Volatility forecasting for low-volatility investing

Christian Conrad*[1], Onno Kleen†[2,3], and Rasmus Lönn‡[2,4]

[1]Heidelberg University

[2]Erasmus University Rotterdam

[3]Tinbergen Institute

[4]Erasmus Research Institute of Management

November 29, 2023

Abstract

Low-volatility investing often involves sorting and selecting stocks based on retrospective risk measures, for example, the historical standard deviation of returns. In this paper, we use the volatility forecasts from a wide spectrum of volatility models to sort and select stocks and estimate portfolio weights. Our portfolios are more closely aligned with the ex-post optimal portfolio and deliver large, significant economic gains compared to traditional benchmarks after transaction costs. Importantly, we find that choosing portfolio weights by optimally combining the volatility forecasts from the different models delivers the strongest forecast and financial performance in real time.

Keywords: Factor investing, low-volatility allocations, volatility forecasts

JEL: C22, C55, C58, G11

*christian.conrad@awi.uni-heidelberg.de
†Corresponding author: kleen@ese.eur.nl
‡lonn@ese.eur.nl

1 Introduction

Low-risk investment strategies are prominent in the financial industry and well-motivated by the low-volatility anomaly—the consistent empirical evidence that points towards an inverse relationship between (idiosyncratic) volatility and expected returns (e.g. Ang et al. (2006); Blitz et al. (2019) among many others). Low-volatility portfolios are straightforward investment strategies that only require estimates of each stock's volatility and no estimates of correlations. Two prominent examples are the S&P 500 Low Volatility Index which is based on the 100 least volatile stocks in the S&P 500 and the S&P 500 Low Volatility Top 80% Index that is based on the 400 least volatile stocks in the S&P 500.[1] These indices assign a weight to each stock inversely proportional to the respective stock's volatility, measured by the standard deviation of daily returns over the preceding year. The methodology of the S&P 500 Low Volatility Top 80% Index is motivated by the empirical observation that the relation between volatility and returns is flat for low- and medium-volatility stocks, but negative for the stocks with the highest volatility. Hence, it pays off to avoid investing in high-volatility stocks (Gu et al., 2020; Blitz et al., 2019).

Nonetheless, from an ex-ante perspective, it remains an open question whether the past standard deviation or volatility forecasts, e.g., based on high-frequency data, should be used to select stocks and estimate portfolio weights. Our paper focuses on improving low-volatility investing through a comprehensive forecast evaluation of a large set of models and forecast combinations in a large cross section of assets. The aim is to provide methodological improvements in the implementation of low-volatility portfolios with monthly re-balancing. We take a new perspective on low-volatility allocations by considering an infeasible post-hoc portfolio as the benchmark. We define the post-hoc portfolio as the portfolio that an investor would choose with perfect foresight of each stock's monthly volatility.

In the financial econometrics literature (see, e.g., Ghysels et al., 2019), it is common to evaluate model forecast performance and to select the best model ex-post based on the out-of-sample forecast performance for individual stocks. Due to the monthly forecast horizon and a small sample,

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

this is not feasible in our setting. Instead, we select the best models and forecast combinations in each month based on the forecast performance in the cross section of stocks. Thus, our approach is implementable in real-time. We then assess which model-based volatility forecasts facilitate constructing low-volatility portfolios that get as close as possible to the post-hoc portfolio. By treating the allocation as a prediction problem of the post-hoc portfolio, we link low-volatility portfolios to the literature on volatility forecasting. We investigate whether state-of-the-art volatility models are useful for anticipating the correct composition of the post-hoc portfolio.

In the literature on low-volatility portfolios, volatility is typically measured using the standard deviation of monthly or daily stock returns over a specific period (e.g., previous year, six months, or month). Bali et al. (2016) provide an overview of commonly used metrics. Instead, following the literature on estimating volatility from high-frequency intraday return data (e.g., Andersen et al., 2003), we use the realized volatility to measure monthly volatility ex-post. To forecast volatility, we employ a wide range of time series models. First, we use simple RiskMetrics models and various generalized autoregressive conditional heteroskedasticity (GARCH)-type models (see, e.g., Glosten et al., 1993). Those models treat the conditional variance as a latent process, and daily (or monthly) returns are used for estimating and forecasting volatilities. Second, we use heterogeneous autoregression (HAR) models (Corsi, 2009) and mixed-frequency data sampling (MIDAS)-type models (Ghysels et al., 2004). Here, the realized variances are modeled directly as a function of past realized variances. Third, we consider forecast combinations; that is, we combine the forecasts from various volatility models according to measures of past forecast performance.

Our empirical application on large US equities suggests that portfolios based on forecast combinations more closely resemble the post-hoc portfolio than those based on individual models. In this regard, our results examining a larger cross-section of assets and with a target horizon of one month, are in line with previous research showing that forecast combinations outperform individual forecasts. Among the different models, those that employ direct forecasting via realized variances outperform all other models. However, our findings provide an important contrast to the broader literature regarding forecast combinations. Whereas it is often found that it is hard to outperform forecast combinations that use equal weights (see, e.g., Stock and Watson, 2004; Smith and Wallis, 2009; Claeskens et al., 2016), we find that the most effective forecast combination relies on model selection, i.e., in each period places all weight on the currently best-performing model. This is optimal because the relative forecasting accuracy across models varies over time, but differences in forecast accuracy are persistent enough to be exploitable in real time.

The out-of-sample performance of the low-volatility portfolios based on state-of-the-art volatility models significantly improves performance relative to simple benchmarks. These findings are robust to the inclusion of estimated effective trading costs. We compute the annualized fee required to make an investor indifferent between the post-hoc allocation and the feasible trading strategies. Comparing the forecast combinations to the simple benchmark allocations, we find differences in fees are as large as one percentage point, annualized, net of transaction costs. This corresponds to a reduction of almost 40 percent. The appraisal ratio associated with the low-volatility portfolios, for an investor that is holding the five Fama-French factors, more than doubles using effective forecast combinations instead of the benchmarks based on daily returns. Our results show that volatility forecasting within a large cross-section of stocks can bring important economic gains. Model forecasting performance varies strongly over time, but model selection can uncover the best models with sufficient accuracy.

The remainder of the paper is organized as follows. Section 2 reviews the previous literature and presents empirical evidence for the low-volatility anomaly. Section 4 introduces the volatility models and Section 5 describes the data. The forecast performance of the volatility models is evaluated in Section 6. Sections 7 and 8 provide a comparison of the various low-volatility portfolios. Section 9 concludes.

2 Related literature

The anomaly that realized returns appear inversely related to volatility is well-documented in the financial literature. Haugen and Heins (1972) and Haugen and Heins (1975) make use of total volatility while (Ang et al., 2006) and Ang et al. (2009) study allocations based on idiosyncratic volatility. Gu et al. (2020) demonstrate that total volatility can be used to predict returns through the use of machine-learning techniques. Likewise, there is a literature that studies risk defined as beta (Haugen and Heins, 1972, 1975; Frazzini and Pedersen, 2014). All of these findings are related as high-beta stocks are typically high-volatility stocks, and total volatility is highly correlated with idiosyncratic volatility (Baker et al., 2011; Bali et al., 2016; Blitz et al., 2019).

Several explanations have been proposed for the low-volatility anomaly. Examples are that investors face leverage constraints (e.g., Frazzini and Pedersen, 2014), regulatory constraints, or constraints on short-selling (see for example Blitz et al. (2014), for an overview). Behavioral explanations include representativeness, overconfidence, or preferences for lottery-like stocks (Barberis and Huang, 2008; Bali et al., 2011; Baker et al., 2011). Asness et al. (2020) find evidence supporting both the leverage and the lottery hypothesis.

Within the Markowitz framework (Markowitz, 1952), the minimum variance allocation often proves to be an empirically robust alternative to the mean-variance alternatives that require estimates of the expected returns. Clarke et al. (2006) explores the comparison on US equities and DeMiguel et al. (2009) find that the global minimum variance allocation tends to generate more stable allocations with higher out-of-sample performance, even with respect to Sharpe ratios. Kirby and Ostdiek (2012) further extend the analysis to include, among else, volatility timing portfolios. They find that these simple allocation rules provide effective active strategies that provide net Sharpe ratios above simple equally weighted alternatives.

In this paper, we take a forward-looking perspective on the inverse volatility allocations by utilizing time-series models that have been widely documented to perform better than trailing volatility measures. Ghysels et al. (2005) employ MIDAS models to derive variance forecasts for the market, providing evidence for a positive relationship between risk and return. Similarly, Fu (2009) employs the exponential generalized autoregressive conditional heteroskedasticity model by Nelson (1991) to forecast idiosyncratic volatilities, which he finds to be positively correlated with returns, contradicting the findings of Ang et al. (2006, 2009). While Ghysels et al. (2005), Fu

(2009), and Gu et al. (2020) demonstrate the usefulness of time-series models for portfolio sorting, their analyses are limited to using daily return data, excluding forecasting models based on realized variances.

We contribute to this literature by providing a comprehensive study of a large spectrum of volatility models based on low-frequency observations as well as intraday returns. In addition, we consider a rich set of forecast combinations to possibly enhance forecasting accuracy further. The literature on intraday data for variance-based portfolio sorting follows the simple trailing volatility approach. Boudt et al. (2015) perform a study similar to ours.[2] However, Boudt et al. (2015) find that there is no (statistically significant) benefit in portfolio returns from using intraday data. In contrast to our study, they do not use volatility models and have a smaller sample. 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.

3 Volatility forecasting and inverse volatility allocation

In this section, we take a new perspective on low-volatility allocations by creating and evaluating the performance of an ex-ante infeasible post-hoc portfolio. As discussed in more detail in Section 5, we use daily stock price data from the Center of Research in Security Prices (CRSP) and combine it with intraday data from the New York Stock Exchange TAQ database. Each month, we use the 500 stocks that are the largest as measured by market capitalization. In total, our analysis includes 1616 stocks because the cross-section is time-varying between 2005:M1 and 2021:M12.

First, we provide evidence that the low-volatility anomaly is much stronger from an ex-post

[2] 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).

Table 1: Returns and Sharpe ratios of equally-weighted quintile portfolios

Quintile		1	2	3	4	5
Trailing volatility sort	Ret	10.31	10.52	10.44	10.78	9.77
	SR	0.98	0.82	0.72	0.67	0.54
Post-hoc volatility sort	Ret	14.93	12.59	12.46	9.71	2.15
	SR	1.45	1.02	0.87	0.62	0.22

Notes: The annualized compounded returns and corresponding Sharpe ratios are calculated based on monthly returns of volatility-sorted portfolios. In each month, we sort stocks either based on a trailing volatility proxy (volatility of 12 months of daily return data) or we sort stocks based on the ex-post measure of monthly volatility calculated using intraday data. The investment universe is the set of 500 largest U.S. common stocks by market capitalization in the previous month. The sample period is 2005:M1 to 2021:M12.

than an ex-ante perspective. We denote the monthly realized variance of stock i, $i = 1, \dots, N$, in month m by $RV_{i,m}$. The monthly realized variance equals the sum of daily realized variances based on intraday data (see Section 5). The previous literature on low-volatility portfolios has focused on volatility measures based on daily data. We follow the construction of the S&P 500 Low Volatility Index and measure volatility of stock i in month m by the standard deviation of daily returns over the last twelve months (i.e., 252 trading days) which we denote by $\text{12m-RV}^d_{i,m}$.[3]

From an ex-ante perspective, the low-volatility anomaly can be illustrated by sorting all stocks in month m according to the ascending ordering of $\text{12m-RV}^d_{i,m}$, forming quintile portfolios and computing returns in month $m + 1$. We refer to this approach as *trailing volatility sort*. Table 1 shows those portfolios' average monthly returns and Sharpe ratios. Clearly, the average returns and Sharpe ratios of the 20% stocks with the highest volatility are the lowest. Next, we take an ex-post perspective. We think of a hypothetical investor who constructs infeasible quintile portfolios that are formed at the end of month m according to the realized volatility from the end of month $m+1$. We refer to this infeasible approach as *post-hoc volatility sort*. As Table 1 shows, the low-volatility anomaly is much more pronounced from an ex-post perspective. Now, average return and Sharpe ratios decline smoothly from the first to the fourth quintile portfolio and then drop dramatically for the fifth quintile portfolio.

