employ as our benchmark. However, the benchmark is at odds with the financial econometrics literature that demonstrated repeatedly the usefulness of intraday data for volatility forecasting.

First, we show that a large number of different time-series models based on intraday data have superior forecasting performance at a monthly horizon in comparison to our benchmark in the years 2001–2019. Our set of models includes Riskmetrics, numerous GARCH- and HAR-type models, and MIDAS regressions. The best-performing models belong to the class of HAR models. In general, forecast performance improves after combining model-based forecasts in real time. Our forecast evaluation is robust against using different loss functions.

Second, it is revealed that superior forecast performance translates into better assessment of the volatility ranking. This is measured both in terms of lower realized variances across stocks inside the low-volatility portfolios and a larger overlap with the infeasible oracle portfolio. Loss-based forecast combination is also beneficial in terms of similarity to the oracle portfolio. We obtain significantly higher returns after transaction costs if we employ partial weight adjustment.

References

Amaya, D., P. Christoffersen, K. Jacobs, and A. Vasquez (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics 118*(1), 135–167.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance 61*(1), 259–299.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2009). High idiosyncratic volatility and low returns: International and further U.S. evidence. *Journal of Financial Economics 91*(1), 1–23.

Asness, C., A. Frazzini, N. J. Gormsen, and L. H. Pedersen (2020). Betting against correlation: Testing theories of the low-risk effect. *Journal of Financial Economics 135*(3), 629–652.

Baker, M., B. Bradley, and J. Wurgler (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal 67*(1), 40–54.

Bali, T. G., N. Cakici, and R. F. Whitelaw (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics 99*(2), 427–446.

Bali, T. G., R. F. Engle, and S. Murray (2016). *Empirical Asset Pricing: The Cross Section of Stock Returns*. New Jersey: John Wiley & Sons.

Bandi, F. M., J. R. Russell, and Y. Zhu (2008). Using high-frequency data in dynamic portfolio choice. *Econometric Reviews 27*(1-3), 163–198.

Barberis, N. and M. Huang (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review 98*(5), 2066–2100.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(2), 253–280.

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operational Research Quarterly 20*(4), 391–410.

Bekaert, G. and M. Hoerova (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics 183*(2), 181–190.

Bertsimas, D. and A. W. Lo (1998). Optimal Control of Execution Costs for Portfolios. *Computing in Science and Engineering 1*(6), 40–53.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business 45*(3), 444–455.

Black, F., M. C. Jensen, and M. Scholes (1972). The capital asset pricing model: Some empirical tests. *Studies in the Theory of Capital Markets 81*(3), 79–121.

Blitz, D. C. and P. van Vliet (2007). The volatility effect. *Journal of Portfolio Management 34*(1), 102–113.

Blitz, D. C., P. van Vliet, and G. Baltussen (2019). The volatility effect revisited. *Journal of Portfolio Management 46*(2), 45–63.

Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen (2018). Risk everywhere: Modeling and managing volatility. *Review of Financial Studies 31*(7), 2730–2773.

Bollerslev, T., S. Z. Li, and B. Zhao (2019). Good volatility, bad volatility, and the cross section of stock returns. *Journal of Financial and Quantitative Analysis 55*(3), 751–781.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2022). Realized semibetas: Disentangling "good" and "bad" downside risks. *Journal of Financial Economics 144*(1), 227–246.

Boudt, K., G. Nguyen, and B. Peeters (2015). The low-risk anomaly revisited on high-frequency data. In G. N. Gregoriou (Ed.), *Handbook of High Frequency Trading*, pp. 397–424. Amsterdam: Elsevier.

Caldeira, J. F., G. V. Moura, F. J. Nogales, and A. A. Santos (2017). Combining multivariate volatility forecasts: An economic-based approach. *Journal of Financial Econometrics 15*(2), 247–285.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Cederburg, S., M. S. O. Doherty, F. Wang, and X. Sterling (2020). On the performance of volatility-managed portfolios. *Journal of Financial Economics 138*(1), 95–117.

Clarke, R., H. de Silva, and S. Thorley (2006). Minimum-variance portfolios in the U.S. equity market. *Journal of Portfolio Management 33*(1), 10–24.

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

Conrad, C. and K. Loch (2015). Anticipating long-term stock market volatility. *Journal of Applied Econometrics 30*(7), 1090–1114.

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

De Pooter, M., M. Martens, and D. Van Dijk (2008). Predicting the daily covariance matrix for S&P 100 stocks using intraday data - But which frequency to use? *Econometric Reviews 27*(1-3), 199–229.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies 22*(5), 1915–1953.

Engle, R. F. and R. Ferstenberg (2007). Execution Risk. *The Journal of Portfolio Management 33*(2), 34 LP – 44.

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*(1-2), 3–27.

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

Engle, R. F., V. K. Ng, and M. Rothschild (1990). Asset pricing with a Factor-ARCH covariance structure. *Journal of Econometrics 45*(1-2), 213–237.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and K. R. French (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies 29*(1), 69–103.

Fleming, J., C. Kirby, and B. Ostdiek (2001). The economic value of volatility timing. *Journal of Finance 56*(1), 329–352.

Fleming, J., C. Kirby, and B. Ostdiek (2003). The economic value of volatility timing using realized volatility. *Journal of Financial Economics 67*(3), 473–509.

Frazzini, A. and L. H. Pedersen (2014). Betting against beta. *Journal of Financial Economics 111*(1), 1–25.

Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics 91*(1), 24–37.

Gârleanu, N. and L. H. Pedersen (2013). Dynamic Trading with Predictable Returns and Transaction Costs. *Journal of Finance 68*(6), 2309–2340.

Ghysels, E., A. Plazzi, R. Valkanov, A. R. Serrano, and A. Dossani (2019). Direct versus iterated multi-period volatility forecasts: Why MIDAS is king. *Annual Review of Financial Economics 11*, 173–195.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The MIDAS touch: Mixed data sampling regression models. *CIRANO Working Papers*.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). There is a risk-return trade-off after all. *Journal of Financial Economics 76*(3), 509–548.

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

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*(5), 1779–1801.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Hansen, P. R. and A. Lunde (2006). Consistent ranking of volatility models. *Journal of Econometrics 131*(1-2), 97–121.

Haugen, R. A. and A. J. Heins (1972). On the evidence supporting the existence of risk premiums in the capital market. *Wisconsin Working Paper*.

Haugen, R. A. and A. J. Heins (1975). Risk and the rate of return on financial assets: Some old wine in new bottles. *Journal of Financial and Quantitative Analysis 10*(5), 775–784.

Hautsch, N., L. M. Kyj, and P. Malec (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics 30*(2), 263–290.

Kleen, O. (2018). mfGARCH: Mixed-frequency GARCH models. R package. Available on CRAN: https://cran.r-project.org/package=mfGARCH.

Kleen, O. and A. Tetereva (2022). A forest full of risk forecasts for managing volatility. *Working Paper*.

Ledoit, O. and M. Wolf (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance 15*(5), 850–859.

Liu, L. Y., A. J. Patton, and K. Sheppard (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics 187*(1), 293–311.

Liu, Q. (2009). On portfolio optimization: How and when do we benefit from high-frequency data? *Journal of Applied Econometrics 24*(4), 560–582.

Moreira, A. and T. Muir (2017). Volatility-managed portfolios. *Journal of Finance 72*(4), 1611–1644.

Moreira, A. and T. Muir (2019). Should long-term investors time volatility? *Journal of Financial Economics 131*(3), 507–527.

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

Nolte, I. and Q. Xu (2015). The economic value of volatility timing with realized jumps. *Journal of Empirical Finance 34*, 45–59.

Novy-Marx, R. and M. Velikov (2016). A taxonomy of anomalies and their trading costs. *Review of Financial Studies 29*(1), 104–147.