[3]Alternatively, volatility might be measured using returns over the previous months (e.g., Ang et al., 2006, 2009) or the previous three years (e.g., Blitz and van Vliet, 2007).

Motivated by these observations, we follow the methodology of the S&P 500 Low Volatility Top 80% Index for the remainder of this paper; that is, we exclude the top 20% volatility stocks from our portfolio allocation and attach weights inversely proportional to an asset's volatility to the remaining assets. We refer to the feasible portfolio that is based on the 80% stocks with the lowest volatility as measured ex-ante by 12m-$\text{RV}^d_{i,m}$ by 12m-RV^d. Because it employs the same sorting and weighting variables, the 12m-RV^d portfolio has a correlation of 98.7% with the S&P 500 Low Volatility Top 80% Index. The small remaining difference is due to the fact that our investment universe is slightly different (largest 500 common stocks vs. S&P 500 constituents) and that we have a different rebalancing frequency (monthly vs. quarterly). For comparison, we construct the hypothetical portfolio that is based on the 80% stocks with the lowest $RV_{i,m+1}$ and choose portfolio weights accordingly. We refer to this infeasible portfolio as *post-hoc portfolio*.

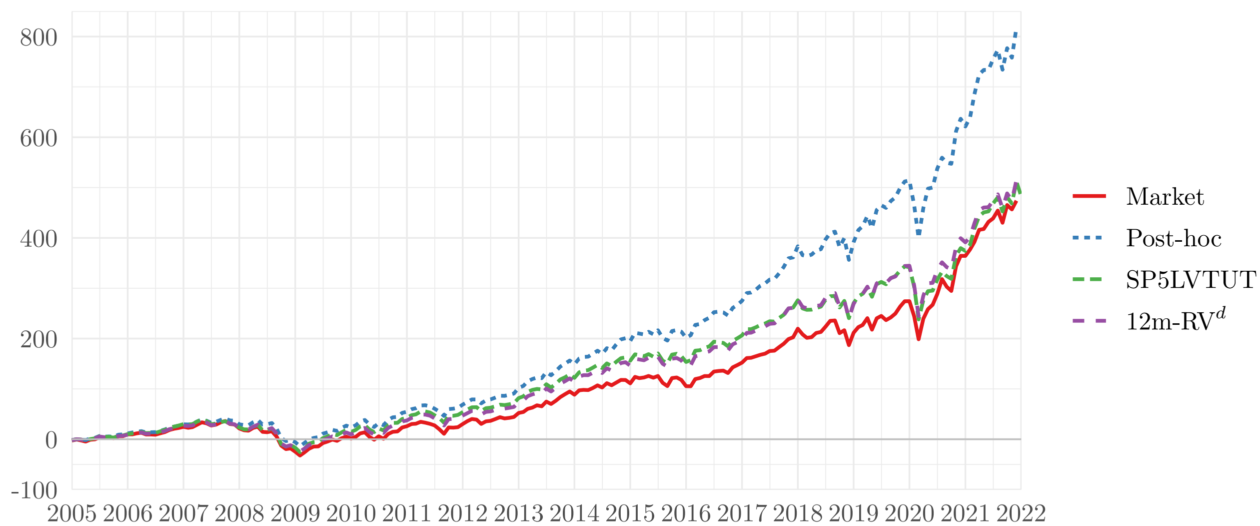
Figure 1 shows the cumulative return of the post-hoc (blue), the 12m-RV^d portfolio (purple), and the CRSP value-weighted market return (red) during the 2005 to 2021 period. For comparison, the figure also depicts the return of the S&P 500 Low Volatility Top 80% Index[4] in green. Over the whole sample period, the post-hoc portfolio clearly outperforms the 12m-RV^d portfolio and the market portfolio. As expected, the returns of the 12m-RV^d portfolio and the S&P 500 Low Volatility Top 80% Index are almost identical.

Table 2 presents summary statistics of the post-hoc portfolio, the 12m-RV^d portfolio along with three feasible alternatives, which use volatility forecasts based on the daily returns over the previous 4 years (4y-RV^d), the previous six months (6m-RV^d) and the previous month (1m-RV^d). Table 2 shows that the post-hoc portfolio has a higher return, a lower volatility and, hence, a considerably higher Sharpe ratio than the four feasible low-volatility portfolios.

An investor pursuing an inverse volatility allocation seeks to replicate the post-hoc portfolio as closely as possible. In the remaining paper, we treat the task of replicating the post-hoc portfolio as a forecasting problem. We forecast the realized variances of the N stocks in month $m+1$ based on information up to the end of the month m and form a portfolio based on the ranking that is

[4] Bloomberg ticker: SP5LVTUT

Figure 1: Cumulative return of volatility-weighted portfolios against market



Notes: Cumulative returns of two volatility-weighted portfolios along the CRSP value-weighted market return in between 2005:M1 to 2021:M12. The construction of the low-volatility portfolios mimics the methodology in the S&P 500 Low Volatility Top 80% Index (Bloomberg ticker SP5LVTUT). For more info about the data underlying the low-volatility portfolios, see Section 5.

implied by the forecasted variances $\widehat{RV}_{i,m+1|m}$, $i = 1, \ldots, N$. Excluding the 20% assets for which the volatility forecasts are the highest, we assign weights inversely proportional to the volatility forecast to those included.

Ultimately, we will address the forecasting problem in three steps. We first estimate various volatility models for each stock and evaluate the forecast performance of each model. Figure 2 below highlights the persistence of the realized volatilities in the cross-section of assets that are included in our analysis. The forecasting step of the analysis allows us to answer the question of which state-of-the-art volatility models provide the best forecasts of *monthly* stock volatility in a large cross-section of returns. While a large amount of literature evaluates daily volatility forecasting, the one-month horizon that is the main focus in our setting is less explored.

Secondly, we evaluate whether the forecasts from the volatility models translate into more accurate inverse volatility weights than the forecasts from the benchmark models. This is measured by the absolute deviations from the post-hoc allocation. This evaluation spans two dimensions: identifying high-volatility assets such that the right assets are excluded and performing accurate volatility forecasts such that the assigned portfolio weights are close to the post-hoc allocation.

Table 2: Performance statistics for low-volatility portfolios

	Ret	Std	SR
Post-hoc	12.81	13.01	0.98
4y-RV^d	10.74	14.13	0.76
12m-RV^d	10.56	13.95	0.76
6m-RV^d	10.75	13.86	0.78
1m-RV^d	10.81	13.73	0.79
Market	8.88	15.88	0.56

Notes: We report arithmetic means of discrete excess returns (Ret), their standard deviation, and the corresponding Sharpe ratio (SR). All measures are annualized. The evaluation period is 2005:M1–2021:M12. The market return is given by the CRSP value-weighted portfolio return.

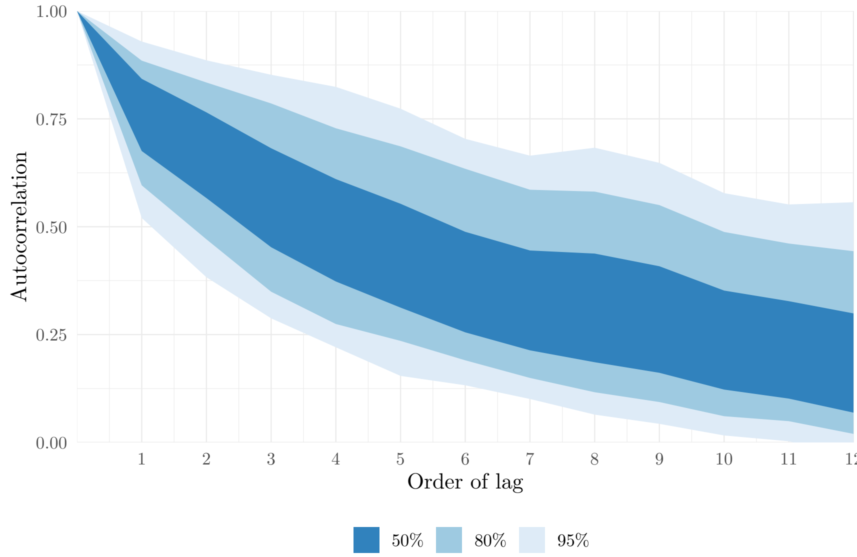
Lastly, we evaluate the out-of-sample performance of the portfolio allocations. We assess how attractive the portfolios are to particular investors both in terms of gross performance and net of transaction costs. To make this evaluation as realistic as possible, we estimate the effective costs of trading each individual asset.

4 Models

We examine an extensive selection of models that are commonly used for modeling volatility. These models can be broadly classified into four categories: RiskMetrics, GARCH, HAR, and MIDAS. The RiskMetrics and GARCH models approach volatility as a latent variable, while the HAR and MIDAS models model the realized variances directly. Next, we provide a brief introduction to each of the model specifications, with a more comprehensive description available in Appendix A.

We use four variants of the RiskMetrics model, which is an exponentially-weighted moving average (EWMA). Two variants employ monthly realized variances based on squared daily returns while the other two employ weighted averages of squared daily returns. The RiskMetrics models also differ in the amount of historical return data they consider, with some models using either six or twelve months of data. It is worth noting that the RiskMetrics models can be viewed as

Figure 2: Autocorrelation of monthly realized variances



Notes: The fan chart depicts the empirical autocorrelation functions (ACF) of monthly realized variances across 582 stocks that (1) were among the 500 largest at some point in between 2000:M1 to 2021:M12 and (2) were continuously traded during this period. The ACF is estimated using the instrumental variables regression proposed by Hansen and Lunde (2014) with 1 to 6 monthly lagged monthly RV as instruments.

restricted GARCH models with fixed ARCH/GARCH parameters and a constant equal to zero.

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). The MEM models differ from other GARCH models because they are estimated using the square-root of realized variances instead of daily return data.

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. Moreover, it has been documented that pooled cross-sectional estimation improves the out-of-sample forecast performance; see, for example, Bollerslev et al. (2018) and Kleen and Tetereva (2022). We denote these models by Panel HAR(-LR).

The MIDAS class of volatility models 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. The only exceptions are the 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. Forecasts are computed for month $m = 1, \ldots, M$.

Ghysels et al. (2019) compare iterated versus direct multi-step-ahead forecasting for GARCH, HAR, and MIDAS models. Accordingly, we follow their recommendations and directly forecast the average 22-day realized variance for all HAR-type and MIDAS models. On the other hand, we compute iterative volatility forecasts for the GARCH and MEM models.

5 Data

Monthly portfolio returns are calculated from monthly total returns taken from the Center of Research in Security Prices (CRSP). For our empirical analysis, we use the stocks that are among the largest 500 measured by market capitalization at some point in our sample. We adjust for CRSP delisting returns to have a survivorship bias-free data set (Shumway, 1997; Bali et al., 2016). Similar to Bollerslev et al. (2019) and Bollerslev et al. (2022), we merge daily CRSP data with

Table 3: Methods included in the empirical evaluation

Model	Description
12m-RV^d	Historical variance based on 12 months of daily data
4y-RV^d	Historical variance based on 4 years of daily data
6m-RV^d	Historical variance based on 6 months of daily data
1m-RV^d	Historical variance based on 1 month of daily data
RM monthly, 12 months	EWMA (J.P. Morgan, 1996) based on 12 months of monthly data
RM monthly, 6 months	EWMA (J.P. Morgan, 1996) based on 6 months of monthly data
RM daily, 12 months	EWMA (J.P. Morgan, 1996) based on 12 months of daily data
RM daily, 6 months	EWMA (J.P. Morgan, 1996) based on 6 months of daily data
GJR-GARCH	Glosten et al. (1993)
Panel GJR-GARCH	Pakel et al. (2011)
Factor GARCH	Engle et al. (1990)
Factor GARCH-VIX	Factor GARCH based on VIX
Factor GARCH-ΔHous	Factor GARCH based on macro-finance forecast
Factor GARCH-TS	Factor GARCH based on macro-finance forecast
MEM	Engle and Gallo (2006)
Panel MEM	Equivalent extension as Panel GJR-GARCH
HAR	Corsi (2009)
HAR-LR	Adapted HAR to monthly horizon
HAR-SPX	Includes a variance forecast for the SPX
HAR-SPX-LR	Adapted HAR-SPX to monthly horizon
HAR-VIX	Bekaert and Hoerova (2014)
HAR-VIX-LR	Adapted HAR-VIX model to monthly horizon
Panel HAR	Bollerslev et al. (2018)
Panel HAR-LR	Adapted Panel HAR model to monthly horizon
MIDAS	Ghysels et al. (2004)
Panel MIDAS	Equivalent extension as Panel GJR-GARCH

Notes: A detailed description of the models can be found in Appendix A.

NYSE TAQ intraday data. Open and close prices per day are taken from the daily CRSP data files. All intraday transaction data is obtained from NYSE TAQ. This trade data is thoroughly cleaned and we include only trades from the exchange that is referenced in the daily CRSP data. The two data sets are merged via the Wharton Research Data Services linking tables.

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.[5] Daily values for the VIX are obtained from the Cboe website.[6]. Observations start in January 2000 and end in December 2021. For the intraday realized variance estimates, we include prices during market hours from 9:30 to 16:00 and calculate 5-minute

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

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

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 since these are common in the literature and because 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 $RV_{i,t}^d$.

Excess market returns $R_{mkt,t}$ and the corresponding risk-free rates $R_{rf,t}$ are obtained from Kenneth R. French's data library.[7] For further factor analyses, we use the Fama-French(-Carhart) four- and five-factor portfolio returns; that is, monthly returns of SMB (Small Minus Big), HML (High Minus Low), MOM (Momentum), RMW (Robust Minus Weak) and CMA (Conservative Minus Aggressive) portfolios (Fama and French, 1993; Carhart, 1997; Fama and French, 2015). These are also obtained from Kenneth R. French's data library website. Last, real-time housing starts data are downloaded from ALFRED[8] and term spread data from the New York Federal Reserve website.[9]

6 Forecast Evaluation and Model Selection

In this section, we introduce two loss functions for evaluating the statistical performance of the models. We then discuss combining and examining the forecasts from a cross-sectional perspective.

[7] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/
[8] https://alfred.stlouisfed.org/series?seid=HOUST
[9] https://www.newyorkfed.org/research/capital_markets/ycfaq.html#/

6.1 Loss functions

For evaluating the variance forecast $\widehat{RV}_{m|m-1}$, we employ robust loss functions $L(RV_m, \widehat{RV}_{m|m-1})$ following Patton (2011).[10] In theory, one would prefer to evaluate the forecasts with respect to the quadratic variation (QV) of stock returns. However, as QV is unobservable, we use the monthly realized variances RV_m as proxies for QV_m.

In response, we employ two popular loss functions that are robust to employing RV_m instead of QV_m: the squared error (SE) loss defined as

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

and the QLIKE loss,

$$L(RV_m, \widehat{RV}_{m|m-1}) = RV_m/\widehat{RV}_{m|m-1} - \log(RV_m/\widehat{RV}_{m|m-1}) - 1.$$

While the SE is a symmetric loss function, the QLIKE is asymmetric and penalizes underestimation more heavily than overestimation. Furthermore, the QLIKE is less affected by extreme observations.

6.2 Forecast evaluation in a large cross-section of stocks

For each stock i, we consider the out-of-sample volatility forecasts $\widehat{RV}^j_{i,m|m-1}$, $m = 1, \ldots, M$ of model j. For each loss function and model j, we define the cross-sectional average loss in month m as,

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

We denote the loss of the benchmark model by L^B_m. The losses L^j_m and L^B_m can be used in real-time to select models. We report five statistics from the cross-sectional losses. First, we report the time-series averages and time-series medians of L^j_m/L^B_m. To reveal for how many months we are

[10]For simplicity in the notation, we drop the index i in this subsection.

outperforming the benchmark model we compute,

$$LR^j = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}_{L_m^j / L_m^B < 1}. \tag{2}$$

Here $\mathbf{1}_{L_m^j / L_m^B < 1}$ equals one if $L_m^j / L_m^B < 1$ and zero else. Hence, LR^j reports the share of months during which model j outperforms the benchmark in the cross-section. In addition, we report for how many months a specific model j is ranked best in terms of L_m^j (denoted by $Rk^j = 1$) or among the top-4 models (denoted by $Rk^j \leq 4$).

To combine forecasts, we consider the approach promoted by Caldeira et al. (2017). For each model j we determine the cross-sectional forecast performance in month m as

$$\bar{L}_m^j = \frac{1}{m} \sum_{k=0}^{m-1} \delta^k L_{m-k}^j, \tag{3}$$

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. All loss ratios are taken into account for $0 < \delta < 1$, but the weights are declining from the most recent to the most distant observation in time. In the interest of brevity, we restrict the evaluation to the cases of $\delta = 0$ and $\delta = 1$.

A natural concern is that some models perform better in high-volatility environments while others perform best in low-volatility environments (Conrad and Kleen, 2020). In the following, we consider forecast combinations as a means to safeguard against such time-varying model performance. Bates and Granger (1969) promote forecast diversification by combining forecasts from different models. [11] The combined forecast for the volatility of stock i, $i = 1, \ldots, N$, for period m is then given by

$$\widehat{RV}_{i,m|m-1}^{cf} = \sum_{j=1}^{J} \lambda_{j,m-1} \widehat{RV}_{i,m|m-1}^{j}, \tag{4}$$

[11] For further discussions see, for example, Timmermann (2006).

where the weights are given by

$$\lambda_{j,m-1} = \frac{(\bar{L}^j_{m-1})^{-\eta}}{\sum_{j=1}^{J}(\bar{L}^j_{m-1})^{-\eta}}. \tag{5}$$

with $\eta \geq 0$. Setting $\eta = 0$ assigns equal weights to the forecasts, while $\eta = \infty$ assigns full weight on a single model for which the loss in Equation (3) is the lowest and all other models receive a weight of zero. Weights are inverse proportional to the loss of the respective model when $\eta = 1$. Note that $\eta = 1/2$ in combination with the SE means that the weights are chosen according to the root mean squared error.

6.3 Volatility forecasting

Table 4 presents the forecast evaluation for individual models in Panel A, and for combined forecasts in Panels B and C. The left-hand side shows the loss statistics for SE and the right-hand side shows the QLIKE loss. The first two columns of each loss function in Table 4 present the monthly mean and median loss ratios L^j_m/L^B_m. The third column of the respective losses reveals the fraction of months the loss is lower than the 12m-RV^d benchmark. In Panel A, we further present the fraction of months a model achieved the lowest loss and is among the four models with the lowest losses, respectively.

Panel A reveals that no single model is dominant with respect to volatility forecasting. The strongest performing models are found within the HAR class, but the Midas framework and the Multiplicative Error Models obtain the lowest SE losses in around 40 percent of the sample months. In Panel B and C of Table 4, we see that combining forecasts produce lower or equal mean and median forecast errors both with respect to SE and QLIKE. However, the lowest median forecast error is associated with forecast combinations that perform model selection rather than weighting forecasts from different models. Furthermore, we find that the most accurate forecasting is associated with combinations that reduce the combination weights to the losses observed in the previous month. Jointly, these results imply that the median-loss optimal forecasts come from using a single model, which can change rapidly with new financial conditions. The best-performing model combination changes if we consider mean losses. By this metric, the most accurate forecasts

Table 4: Forecast comparison of model performance

Model		SE Mean	SE Med	SE LR^j	SE $Rk^j=1$	SE $Rk^j \leq 4$	QLIKE Mean	QLIKE Med	QLIKE LR^j	QLIKE $Rk^j=1$	QLIKE $Rk^j \leq 4$
Panel A: Model-based forecasts											
12m-RV^d		—	—	—	0.01	0.06	—	—	—	0.00	0.03
4y-RV^d		3.78	1.50	0.32	0.01	0.05	1.82	1.52	0.30	0.02	0.04
6m-RV^d		1.24	1.00	0.49	0.00	0.06	0.98	0.98	0.56	0.00	0.09
1m-RV^d		2.61	1.43	0.31	0.01	0.01	1.38	1.23	0.34	0.00	0.04
RM monthly, 12 months		0.98	0.98	0.68	0.01	0.06	0.98	0.98	0.76	0.00	0.03
RM monthly, 6 months		1.23	0.99	0.50	0.00	0.06	0.97	0.97	0.56	0.00	0.08
RM daily, 12 months		1.48	1.00	0.50	0.00	0.05	0.95	0.91	0.60	0.01	0.05
RM daily, 6 months		1.53	1.02	0.49	0.00	0.05	0.97	0.91	0.59	0.00	0.06
GJR-GARCH		1.48	0.99	0.52	0.00	0.01	1.00	1.00	0.50	0.00	0.01
Panel GJR-GARCH		1.79	1.14	0.43	0.00	0.03	1.10	1.00	0.50	0.00	0.03
Factor GARCH		1.31	0.98	0.50	0.00	0.03	1.02	1.01	0.50	0.01	0.03
Factor GARCH-VIX		1.05	0.85	0.59	0.00	0.03	0.94	0.93	0.59	0.00	0.01
Factor GARCH-ΔHous		1.06	0.89	0.60	0.00	0.05	0.92	0.94	0.63	0.00	0.01
Factor GARCH-TS		1.05	0.88	0.59	0.00	0.05	0.93	0.94	0.62	0.00	0.01
MEM		0.95	0.78	0.60	0.10	0.23	7.26	5.42	0.19	0.02	0.10
Panel MEM		1.05	0.71	0.61	0.08	0.23	7.64	4.01	0.29	0.11	0.16
HAR		0.95	0.59	0.77	0.01	0.18	0.87	0.55	0.86	0.01	0.31
HAR-LR		2.52	0.60	0.75	0.05	0.21	1.72	0.59	0.82	0.08	0.30
HAR-SPX		0.89	0.53	0.78	0.01	0.25	1.11	0.59	0.82	0.02	0.20
HAR-SPX-LR		2.40	0.58	0.71	0.04	0.30	1.89	0.60	0.78	0.03	0.24
HAR-VIX		0.90	0.50	0.78	0.07	**0.39**	1.02	0.56	0.83	0.10	0.29
HAR-VIX-LR		2.39	0.58	0.71	**0.14**	0.34	2.11	0.62	0.75	0.05	0.20
Panel HAR		0.74	0.55	0.78	0.06	0.29	0.63	0.57	0.84	0.08	0.38
Panel HAR-LR		**0.63**	0.52	**0.84**	0.14	0.34	0.59	**0.54**	0.90	**0.19**	**0.45**
MIDAS		0.76	**0.49**	0.82	0.09	0.32	**0.58**	0.54	**0.92**	0.08	0.45
Panel MIDAS		0.88	0.58	0.75	0.12	0.30	0.74	0.55	0.85	0.15	0.37
Panel B: Loss-based combined forecasts $\delta = 0$											
$\eta = 0$		0.68	0.57	0.84	—	—	0.68	0.66	0.90	—	—
$\eta = 1/2$	SE	0.59	0.47	0.92	—	—	0.62	0.60	**0.93**	—	—
	QLIKE	0.67	0.56	0.85	—	—	0.65	0.65	0.90	—	—
$\eta = 1$	SE	**0.55**	0.41	**0.93**	—	—	**0.58**	0.56	0.95	—	—
	QLIKE	0.64	0.53	0.89	—	—	0.63	0.62	0.91	—	—
$\eta = \infty$	SE	0.70	**0.39**	0.86	—	—	0.82	0.54	0.81	—	—
	QLIKE	0.61	0.39	0.88	—	—	0.59	**0.52**	0.87	—	—
Panel C: Loss-based combined forecasts $\delta = 1$											
$\eta = 1/2$	SE	0.64	0.53	0.88	—	—	0.66	0.64	0.92	—	—
	QLIKE	0.69	0.58	0.85	—	—	0.68	0.66	0.89	—	—
$\eta = 1$	SE	0.62	0.51	0.89	—	—	0.65	0.62	0.92	—	—
	QLIKE	0.68	0.58	0.86	—	—	0.67	0.65	0.90	—	—
$\eta = \infty$	SE	0.71	0.48	0.86	—	—	0.63	0.56	0.86	—	—
	QLIKE	0.63	0.52	0.87	—	—	0.59	0.53	0.90	—	—