Obizhaeva, A. A. and J. Wang (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets 16*(1), 1–32.

Patton, A. J. (2006). Volatility forecast comparison using imperfect volatility proxies. Quantitative Finance Research Centre, University of Technology Sydney, Research Paper 175.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics 160*(1), 246–256.

Shumway, T. (1997). The delisting bias in CRSP data. *Journal of Finance 52*(1), 327–340.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 135–196. Amsterdam: Elsevier.

A Description of time-series models

Because months have different numbers of days, all models forecast the 22-day-ahead average realized variance which is then evaluated against the average realized variance in that month. Let $\mathcal{F}_t$ denote the information set up to time t.

HAR-type models

- **HAR**: The HAR model (Corsi, 2009) employs the realized variances directly. In this model, realized variances are regressed on past realized variances aggregated on a daily, weekly, and monthly frequency. The model for forecasting the 22-day-ahead cumulative variance is given by

$$RV_{i,t+1:t+22} = b_0 + b_d RV_{i,t} + b_w RV_{i,t-4:t} + b_m RV_{i,t-21:t} + \eta_{i,t}$$

  with $RV_{i,t+1:t+l} = \sum_{k=1}^{l} RV_{i,t+k}$ and $\mathbf{E}[\eta_{i,t}|\mathcal{F}_{t-1}] = 0$.

- **HAR-SPX:** Now, let $RV_{mkt,t}$ denote the realized variance of the S&P 500 index. Then the HAR-SPX model is the HAR model from above augmented by a HAR model forecast for the market itself,

$$RV_{i,t+1:t+22} = b_0^S + b_d^S RV_{i,t} + b_w^S RV_{i,t-4:t} + b_m^S RV_{i,t-21:t} + b_{mkt}^S \widehat{RV}_{mkt,t+1:t+22|t} + \eta_{i,t}^S$$

  with $\mathbf{E}[\eta_{i,t}^S|\mathcal{F}_{t-1}] = 0$.

- **HAR-LR:** Given that we are only interested in monthly volatility forecast, we employ a long-run version of the HAR model that includes a quarterly and semi-annual component:

$$\begin{aligned} RV_{i,t+1:t+22} = b_0^L &+ b_d^L RV_{i,t} + b_w^L RV_{i,t-4:t} + b_m^L RV_{i,t-21:t} \\ &+ b_q^L RV_{i,t-65:t} + b_s^L RV_{i,t-131:t} + \eta_{i,t}^L \end{aligned}$$

  with $\mathbf{E}[\eta_{i,t}^L|\mathcal{F}_{t-1}] = 0$.

- **HAR-SPX-LR**: As we did in the HAR-SPX, we can als define a HAR-SPX-LR model which employs both the long-run and the market component,

$$\begin{aligned} RV_{i,t+1:t+22} = b_0^{SL} &+ b_d^{SL} RV_{i,t} + b_w^{SL} RV_{i,t-4:t} + b_m^{SL} RV_{i,t-21:t} \\ &+ b_q^{SL} RV_{i,t-65:t} + b_s^{SL} RV_{i,t-131:t} + b_{mkt}^{SL} \widehat{RV}_{mkt,t+1:t+22|t} + \eta_{i,t}^{SL} \end{aligned}$$

with $\mathbf{E}[\eta^{SL}_{i,t}|\mathcal{F}_{t-1}] = 0$.

- **Panel HAR:** The HAR model can also be estimated in a panel if the individual realized variances are demeaned first. Let $\overline{RV}_i$ be the average realized variance of stock i in the estimation period. Then we estimate Panel HAR coefficients

  $$RV_{i,t+1:t+22} - \overline{RV}_i = b^P_d(RV_{i,t} - \overline{RV}_i) + b^P_w(RV_{i,t-4:t} - \overline{RV}_i) + b^P_m(RV_{i,t-21:t} - \overline{RV}_i) + \eta^P_{i,t}$$

  with $\mathbf{E}[\eta^{P}_{i,t}|\mathcal{F}_{t-1}] = 0$. For forecasting the individual stock's realized variance, we re-add $\overline{RV}_i$ in the end.