Notes: For each loss function SE and QLIKE, the first two columns report the time-series mean and median of the loss ratio L^j_m/L^B_m. The column LR^j reports the proportion of months in which the cross-sectional loss L^j of model j is lower than the one of the 12m-RV^d benchmark forecast. $Rk^j \leq 1$ and $Rk^j \leq 4$ report the proportion of the respective model being the best or among the four best-performing models as measured by L^j_m. In Panel B and C, we report results for combined forecasts with $\delta = 0$ and $\delta = 1$, respectively. The combined forecast with equal weights ($\eta = 0$) is listed only for $\delta = 0$ because equal weights are independent of the smoothing parameter δ. In Panel A, numbers in bold highlight the lowest average and median loss ratio across models. Similarly, we highlight the highest LR^j, $Rk^j = 1$, and $Rk^j \leq 4$. We jointly apply the same highlighting to Panel B and C that are based on forecast combinations. The evaluation period is 2005:M1–2021:M12.

come from utilizing a model combination of all models where weights are inversely proportional to the losses they obtained in the previous month. We do not find that equally weighting models deliver competitive forecasting performance.

Within the set of model-based forecasts, it is evident that the short-term realized volatility models (1m-RV^d and 6m-RV^d) perform worse than the 12-month benchmark regarding the mean SE loss. Likewise, they fail to produce forecasts that outperform the benchmark to any greater extent with respect to the mean and median QLIKE losses. In the same way, we only find minor forecast improvements over the benchmark from the RiskMetrics methods (RM).

In contrast, the models based on intraday information such as MEM, HAR, and MIDAS models, generally perform better. Although the mean SE loss ratio for the Panel MEM is 105%, the median SE loss ratio is only 71%. Only the Panel MEM and individual HAR models with long-run components have average loss ratios that are worse than the benchmark. Individual HAR-LR models have an average loss ratio above 250% relative to the benchmark. However, the median loss associated with these models is lower than the benchmark, highlighting a very meaningful skewness in the losses over time. Among the HAR models, the Panel HAR-LR model stands out, with the lowest average SE loss ratio of 0.63, which is 11 percentage points lower than the Panel HAR model without quarterly and bi-annual RV. These results suggest that it is informative to incorporate lagged RVs of more than one month for monthly forecast horizons, despite the increased estimation uncertainty. Nonetheless, the increased estimation uncertainty needs to be addressed; for example, by the panel estimation approach.

Concerning the GARCH specifications, we find that the GARCH models that incorporate macroeconomic information (i.e., VIX, housing starts, or term spread) outperform the benchmark with respect to the median loss. However, the GARCH-type models only produce higher SE forecast accuracy than the benchmark between 43% and 61% of the months, significantly below the loss rate LR^j of the high-accuracy HAR- and MIDAS-type models, which ranges from 71% to 84%.

With respect to the QLIKE loss, we find that our forecast results align closely with those of the

SE results for the HAR- and MIDAS-type models, but with some notable exceptions. Foremost, the asymmetry of the QLIKE loss leads to remarkably high mean (and median) loss ratios that exceed 700% for the MEM models. This is because the MEM models frequently imply forecasts that are significantly lower than the realized volatility, and the QLIKE loss assigns greater weight to underprediction than to overprediction. With the exception of the MEM, HAR-SPX and HAR-VIX models, we find that the average loss ratio to the benchmark is lower under the QLIKE measure. Thus, the benchmark model appears more likely to underpredict volatility than the model-based alternatives.

Panel B and C of Table 4 present the forecast performance of the combined forecasts. The ex-post best forecast combination, in terms of the median SE-loss ratio, is 10 percentage points lower than the corresponding figure for the ex-post best model (the univariate MIDAS model). Furthermore, the forecast combinations that assign weights based on time-series averages of losses do not perform as well as the combinations using solely the last months losses. However, this is somewhat attributable to the limited time series available at the initial stages of the out-of-sample period. In Appendix B, we perform a forecast evaluation on a sub-sample that ranges from 2015 to 2021. The results are can be found in Table B.1. In this instance we find that the Panel-HAR models and the MIDAS approach produce lower mean forecast errors than the model combinations based on the most recent losses. In Table B.1, we further observe that selecting the best model based on long-term average performance (that is, $\delta = 1$ and $\eta = \infty$) yields the best performance among all model combinations and is now as efficient as the ex-post best model.

7 Comparison of Low-Volatility Portfolios

7.1 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 80% stocks with the lowest volatility are included in the portfolio for month m.

Stocks receive weights inverse proportional to their forecasted volatility, which is in line with the construction of the S&P 500 Low Volatility Top 80% Index, and closely aligns the portfolio allocation with volatility forecasts. We denote the individual weight of stock i w.r.t. forecasts from model j by

$$w_{i,m}^{j} \propto \frac{1}{\sqrt{\widehat{RV}_{i,m|m-1}^{j}}}.$$

The weights are standardized such that the sum of all weights is equal to 1. All remaining stocks that are in the highest volatility quintile receive a weight of zero. The inverse volatility allocation does not necessarily minimize the aggregate portfolio variance. However, assuming zero correlation between assets provides an allocation that is more stable and thus attracts less allocation costs (Kirby and Ostdiek, 2012). Furthermore, excluding covariances from the estimation of the portfolio weights reduces the variance of the estimates. This comes at the cost of a possible bias, but since the cross-section of assets is large in our application we have reason to believe that the net effect with respect to portfolio performance will be positive.

7.2 Portfolio evaluation

We measure the performance of the portfolios by their out-of-sample returns, volatilities, Sharpe ratios, turnover, costs and utility for a hypothetical investor. The Sharpe ratios are the average out-of-sample excess return divided by the out-of-sample volatility. In the tables we report annualized values by simple 12 and $\sqrt{12}$ scaling.

7.2.1 Portfolio turnover and transaction costs

We calculate the turnover of the portfolio allocations by the average fraction of wealth reallocated at re-balancing. Denoting the portfolio turnover TO_m, we follow 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 $m-1$. Based on the volatility forecasts for month $m+1$, the new desired weights are $w^j_{i,m+1}$. Before the next re-balancing 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 is the vector of equity returns; that is, $R_m = (R_{1,m}, \ldots, R_{n,m})'$. Hence, the turnover due to portfolio re-balancing at the end of month m is given by

$$TO^j_m = \sum_{i=1}^{N} \left| w^j_{i,m+1} - w^j_{i,m} \frac{1+R_{i,m}/100}{1+(w^j_m)'R_m/100} \right|. \tag{6}$$

The quantity TO_m can be interpreted as the proportion of wealth reallocated at the end of month m.

To assess the impact of transaction costs on portfolio performance, we follow the approach of Novy-Marx and Velikov (2016). Effective costs are estimated through a generalized trading-cost model proposed by Hasbrouck (2009) where daily log returns $r_{i,t}$ are modeled as

$$r_{i,t} = c_i \Delta q_{i,t} + \beta_i r_{mkt,t} + u_{i,t}, \tag{7}$$

with $r_{mkt,t}$ being the market factor in log returns, $q_{i,t}$ being an indicator for the trade direction in asset i and c_i denoting the cost. The error terms $u_{i,t}$ are assumed to be i.i.d. normal with variance σ^2_u. The cost is treated as unobserved and estimated using a Gibbs sampler.[12] Estimates are annual using daily data. We treat missing values as proposed by Novy-Marx and Velikov (2016). We match assets with respect to the Euclidean distance of rank-transformed idiosyncratic volatility and market capitalization. We calculate the ranks based on in-month average idiosyncratic volatilities. The latter are obtained from 90-day rolling window regressions.

With the portfolio turnover and the effective trading cost estimates we compute the portfolio returns in excess of the one-month Treasury bill rate,

$$R^j_{p,m} = \frac{W^j_m}{W^j_{m-1}} - 1 - R_{rf,m}. \tag{8}$$

[12] We are grateful to Hasbrouck for making SAS code available (https://pages.stern.nyu.edu/ jhasbrou/).

Here, W_m^j is the wealth of the model/loss-based portfolio, which can be obtained as

$$W_m^j = W_{m-1}^j \cdot (1 + (w_m^j)' R_m) \cdot (1 - TC_m^j), \tag{9}$$

with TC_m^j being the turnover-implied transaction costs per dollar traded at the end of month m. In the tables we report both gross and net performance of the allocations.

7.3 Utility

To highlight the economic significance of the various portfolios we make use of a utility framework. We follow Fleming et al. (2001, 2003) and compute the fee required to make a hypothetical investor indifferent with respect to the portfolios and the post-hoc allocation. Using a quadratic utility function with risk-aversion parameter γ, the monthly utility generated by a portfolio based on model j is given by

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

This utility is compared to the utility obtained from the post-hoc portfolio under a fee. Denoting the return of the post-hoc portfolio by $R_{p,m}^o$, we compute the maximum fee Δ_γ^j that an investor would be willing to pay in order to switch from portfolio j to the post-hoc portfolio by solving

$$\sum_{m=1}^{M} U_\gamma(R_{p,m}^j) = \sum_{m=1}^{M} U_\gamma(R_{p,m}^o - \Delta_\gamma^j). \tag{10}$$

Portfolio with a comparably small Δ_γ^j are more closely mimicking the utility of the post-hoc portfolio. We report the fee Δ_γ^j in Table 5 in annualized percentage points for $\gamma = 4$.

8 Out-of-sample portfolio performance

8.1 Portfolio performance

Table 5 shows the annualized returns of each portfolio with and without trading costs. The 12m-RV^d serves as our benchmark portfolio. We report the annualized return and test whether there is a significant difference between the return of the respective model-/loss-based portfolio and the benchmark. Inference is based on Newey-West standard errors. We also test for differences in Sharpe ratios and volatilities between the respective portfolios and the benchmark using tests proposed by Ledoit and Wolf (2008, 2011).

The financial performance relative to the benchmark depends greatly on whether transaction costs are included or not. With respect to the annualized Sharpe ratios obtained under the model-based portfolios, we find that the statistical difference to the RV benchmark is greatly reduced once transaction costs are accounted for. Without transaction costs, all HAR specifications, MEM and MIDAS models obtain Sharpe ratios that are significantly higher than the Sharpe ratios of the benchmark on all conventional significance levels. However, the number of model- and loss-based portfolios that remain significantly different at the 5% level drops from 24 to 8 portfolios when accounting for transaction costs.

The highest Sharpe ratios are associated with the models that deliver the lowest forecast errors in Section 6. However, the differences in Sharpe ratios between the loss-based and the HAR- and MIDAS-based portfolios are only minor. In contrast to the forecasting performance, we find that the portfolios formed from model combinations using $\delta = 1$ and $\eta = \infty$ deliver lower out-of-sample volatility and higher net expected returns than the other loss-based portfolios. Together with the Panel-HAR-LR and HAR-VIX approaches, a forecast combination with $\delta = 1$ and $\eta = \infty$ provides the most financially effective portfolio in our evaluation. It is noteworthy that the equally-weighted forecast combination neither delivers low forecasting errors (see Table 4) nor high financial performance in terms of mean returns, volatilities, or Sharpe ratios.

Turning to the net-return volatility we find that among the model-based portfolios, it is only the

Table 5: Returns of low-volatility portfolios

		Without TC				With TC			
		Ret	Std	SR	Δ_4	Ret	Std	SR	Δ_4
Post-hoc		12.81	13.01	0.98	—	12.24	13.02	0.94	—
Panel A: Model-based portfolios									
12m-RV^d		10.56	13.95	0.76	2.78	10.33	13.97	0.74	2.44
4y-RV^d		10.74	14.13**	0.76	2.70	10.54	14.14**	0.75	2.33
6m-RV^d		10.75	13.86*	0.78	2.53	10.45	13.88*	0.75	2.27
1m-RV^d		10.81	13.73*	0.79	2.40	9.82*	13.74*	0.72	2.81
RM monthly, 12 months		10.58	13.95	0.76	2.75	10.34	13.96	0.74	2.42
RM monthly, 6 months		10.75	13.86	0.78	2.54	10.44	13.88	0.75	2.28
RM daily, 12 months		10.73	13.87	0.77	2.56	10.26	13.88	0.74	2.46
RM daily, 6 months		10.72	13.83	0.77	2.55	10.24	13.85	0.74	2.46
GJR-GARCH		10.79	13.84	0.78	2.48	10.30	13.86	0.74	2.41
Panel GJR-GARCH		10.71	13.86	0.77	2.58	10.27	13.87	0.74	2.44
Factor GARCH		10.85	14.02	0.77	2.53	10.48	14.03	0.75	2.33
Factor GARCH-VIX		10.86	14.10*	0.77	2.57	10.47	14.11*	0.74	2.38
Factor GARCH-ΔHous		10.90*	14.13**	0.77	2.54	10.51	14.14**	0.74	2.36
Factor GARCH-TS		10.87	14.15**	0.77	2.58	10.48	14.16**	0.74	2.40
MEM		11.23**	13.89	0.81**	2.08	10.75	13.91	0.77	1.98
Panel MEM		11.13**	13.82	0.81**	2.14	10.68	13.83	0.77*	2.01
HAR		11.01*	13.75*	0.80**	2.21	10.59	13.76	0.77*	2.06
HAR-LR		11.20***	13.72*	0.82***	2.00	10.73*	13.73*	0.78**	1.91
HAR-SPX		11.07**	13.78	0.80**	2.17	10.65	13.79	0.77	2.02
HAR-SPX-LR		11.13**	13.68*	0.81**	2.05	10.64	13.70*	0.78*	1.98
HAR-VIX		11.15***	13.77*	0.81***	2.08	10.75**	13.78	0.78**	1.91
HAR-VIX-LR		**11.23*** **	**13.68***	**0.82*** **	**1.95**	10.75*	**13.69***	0.79**	1.86
Panel HAR		11.14**	13.80*	0.81**	2.11	10.77*	13.82*	0.78**	1.91
Panel HAR-LR		11.20**	13.76**	0.81***	2.03	**10.86*** **	13.77**	**0.79*** **	**1.80**
MIDAS		11.19**	13.80	0.81***	2.06	10.80**	13.82	0.78**	1.88
Panel MIDAS		11.06*	13.83	0.80*	2.21	10.69	13.85	0.77	2.01
Panel B: Loss-based portfolios $\delta = 0$									
$\eta = 0$		10.92**	13.84*	0.79**	2.35	10.59	13.85*	0.76*	2.11
$\eta = 1/2$	SE	10.90*	13.85	0.79**	2.38	10.57	13.86	0.76*	2.14
	QLIKE	10.90*	13.87	0.79**	2.39	10.57	13.88	0.76*	2.15
$\eta = 1$	SE	10.94**	13.85	0.79**	2.34	10.60	13.86*	0.76*	2.11
	QLIKE	10.93**	13.86	0.79**	2.35	10.60	13.87	0.76*	2.12
$\eta = \infty$	SE	11.05**	13.87	0.80**	2.24	10.56	13.89	0.76	2.17
	QLIKE	10.86	13.88	0.78*	2.43	10.41	13.90	0.75	2.32
Panel C: Loss-based portfolios $\delta = 1$									
$\eta = 0$		10.92**	13.84*	0.79**	2.35	10.59	13.85*	0.76*	2.11
$\eta = 1/2$	SE	10.89*	13.85*	0.79**	2.38	10.56	13.86*	0.76*	2.14
	QLIKE	10.90*	13.84*	0.79**	2.38	10.57	13.85*	0.76*	2.13
$\eta = 1$	SE	10.90*	13.86*	0.79**	2.38	10.57	13.87*	0.76*	2.14
	QLIKE	10.92*	13.84*	0.79**	2.35	10.60	13.85*	0.77*	2.10
$\eta = \infty$	SE	**11.24*** **	**13.75*** **	**0.82*** **	**1.98**	**10.88*** **	13.77**	**0.79*** **	**1.77**
	QLIKE	11.19**	13.76**	0.81***	2.04	10.84**	**13.77*** **	0.79**	1.82

Notes: Average annualized excess 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 post-hoc portfolio; see Equation (10). 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) and the volatility test according to Ledoit and Wolf (2011). Statistical significance at the 10%, 5%, and 1% level are indicated by *, **, and *** respectively. In Panel A, numbers in bold represent the highest return, the lowest standard deviation, the highest SR and the lowest fee Δ_4 across all models. We jointly apply the same highlighting to all forecast combinations in Panel B and C. The evaluation period is 2005:M1–2021:M12.

HAR models along with the 1m-RV^d and 6m-RV^d that deliver statistically significant reductions in portfolio volatility compared to the benchmark allocation. In this set of models, it is only the Panel HAR-LR that reduces volatility to an extent that is significant at the 1% level. That reduction is of 0.2 percentage points, and, together with an increase in mean excess returns of slightly above 0.5 percentage points, we find a Sharpe ratio increase of about 0.05, annualized.

Among the portfolios formed from forecast combinations, we found that the combinations that use short-term losses to perform model selection produced the lowest level of forecast error. However, the inverse volatility allocations from these forecasts exhibit greater costs than the other loss-based portfolios. The difference between the gross and net returns falls from 11.05 to 10.56, whereas the combination relying on selection using long-term average losses only decreases from 11.24 to 10.88. Performing model selection using the short-term losses fails to improve upon the benchmark Sharpe ratio in a manner that is statistically significant at any significance level. In Appendix C, we report additional results for value-weighted portfolios with stock exclusion and inverse-volatility-weighted portfolios without stock exclusion. It becomes evident that our combination of stock exclusion and volatility-weighting performs best in terms of returns and Sharpe ratios.

8.2 Economic significance

To contextualize the improvements to financial performance we assess the fee required to make an investor indifferent between the post-hoc allocation and the feasible allocations. In most cases, we see that the required fee Δ_γ decreases when costs are introduced. This implies that while the post-hoc portfolio obtains the highest net performance, it is comparably expensive to maintain. Independent of whether costs are accounted for or not, the greatest utility is observed for the loss-based approach when setting $\eta = \infty$, $\delta = 1$, and using the SE-loss. Allowing for costs, an investor holding the post-hoc portfolio would prefer this allocation if the annualized fee on the infeasible allocation exceeds 1.77%. This fee is around 37% lower than the highest fee that we obtain, which is for the 1m-RV^d equal to 2.81 %. A striking finding is that the fee associated with this allocation increases once costs are introduced, contrary to most other methods. Thus, this method allocates more wealth than the post-hoc to assets that are associated with higher trading costs.

Since the analysis that incorporates costs is the most realistic, we will focus our discussion on these results in the following. We find that using the ex-ante standard deviation of daily returns as volatility measures and the RiskMetrics variants often produce worse results in terms of net utility than model-based approaches. The strongest performance of the model-based portfolios within this utility framework is found within the HAR category. The fees obtained from these estimators are consistently low relative to the other methods evaluated. Within the set of HAR models, the lowest fee is associated with the approach that incorporates long-run components and the VIX index (HAR-VIX-LR).

Most of the loss-based portfolios do not obtain a level of utility as high as those of the best-performing HAR models. The exception is when an investor averages forecast performance over all previous months ($\delta = 1$) and puts all weight on the forecast with the lowest loss ($\eta = \infty$). This is again in stark contrast to the common finding that equally-weighted forecasts are hard to outperform. A possible explanation is that averaging the loss over the previous months reduces the variance of the loss estimate.[13] As such, an investor can rely on the time-series model with the lowest loss. If we set $\delta = 0$, such that the loss simply becomes that of the previous month, we find relatively poor performance when $\eta = \infty$. In this case, the equally-weighted forecasts provide a better alternative.

8.3 Comparison to the post-hoc portfolio and trading costs

To compare the respective portfolios with the infeasible post-hoc allocation we compute absolute deviations in the portfolio weights (WAD). We summarize the cross-section of WAD by the mean and median deviations along with the minimum, maximum, $20th$ and $80th$ percentiles. Table 6 presents the time-series averages of these statistics. To complement the summary of the cross-sectional deviations from the post-hoc portfolio, we compute the variation in weights over time for

[13] This reduction comes at the cost of a potential bias in the case where the loss is time-varying.

the respective portfolios. This is the turnover, which we contextualize economically through the transaction costs.

The minimum deviations across the model-based and loss-based portfolios are naturally zero. These are the cases where both the post-hoc and the forecasted allocations agree on excluding certain assets. Likewise, we see that the maximum is magnitudes larger than the average $80th$ percentile. These are cases where the allocations disagree on including assets or not. The average maximum deviations are greater in the cases of disagreement and where one of the allocations assigns a high weight (i.e., low volatility) to the asset. These extreme disagreements are around of 4.8 percent and foremost within the set of model-based portfolios and the loss-based portfolios that maintain $\delta = 1$ and $\eta = \infty$. These loss-based portfolios select a single model based on the average loss. Two outliers within the set of large deviations are the 1m-RV^d and the Panel HAR allocations. These exhibit average maximum deviations of 4.19 and 4.24 percent respectively, which is far below the level we find across the other methods.

Disagreements about including certain assets create a strong positive skewness in the distribution of WAD. For example, the median deviation under the Panel MEM approach is more than half of the mean deviation. It is very striking that while the 1m-RV^d exhibit low maximum deviations from the post-hoc portfolio, the mean $80th$ percentile exceeds that of the other allocations. It is around 0.14 percent, which is more than double that of many loss-based combinations and HAR-based portfolios.

Focusing on the median deviations, we find that the loss-based, HAR-, and MIDAS-based portfolio weights exhibit lower deviations. These as the same models that exhibit high forecasting performance in Table 4. Loss-based portfolios that maintain $\delta = 1$ and $\eta = \infty$ again provide an interesting case where the average median deviation is low, 0.02 percent. Likewise, we see that the difference of the average $20th$ and average $80th$ percentiles is comparably low at around 0.05 percent. The implications is that using solely the most accurate forecast based on long-run average losses will give you an allocation comparably close to the post-hoc allocation, but with a small number of large misallocations.

Turning to the turnover of the respective portfolios we find that the portfolios formed on the realized volatilities using daily returns over the previous month (1m-RV^d) require an, in this context, extreme degree of trading. On average 42.83 percent of wealth is traded to maintain this portfolio. The cost associated with this turnover is 8.18 basis points per month. Strikingly, the second-highest turnover is associated with the post-hoc portfolio. Economically this implies that the post-hoc foresight of the inverse volatility allocation produces a portfolio that is impractical and expensive to maintain; that is, monthly trading costs of 4.74 basis points on average.

Naturally, the allocations that rely on averages of squared daily returns over a 12 or 48-month period are by contrast very stable, with turnover as low as 8.32 percent. We also find low turnover within the set of loss-based allocations, with the exception of model selection using solely short-term losses. The costs of the other loss-based portfolios range from 2.73 to 2.86 basis points. Within the set of model-based allocations, the Panel HAR-LR, long-run realized volatilities, and the RiskMetrics utilizing monthly data obtain the same level of costs.