- **Panel HAR-LR**: The Panel HAR-LR model is then the long-run analogue of the Panel HAR:

  $$\begin{aligned} RV_{i,t+1:t+22} - \overline{RV}_i &= b^{PL}_d(RV_{i,t} - \overline{RV}_i) + b^{PL}_w(RV_{i,t-4:t} - \overline{RV}_i) \\ &+ b^{PL}_m(RV_{i,t-21:t} - \overline{RV}_i) + b^{PL}_q(RV_{i,t-65:t} - \overline{RV}_i) + b^{PL}_s(RV_{i,t-131:t} - \overline{RV}_i) + \eta^{PL}_{i,t} \end{aligned}$$

  with $\mathbf{E}[\eta^{PL}_{i,t}|\mathcal{F}_{t-1}] = 0$.

- **HAR-VIX**: All models above are only backward-looking time series models and make no use of expectations on future volatility; for example, those implied by option prices. Hence, we include the squared VIX as a model-free risk-neutral measure of next-month's volatility of market returns,

  $$RV_{i,t+1:t+22|t} = b^V_0 + b^V_d RV_{i,t} + b^V_w RV_{i,t-4:t} + b^V_m RV_{i,t-21:t} + b_{\text{vix}} VIX^2_t + \eta^V_{i,t}.$$

  with $\mathbf{E}[\eta^{V}_{i,t}|\mathcal{F}_{t-1}] = 0$. Bekaert and Hoerova (2014) use the same approach for forecasting aggregate stock market volatility instead of individual stocks. Of course, one could derive individual option-implied volatilities from each stock's option prices but that is beyond the scope of this paper.

- **HAR-VIX-LR**: The HAR-VIX model may also be augmented by our two long-run components:

  $$\begin{aligned} RV_{i,t+1:t+22|t} =& b^{VL}_0 + b^{VL}_d RV_{i,t} + b^{VL}_w RV_{i,t-4:t} + b^{VL}_m RV_{i,t-21:t} \\ &+ b^{VL}_q RV_{i,t-66:t} + b^{VL}_s RV_{i,t-132:t} + b^{VL}_{\text{vix}} VIX^2_t + \eta^{VL}_{i,t}. \end{aligned}$$

  with $\mathbf{E}[\eta^{VL}_{i,t}|\mathcal{F}_{t-1}] = 0$.

All HAR models are estimated by ordinary least squares estimation.

GARCH-type models

Let $\varepsilon_{mkt,t}$ and $\varepsilon_{i,t}$ denote the demeaned market and individual stock log returns. Likewise, let $\bar{\sigma}^2_{mkt}$ and $\bar{\sigma}^2_i$ denote the empirical variances of the two in the corresponding estimation sample.

- **GJR-GARCH:** The GARCH specification of Glosten et al. (1993) of returns $\varepsilon_{i,t} = \sqrt{h^{GJR}_{i,t}} Z^{GJR}_{i,t}$, $Z^{GJR}_{i,t} \sim \mathcal{D}(0,1)$, is given by

  $$h^{GJR}_{i,t} = (1-\alpha^{GJR}_i - \beta^{GJR}_i - \gamma^{GJR}_i/2)\bar{\sigma}^2_i + \alpha^{GJR}_i \varepsilon^2_{i,t-1} + \gamma^{GJR}_i \mathbb{1}_{\{\varepsilon_{i,t-1}<0\}} \varepsilon^2_{i,t-1} + \beta^{GJR}_i h^{GJR}_{i,t-1}.$$

  We determine the rolling-window coefficients by quasi-maximum-likelihood estimation (QMLE).