8.4 Portfolio characteristics over time

The portfolio allocations that we form by model selection deliver better out-of-sample performance than those formed using 12 months of daily returns. In this section we further dissect the differences between the competing procedures by analyzing the financial characteristics of the portfolios. Using accounting data from CRSP and Computstat, we compute the market capitalization, book-to-market ratio, operating profitability and investment characteristics for each firm in the sample. For this, we follow the same procedure and conventions used by Green et al. (2017). Aggregating the firm characteristics in the portfolios—weighted by the relative wealth invested in each stock—lets us compute the characteristics of the portfolios. Fama and French (2015) find monotonic relations between these characteristics and expected returns among stocks with high market capitalization, which provides the basis for our comparison since the volatility sorts are restricted to the largest firms on the market.

Figure 3 presents the magnitude of the characteristics associated with the portfolio formed

Table 6: Portfolio characteristics

		WAD						TO	TC
		Min	$q_{20\%}$	Mean	Median	$q_{80\%}$	Max	—	—
Post-hoc		—	—	—	—	—	—	24.33	4.74
Panel A: Model-based portfolios									
12m-RV^d		0.00	0.01	0.06	0.03	0.10	4.82	9.96	1.99
4y-RV^d		0.00	0.01	0.07	0.04	0.12	4.77	8.32	1.69
6m-RV^d		0.00	0.01	0.06	0.03	0.10	4.83	13.04	2.56
1m-RV^d		0.00	0.01	0.07	0.04	0.14	4.19	42.83	8.18
RM monthly, 12 months		0.00	0.01	0.06	0.03	0.10	4.82	9.99	1.99
RM monthly, 6 months		0.00	0.01	0.06	0.03	0.10	4.83	13.07	2.57
RM daily, 12 months		0.00	0.01	0.06	0.03	0.11	4.78	20.03	3.89
RM daily, 6 months		0.00	0.01	0.06	0.03	0.11	4.78	20.54	3.96
GJR-GARCH		0.00	0.01	0.07	0.04	0.11	4.73	20.03	4.12
Panel GJR-GARCH		0.00	0.01	0.06	0.04	0.11	4.67	17.46	3.60
Factor GARCH		0.00	0.01	0.07	0.04	0.11	4.81	14.72	3.10
Factor GARCH-VIX		0.00	0.01	0.07	0.04	0.11	4.81	15.38	3.23
Factor GARCH-ΔHous		0.00	0.01	0.07	0.04	0.11	4.82	15.49	3.24
Factor GARCH-TS		0.00	0.01	0.07	0.04	0.11	4.82	15.44	3.23
MEM		0.00	0.01	0.06	0.03	0.10	4.44	19.40	3.96
Panel MEM		0.00	0.00	0.05	0.02	0.07	4.64	17.91	3.73
HAR		0.00	0.00	0.04	0.02	0.06	4.41	16.70	3.49
HAR-LR		0.00	0.00	0.05	0.02	0.06	4.69	18.40	3.97
HAR-SPX		0.00	0.00	0.05	0.02	0.06	4.76	16.72	3.53
HAR-SPX-LR		0.00	0.00	0.05	0.03	0.07	4.78	19.13	4.13
HAR-VIX		0.00	0.00	0.05	0.02	0.06	4.78	15.76	3.31
HAR-VIX-LR		0.00	0.01	0.05	0.03	0.07	4.77	18.46	4.01
Panel HAR		0.00	0.00	0.04	0.02	0.05	4.24	15.11	3.08
Panel HAR-LR		0.00	0.00	0.04	0.02	0.05	4.84	14.14	2.83
MIDAS		0.00	0.00	0.04	0.02	0.06	4.43	15.57	3.23
Panel MIDAS		0.00	0.00	0.04	0.02	0.05	4.68	15.02	3.05
Panel B: Loss-based combined portfolios $\delta = 0$									
$\eta = 0$		0.00	0.00	0.05	0.03	0.07	4.76	13.46	2.73
$\eta = 1/2$	SE	0.00	0.00	0.05	0.02	0.06	4.76	13.69	2.79
	QLIKE	0.00	0.00	0.05	0.02	0.06	4.76	13.69	2.79
$\eta = 1$	SE	0.00	0.00	0.05	0.02	0.06	4.76	13.98	2.86
	QLIKE	0.00	0.00	0.05	0.02	0.06	4.76	13.98	2.86
$\eta = \infty$	SE	0.00	0.00	0.05	0.02	0.06	4.43	20.18	4.14
	QLIKE	0.00	0.00	0.05	0.02	0.06	4.43	20.18	4.14
Panel C: Loss-based combined portfolios $\delta = 1$									
$\eta = 1/2$	SE	0.00	0.00	0.05	0.03	0.07	4.76	13.49	2.73
	QLIKE	0.00	0.00	0.05	0.03	0.07	4.76	13.49	2.73
$\eta = 1$	SE	0.00	0.00	0.05	0.02	0.07	4.76	13.56	2.74
	QLIKE	0.00	0.00	0.05	0.02	0.07	4.76	13.56	2.74
$\eta = \infty$	SE	0.00	0.00	0.04	0.02	0.05	4.84	14.88	3.01
	QLIKE	0.00	0.00	0.04	0.02	0.05	4.84	14.88	3.01

Notes: Summary measures of the model-based and loss-based portfolios are reported. WAD refers to the absolute distance between the model-based portfolio weights and the weights of the infeasible post-hoc portfolio. We report the time-series average of the cross-sectional minimum WAD (Min), 20%-quantile of WAD ($q_{20\%}$), mean WAD (Mean), median WAD (Median), 80%-quantile of WAD ($q_{80\%}$), and the maximum WAD (Max). For the definition of turnover (TO) see Subsection 7.2.1. WAD and TO are reported in percentages. TC are reported in basis points. The evaluation period is 2005:M1–2021:M12.

using model selection relative to the allocation based on 12m-RV^d. While there is significant variation in the ratio over time, some striking patterns stand out. First, using model selection delivers portfolios that have lower market capitalization than the portfolios based on 12m-RV^d. There are periods in which the size of the model selection portfolio is higher, but the average size is about 95% of the market capitalization in the benchmark allocation. At times the ratio falls as low as 80%. Thus, the increased forecasting accuracy as reported in Table 4, reveals a greater number of large firms to be more volatile than 12m-RV^d. These firms are thus excluded or receive lower weight in the portfolio.

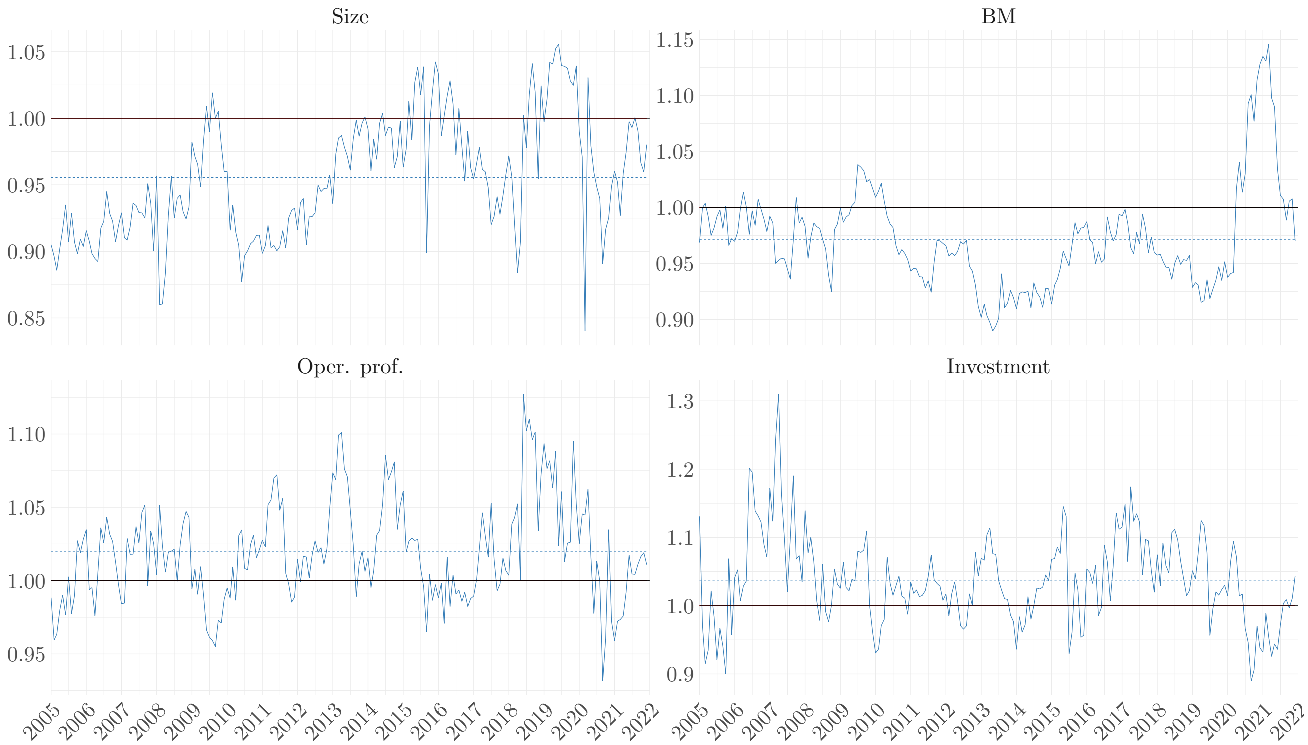
Fama and French (1995) highlight that typical high book-to-market firms are relatively distressed, and thus attract a premium. Over the greater part of our sample, we find that the book-to-market ratio of the model selection portfolio is lower than the benchmark. The striking exception is the period coinciding with the Covid-19 pandemic, during which the book-to-market ratio in the allocation based on model selection spikes. We briefly discuss this period further in Section 8.6.

Our Figure 3 also reveals that model selection delivers portfolios that are in general more robust with respect to their operating profitability, and have higher levels of investments. The investment characteristics is as much as 30% higher than the benchmark during the period that precedes the financial crisis. The greatest differences in operating profitability are in the latter part of the sample and at times exceeding 10%.

Complementary to the lower volatilities reported in Table 5, these results indicate the following: Increasing the forecasting accuracy through model selection typically attracts different premia than the benchmark allocation with regards to these 4 characteristics. Under model selection, the smaller size and higher operating profitability attract premia, relative to the benchmark. The allocation based on 12m-RV^d typically finds relative premia through higher book-to-market and lower investment characteristics.

A complete description of the differences in financial characteristics that emerge under different forecasting methods is beyond the scope of this paper. However, we highlight that there is variation in the financial characteristics of the portfolios formed from forecasting models that perform selection, compared to the benchmark allocation that rely on 12m-RV^d.

Figure 3: Ratio of weighted average firm characteristic of SE-based portfolio divided by weighted average firm characteristic of 12m-RV^d portfolio



Notes: Monthly ratio of weighted stock characteristics. Time-series averages depicted as dashed lines. We merge our portfolio data with firm characteristics as calculated in Green et al. (2017). Each month, we calculate the weighted sum of firm characteristics per 12m-RV^d portfolio and the best-performing SE-based portfolio (i.e., $\eta = \infty$, $\delta = 1$). We calculate the ratio with the average characteristic of 12m-RV^d in the denominator; that is, ratios above one indicate that the SE-based portfolio has a higher weighted characteristic as the 12m-RV^d benchmark in that month.

8.5 Factor analysis

In this section, we evaluate the low-volatility allocations in the presence of a small number of factors. By estimating time-series regressions of the low-volatility allocations onto the five Fama-French factors and Momentum, we find whether the unexplained expected returns increase when using our loss-based portfolios. The regressions are defined as,

$$R_{p,m} = \alpha_p + \beta_p' f_m + e_{p,m}, \tag{11}$$

where $R_{p,m}$ denotes again the monthly excess return on the low-volatility portfolios. Exposures to the factors f_m are denoted β_p and the unexplained expected return is α_p. The standard deviation of the error term $e_{p,m}$ is denoted by $\sigma_{p,m}$.

We restrict the evaluation to our loss-based portfolios using $\eta = \infty$ and $\delta = 1$, and the 12m-RV^d allocation as a benchmark. With respect to the factors, we make use of the CAPM, Fama-French-Carhart (FFC) four-factor model (Fama and French, 1993; Carhart, 1997) and the Fama-French five-factor model (Fama and French, 2015). All inference is based on Newey-West standard errors and may be regarded as conservative in this context. The findings are reported in Table 7.

First, we observe that the alphas associated with the loss-based portfolios are statistically significant at conventional levels of significance across all factor models. The weakest evidence against a zero alpha is with respect to the QLIKE-based portfolio under the Fama-French five factor model. The t-statistic associated with the intercept of 1.31 percent, annualized, is 1.86. Under the CAPM and the four factor model the low-volatility portfolio retain annualized alphas above 2% with t-statistics ranging from 2.66 to 3.53. The benchmark portfolio does not produce similar findings; only when using the four factor model, the alpha becomes statistically different from zero.

The appraisal ratio $AR_p = \alpha_p/\sigma_{p,e}$ tells us how an investor holding the factor strategies improves the slope of the attainable mean-variance frontier by adding a portfolio with a long position in the low-volatility portfolio and an equal short position in the risk-free rate. Contrasting the estimates for our loss-based allocations to the benchmark we see that the appraisal ratios are consistently higher across all factor models. The highest appraisal ratio is found in the SE-loss portfolio in the context of the four factor model. In the case of the five factor model we find that the AR increases from the benchmark 0.23 to 0.48 (SE) and 0.46 (QLIKE). This implies that the residual volatility of the factor model increases with the loss-based approach but not to the same degree as the increase in alpha, since the unexplained expected returns more than double compared to the benchmark.

With respect to the individual factors, the estimates of the five factor model reveal that only exposures to the market and profitability factor are significant for the loss-based and benchmark portfolios. In no model specification do we find strong conditional relations between the low-volatility portfolios and the size and value factors. The residual variance is greatly decreased by adding the Fama-French and momentum factor, but the exposure to the market factor remains stable at around 0.9.

Table 7: Low-volatility portfolio returns and factor loadings

	CAPM			FFC						FF5						
	α	β_{MKT}	AR	α	β_{MKT}	β_{SMB}	β_{HML}	β_{MOM}	AR	α	β_{MKT}	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}	AR
12m-RV^d	1.27	0.90	0.41	1.37	0.90	-0.02	0.05	0.02	0.45	0.66	0.91	0.02	0.02	0.14	0.06	0.23
	(1.54)	(39.05)		(2.04)	(41.09)	(-0.45)	(1.23)	(0.77)		(0.88)	(37.16)	(0.71)	(0.52)	(2.91)	(1.16)	
SE	2.08	0.89	0.69	2.11	0.89	-0.01	0.04	0.02	0.70	1.38	0.90	0.03	0.00	0.15	0.05	0.48
	(2.75)	(41.91)		(3.53)	(47.68)	(-0.29)	(0.84)	(1.09)		(1.97)	(39.36)	(1.07)	(0.07)	(2.94)	(0.91)	
QLIKE	2.02	0.89	0.67	2.04	0.89	-0.01	0.03	0.02	0.68	1.31	0.90	0.03	-0.00	0.15	0.05	0.46
	(2.66)	(42.12)		(3.40)	(48.03)	(-0.21)	(0.77)	(1.02)		(1.86)	(39.47)	(1.17)	(-0.03)	(2.90)	(0.97)	

Notes: Regressions of the low-volatility portfolio returns on different factor models in the leading case of $\eta = \infty$ and $\delta = 1$. 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 appraisal ratio (AR) is the annualized ratio of the estimated excess return and the standard deviation of the factor-model-specific residuals. The evaluation period is 2005:M1–2021:M12.

8.6 Sector concentration and Covid-19 period

To further assess the portfolio allocations features relative to post-hoc allocation, we examine whether our low-volatility investing strategies may generate high exposure to a small set of industries. In Figure 4, we depict histograms of the average sector concentration by primary SIC codes. We report numbers for the post-hoc portfolio, the 12m-RV^d benchmark, and the SE- and QLIKE-based portfolios for the leading case with $\eta = \infty, \delta = 1$. For brevity, the latter two are considered to be representative of our model-based strategies. We use real-time 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 divestitures like the sale of McGraw-Hill Education in 2013.

In Figure 4, we see that 40.5% of the weight in the post-hoc portfolio belongs to stocks classified as "Manufacturing."[14] The second largest industry is "Transportation, Communications, Electric, Gas, and Sanitary service" (17.1%), followed by "Finance, Insurance, and Real Estate" (14.6%), "Services" (13.8%), "Trade" (11.3%), and "Mining and Construction" (2.26%). 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 post-hoc portfolio.

The low weight for the sector "Mining and Construction" provides a practical distinction between our inverse volatility portfolios and minimum-variance allocations. Blitz et al. (2019) find that these assets are relatively high risk, but with low correlation with other stocks. Hence, a minimum-variance portfolio may include such high-risk-low-correlation stocks in order to minimize the overall portfolio variance. We find that our inverse volatility portfolios have very low exposure to this section, in line with the results of Blitz et al. (2019).

In Figure 5 we plot the monthly aggregate weight of sectors that are identified by S&P Global Market Intelligence to be most affected by the Covid-19 pandemic; for example, the car industry.[15] We report these aggregate weights for the 12m-RV^d portfolio and the SE-based portfolio (for $\eta = \infty$, $\delta = 1$). We see that the reaction of the SE-based strategy to the Covid-19 epidemic crash in March 2020 is way more timely. We observe a stable level shift in the weight attached to these industries for the SE-based portfolio in contrast to the 12m-RV^d portfolio. The latter seems to "overshoot" in its reaction to the stock market crash in March 2020.

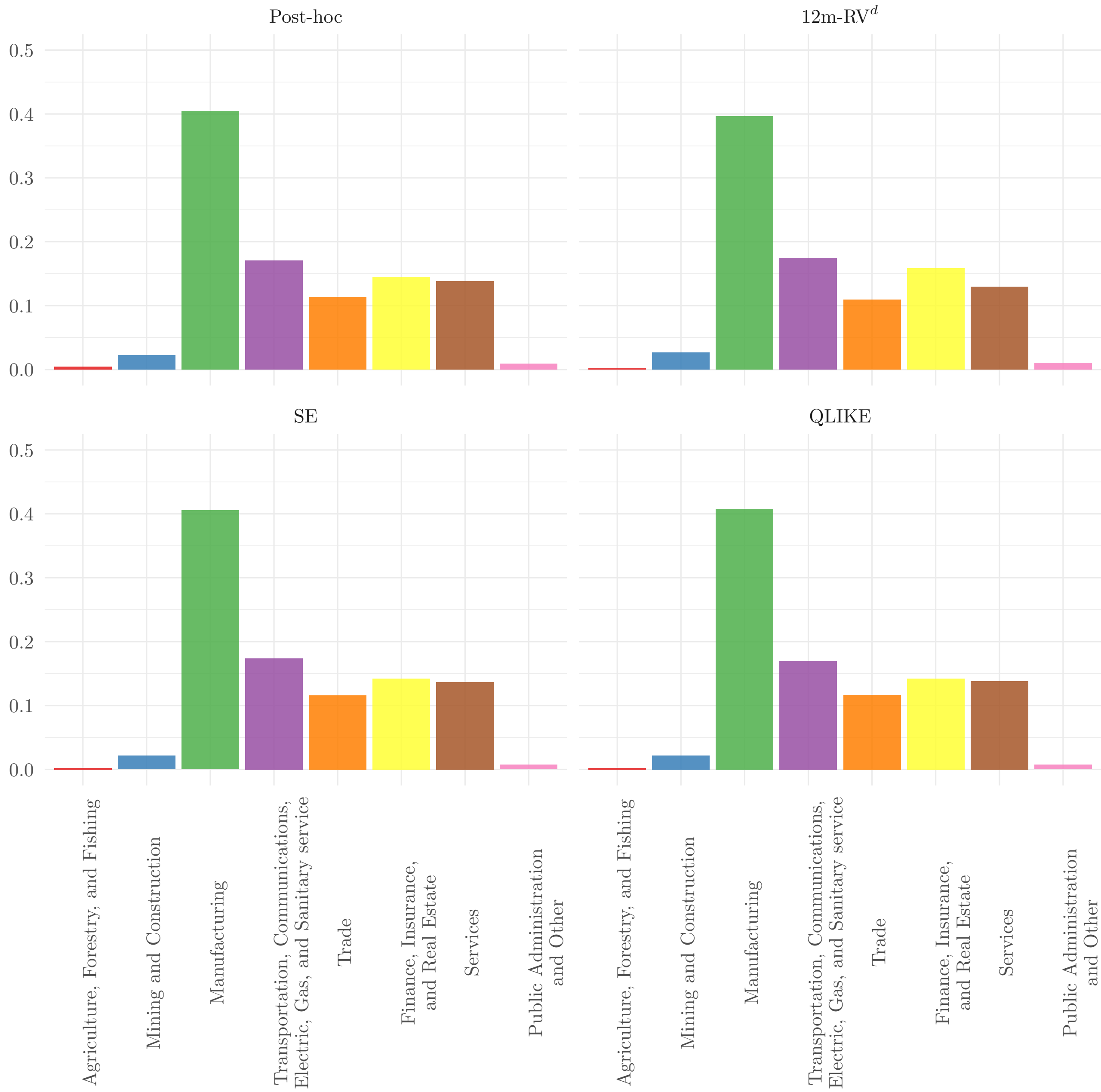
9 Conclusion

We investigate the impact of incorporating volatility models on low-volatility investments in the US stock market. The low-volatility allocation is typically exploited by sorting stocks based on historical volatility. In contrast, we employ a variety of time-series models based on intraday data,

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

[15]https://www.spglobal.com/marketintelligence/en/news-insights/blog/industries-most-and-least-impacted-by-covid-19-from-a-probability-of-default-perspective-january-2022-update
SIC codes: 13, 37, 45, 49, 50, 55, 58, 70, 79

Figure 4: Sector concentration of post-hoc and feasible low-volatility portfolios



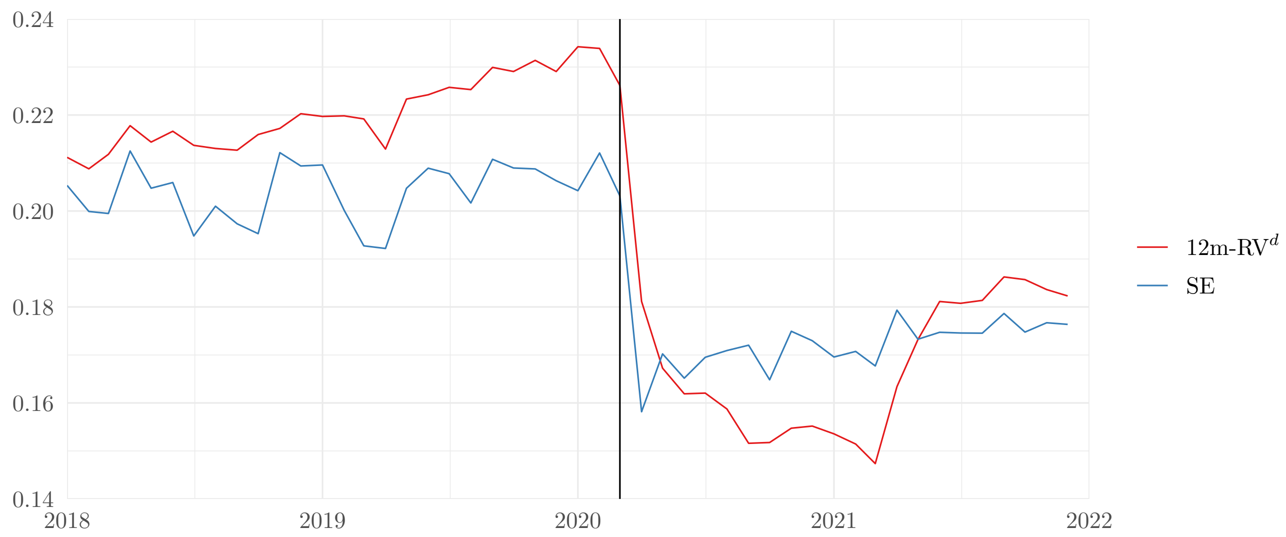
Notes: Sector concentration by real-time Standard Industrial Classification (SIC). We report the time-average share inside each SIC class for the infeasible post-hoc portfolio along the corresponding numbers for three exemplary ex-ante feasible portfolios: the benchmark 12m-RV^d portfolio and the SE-based and QLIKE-based portfolios with $\eta = \infty$ and $\delta = 1$. 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 2005:M1–2021:M12.

including Riskmetrics, GARCH, HAR, and MIDAS regressions. We compare their forecasting and portfolio performance against our benchmark.