- **Panel GJR-GARCH:** Instead of estimating the GARCH coefficients for every stock separately, we can estimate a Panel GJR-GARCH in which

  $$\begin{aligned} h^{PGJR}_{i,t} = (1 - \alpha^{PGJR} - \beta^{PGJR} - \gamma^{PGJR}/2)\bar{\sigma}^2_i + \alpha^{PGJR} \varepsilon^2_{i,t-1} \\ + \gamma^{PGJR} \mathbb{1}_{\{\varepsilon_{i,t-1}<0\}} \varepsilon^2_{i,t-1} + \beta^{PGJR} h^{PGJR}_{i,t-1}. \end{aligned}$$

  Under the assumption of the innovation terms being independent, the Panel GJR-GARCH is estimated via QMLE by summing up the individual log-likelihoods.

- **Factor GARCH:** In this model introduced by Engle et al. (1990), the market return is modeled as a GJR-GARCH,

  $$\varepsilon_{mkt,t} = \sqrt{h^{CG}_{mkt,t}} Z^{CG}_{mkt,t}$$

  with $Z_{mkt,t} \sim \mathcal{D}(0,1)$ and

  $$\begin{aligned} h_{mkt,t} = (1 - \alpha^{CG}_{mkt} - \beta^{CG}_{mkt} - \gamma^{CG}_{mkt}/2)\bar{\sigma}^2_{mkt} + \alpha^{CG}_{mkt} \varepsilon^2_{mkt,t-1} \\ + \gamma^{CG}_{mkt} \mathbb{1}_{\{\varepsilon_{mkt,t-1}<0\}} r^2_{mkt,t-1} + \beta^{CG}_{mkt} h^{CG}_{mkt,t-1}. \end{aligned}$$

  The individual demeaned stock return is given by

  $$\varepsilon_{i,t} = \beta^{CG}_i r_{mkt,t} + \eta^{CG}_{i,t} = \beta^{CG}_i r_{mkt,t} + \sqrt{h^{CG}_{i,t}} Z^{CG}_{i,t}$$

with $Z_{i,t}^{CG} \sim \mathcal{D}(0,1)$ and

$$h_{i,t}^{CG} = (1 - \alpha_i^{CG} - \beta_i^{CG})\bar{\omega}_i + \alpha_i^{CG}\eta_{i,t-1}^2 + \beta_i h_{i,t-1}^{CG},$$

where $\bar{\omega}_i$ denotes the empirical variance of the stock-specific CAPM residuals. Under the assumption of independence of $Z_{\text{mkt},t}^{CG}$ and $Z_{i,t}^{CG}$, the forecast of the individual stock's conditional variance is given by

$$\left(\beta_i^{CG}\right)^2 h_{\text{mkt},t+1:t+22|t}^{CG} + h_{i,t+1:t+22|t}^{CG}$$

where $h_{\text{mkt},t+1:t+22|t}^{CG}$ and $h_{i,t+1:t+22|t}^{CG}$ are the cumulated daily GARCH forecasts. The β_i^{CG}s are estimated separately for each stock in the respective rolling window as well as the GARCH models for the market and the CAPM-residuals.

- The **Factor GARCH-MIDAS** model is the same as the CAPM GARCH model but the market return is now given by a GARCH-MIDAS model. It includes either the VIX, changes in housing starts, or the term spread as a covariate and estimation has been carried out using QMLE, see Engle et al. (2013), using the R-package *mfGARCH* by Kleen (2018).

  More specifically, the standardized demeaned market return $\varepsilon_{mkt,t}$ is now modeled as

$$\frac{\varepsilon_{mkt,t}}{\sqrt{\tau_m}} = \sqrt{g_{mkt,t}} Z_{mkt,t},$$

  where τ_m is specified as a function of a monthly explanatory variable X_m, $g_{mkt,t}$ follows a daily GARCH equation, and $Z_{mkt,t}$ is an *i.i.d.* innovation process with mean zero and variance one. The short-term component is assumed to follow a mean-reverting unit-variance GJR-GARCH process:

$$\begin{aligned} g_{mkt,t} = {} & (1 - \alpha^{CGM} - \gamma^{CGM}/2 - \beta^{CGM}) \\ & + \left(\alpha^{CGM} + \gamma^{CGM} \mathbb{1}_{\{\varepsilon_{mkt,t-1}<0\}}\right) \frac{\varepsilon_{mkt,t-1}^2}{\tau_m} + \beta^{CGM} g_{mkt,t-1}. \end{aligned}$$

  The long-term component τ_m in month m is given by

$$\tau_m = \exp\left(m^{CGM} + \theta^{CGM} \sum_{l=1}^{K} \varphi_l(w_1^{CGM}, w_2^{CGM}) X_{m-l}\right).$$

where the weights $\varphi_l(w_1, w_2) \geq 0$ are parameterized via the Beta weighting scheme

$$\varphi_l(w_1, w_2) = \frac{(l/(K+1))^{w_1-1} \cdot (1 - l/(K+1))^{w_2-1}}{\sum_{j=1}^{K} (j/(K+1))^{w_1-1} \cdot (1 - j/(K+1))^{w_2-1}}. \tag{13}$$

In our versions with either changes in housing starts or the term spread as the explanatory variable X_m, we choose $K = 36$. In case of the VIX, we choose $K = 3$. For more details see Conrad and Kleen (2020). We name our Factor GARCH-MIDAS models accordingly to the covariate employed: *Factor GARCH-VIX*, *Factor GARCH-*Δ*Hous*, and *Factor GARCH-TS*.

MIDAS-type models

- **MIDAS:** The class of MIDAS models was introduced by Ghysels et al. (2004, 2005, 2006) which are very flexible distributed lag models that potentially employ data sampled on different frequencies (see the CAPM GARCH-MIDAS above). In our case, it is defined as

$$RV_{i,t+1:t+22|t} - \overline{RV}_i = \theta_i^M \sum_{l=0}^{K-1} \varphi_l(1, w_{i,2}^M) \cdot (RV_{i,t-l} - \overline{RV}_i) + \eta_{i,t}^M.$$

  The weighting scheme is a Beta weighting scheme as in Equation (13) with $w_1 = 1$ and we choose $K = 132$ to match the long-run HAR models. We assume $\mathbf{E}[\eta_{i,t}^M | \mathcal{F}_{t-1}] = 0$. The parameters are obtained by minimizing the squared residuals.

- **Panel MIDAS:** Similar to our other panel variations for HAR and GARCH models, we include a Panel MIDAS by restricting the scaling parameter θ_i^M and the weighting parameter $w_{i,2}^M$ to be the same for all stocks,

$$RV_{i,t+1:t+22|t} - \overline{RV}_i = \theta^{PM} \sum_{l=0}^{K-1} \varphi_l(1, w_2^{PM}) \cdot (RV_{i,t-l} - \overline{RV}_i) + \eta_{i,t}^{PM}.$$

  We assume $\mathbf{E}[\eta_{i,t}^{PM} | \mathcal{F}_{t-1}] = 0$. This is again estimated by minimizing the squared residuals.

Riskmetrics

Our Riskmetrics forecasts are based either on monthly (indexed by m) or daily data (indexed by t). In total we employ four different versions. The first is *RM monthly, 12* *months* and the forecasts are given by

$$RV^d_{m+1|m} = \frac{1}{\sum_{k=0}^{K-1} \lambda^k} \sum_{k=0}^{K-1} \lambda^k RV^d_{m-k}$$

with $K = 12$ and RV^d_m being the realized variance in month m based on squared daily returns. *RM monthly, 6 months* is the same but with $K = 6$. *RM daily, 12 months, and RM daily, 6 months* are similar but they use daily squared returns on the right hand side with the corresponding number of lags to match the data of the monthly RM models. We choose $\lambda = 0.97$ because we target the monthly horizon.

All models are reestimated at the end of each month. In a handful of cases, the forecast is unreasonable (e.g., negative for some stocks in the Panel HAR model). Thus, we apply a rolling "sanity filter" which truncates forecasts by one-third of the 1%- and three times the 99%-quantile of the stock-specific monthly RVs in the estimation window.