We evaluate the forecast performance in a large cross-section of 1616 stocks and, in comparison to previous studies, with a focus of one-month-ahead predictions. Our empirical analysis reveals compelling results, illustrating the substantial out-of-sample forecast performance of HAR and MIDAS models. Consequently, we observe that incorporating model-based forecasts enhances the financial performance of low-volatility investments. This improvement can be attributed to the models' enhanced ability to exclude high-risk assets more effectively. Additionally, our findings demonstrate that forecast combinations relying on model selection are particularly effective, both in terms of financial performance and forecast accuracy. These results are statistically significant and robust, even when accounting for estimated effective trading costs. The superior performance of model selection is contrary to the "forecast combination puzzle" and can be explained by our use of large cross-sectional data.

Even though our set of time series models is large, future research might still look at alternative models to improve low-volatility investments. Secondly, in this study, we have not included volatility timing strategies. Kirby and Ostdiek (2012) propose volatility timing strategies as simple active alternatives that can deliver reliable out-of-sample mean-variance performance. The parameters that control timing aggressiveness are key in these strategies. It is possible that the forecasting exercise can be extended to also incorporate these parameters.

Figure 5: Monthly portfolio weights on a subset of industries at the start of the Covid-19 pandemic



Notes: Weight on firms for which SIC codes start with 13, 37, 45, 49, 50, 55, 58, 70, 79. We depict these aggregated weights for the best-performing SE portfolio ($\eta = \infty$, $\delta = 1$) and the 12m-RV^d portfolio. The vertical line depicts March 2020.

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.

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.

Blitz, D., E. Falkenstein, and P. Van Vliet (2014). Explanations for the volatility effect: An overview based on the CAPM assumptions. *Journal of Portfolio Management 40*(3), 61–76.

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.

Claeskens, G., J. R. Magnus, A. L. Vasnev, and W. Wang (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting 32*(3), 754–762.

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 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 (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance 50*(1), 131–155.

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

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.

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.

Green, J., J. R. Hand, and X. F. Zhang (2017). The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies 30*(12), 4389–4436.

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.

Hansen, P. R. and A. Lunde (2014). Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory 30*(1), 60–93.

Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *Journal of Finance 64*(3), 1445–1477.

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.

J.P. Morgan (1996). RiskMetrics. *Technical report, 4th edition*.

Kirby, C. and B. Ostdiek (2012). It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis 47*(2), 437–467.

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. *Available at SSRN: 4161957*.

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

Ledoit, O. and M. Wolf (2011). Robust performances hypothesis testing with the variance. *Wilmott 2011*(55), 86–89.

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.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance 7*(1), 77–91.

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.

Pakel, C., N. Shephard, and K. Sheppard (2011). Nuisance parameters, composite likelihoods and a panel of GARCH models. *Statistica Sinica 21*(1), 307–329.

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.

Smith, J. and K. F. Wallis (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics 71*(3), 331–355.

Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*(6), 405–430.

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.

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

- **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

$$RV_{i,t+1:t+22} - \overline{RV}_i = b_d^P(RV_{i,t} - \overline{RV}_i) + b_w^P(RV_{i,t-4:t} - \overline{RV}_i) + b_m^P(RV_{i,t-21:t} - \overline{RV}_i) + \eta_{i,t}^P$$

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

$$\begin{aligned} RV_{i,t+1:t+22} - \overline{RV}_i &= b_d^{PL}(RV_{i,t} - \overline{RV}_i) + b_w^{PL}(RV_{i,t-4:t} - \overline{RV}_i) \\ &+ b_m^{PL}(RV_{i,t-21:t} - \overline{RV}_i) + b_q^{PL}(RV_{i,t-65:t} - \overline{RV}_i) + b_s^{PL}(RV_{i,t-131:t} - \overline{RV}_i) + \eta_{i,t}^{PL} \end{aligned}$$

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

- **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,

$$RV_{i,t+1:t+22|t} = b_0^V + b_d^V RV_{i,t} + b_w^V RV_{i,t-4:t} + b_m^V RV_{i,t-21:t} + b_{\text{vix}} VIX_t^2 + \eta_{i,t}^V.$$

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

$$\begin{aligned} RV_{i,t+1:t+22|t} =& b_0^{VL} + b_d^{VL} RV_{i,t} + b_w^{VL} RV_{i,t-4:t} + b_m^{VL} RV_{i,t-21:t} \\ &+ b_q^{VL} RV_{i,t-66:t} + b_s^{VL} RV_{i,t-132:t} + b_{\text{vix}}^{VL} VIX_t^2 + \eta_{i,t}^{VL}. \end{aligned}$$

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

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

$$h^{CG}_{i,t} = (1 - \alpha^{CG}_i - \beta^{CG}_i)\bar{\omega}_i + \alpha^{CG}_i \eta^2_{i,t-1} + \beta_i h^{CG}_{i,t-1},$$

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

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

where $h^{CG}_{mkt,t+1:t+22|t}$ and $h^{CG}_{i,t+1:t+22|t}$ 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:

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

  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

$$\varphi_l(w_1, w_2) = \frac{(l/(K+1))^{w_1 - 1} \cdot (1 - l/(K+1))^{w_2 - 1}}{\sum_{j=1}^{K} (j/(K+1))^{w_1 - 1} \cdot (1 - j/(K+1))^{w_2 - 1}}. \tag{12}$$

  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-ΔHous*, and *Factor GARCH-TS*.

- **MEM:** The Multiplicative Error Model (MEM) by Engle and Gallo (2006) employs as the dependent variable not (demeaned) returns but the demeaned square-root of the realized measure itself denoted by $RVol_{i,t}$,

$$RVol_{i,t} = h_{i,t}^{MEM} Z_{i,t}^{MEM}, \qquad Z_{i,t}^{MEM} \sim \mathcal{D}(0,1),$$

  and

$$h_{i,t}^{MEM} = (1 - \alpha_i^{MEM} - \beta_i^{MEM})\overline{RVol_i^2} + \alpha_i^{MEM} RVol_{i,t-1}^2 + \beta_i^{MEM} h_{i,t-1}^{MEM},$$

  with $\overline{RVol_i^2}$ being the average $RVol_{i,t}^2$ in the corresponding rolling estimation sample.

- **Panel MEM:** As in the Panel GARCH, we can estimate one parameter vector for all stocks in a Panel MEM model by summing up the log-likelihoods with respect to all centered conditional variance equations jointly,

$$h_{i,t}^{PMEM} = (1 - \alpha^{PMEM} - \beta^{PMEM})\overline{RVol_i^2} + \alpha^{PMEM} RVol_{i,t-1}^2 + \beta^{PMEM} h_{i,t-1}^{PMEM}.$$

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 (12) 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.

Appendix B   Additional forecast evaluation results

Table B.1: Forecast comparison of model performance starting 2015

	SE					QLIKE				
Model	Mean	Med	LR^j	$Rk^j = 1$	$Rk^j \leq 4$	Mean	Med	LR^j	$Rk^j = 1$	$Rk^j \leq 4$
Panel A: Model-based forecasts										
12m-RV^d	—	—	—	0.01	0.05	—	—	—	0.00	0.01
4y-RV^d	1.01	0.90	0.56	0.02	0.07	1.10	1.03	0.49	0.02	0.07
6m-RV^d	1.46	1.09	0.42	0.00	0.01	1.00	0.99	0.51	0.01	0.05
1m-RV^d	2.92	1.32	0.36	0.02	0.02	1.38	1.14	0.38	0.00	0.01
RM monthly, 12 months	1.00	1.00	0.55	0.01	0.04	0.99	0.98	0.64	0.00	0.01
RM monthly, 6 months	1.46	1.09	0.43	0.00	0.04	1.00	1.00	0.50	0.00	0.02
RM daily, 12 months	1.75	0.98	0.51	0.00	0.04	0.98	0.86	0.64	0.01	0.04
RM daily, 6 months	1.81	0.99	0.51	0.01	0.05	0.99	0.85	0.62	0.00	0.05
GJR-GARCH	1.29	0.75	0.69	0.01	0.04	0.88	0.84	0.61	0.00	0.04
Panel GJR-GARCH	1.30	0.80	0.64	0.01	0.02	0.97	0.85	0.62	0.00	0.05
Factor GARCH	1.18	0.80	0.71	0.01	0.05	0.89	0.83	0.64	0.02	0.06
Factor GARCH-VIX	1.04	0.70	0.73	0.01	0.02	0.82	0.77	0.70	0.01	0.01
Factor GARCH-ΔHous	1.04	0.71	0.76	0.00	0.04	0.82	0.77	0.76	0.00	0.01
Factor GARCH-TS	1.04	0.72	0.74	0.00	0.01	0.82	0.79	0.74	0.00	0.00
MEM	1.11	0.88	0.54	0.02	0.04	10.73	8.83	0.07	0.00	0.01
Panel MEM	1.17	0.82	0.58	0.01	0.06	10.49	8.63	0.15	0.01	0.02
HAR	0.64	0.36	0.89	0.01	0.23	1.19	0.51	0.90	0.01	0.31
HAR-LR	2.30	0.48	0.77	0.01	0.15	2.34	0.58	0.82	0.04	0.19
HAR-SPX	0.62	0.32	0.86	0.02	0.35	1.65	0.50	0.87	0.06	0.23
HAR-SPX-LR	2.19	0.44	0.73	0.05	0.26	2.58	0.59	0.77	0.02	0.23
HAR-VIX	0.67	0.32	0.85	0.07	0.42	1.45	0.52	0.86	0.08	0.31
HAR-VIX-LR	2.21	0.45	0.70	0.12	0.26	2.90	0.58	0.77	0.06	0.19
Panel HAR	0.42	**0.30**	**0.93**	0.06	0.48	0.51	**0.46**	**0.95**	0.10	0.60
Panel HAR-LR	**0.41**	0.32	**0.93**	**0.23**	**0.52**	**0.50**	0.46	**0.95**	**0.27**	**0.64**
MIDAS	0.53	0.33	0.89	0.14	0.38	0.54	0.50	0.93	0.08	0.40
Panel MIDAS	0.86	0.35	0.83	0.12	0.37	0.59	0.47	0.92	0.18	0.44
Panel B: Loss-based combined forecasts $\delta = 0$										
$\eta = 0$	0.61	0.42	0.88	—	—	0.63	0.60	0.93	—	—
$\eta = 1/2$ SE	0.55	0.34	0.92	—	—	0.59	0.56	0.94	—	—
QLIKE	0.62	0.41	0.88	—	—	0.61	0.59	0.93	—	—
$\eta = 1$ SE	0.54	0.31	0.92	—	—	0.56	0.53	0.94	—	—
QLIKE	0.60	0.39	0.89	—	—	0.59	0.57	0.92	—	—
$\eta = \infty$ SE	0.69	0.30	0.82	—	—	0.62	0.53	0.83	—	—
QLIKE	0.62	**0.30**	0.88	—	—	0.56	0.50	0.87	—	—
Panel C: Loss-based combined forecasts $\delta = 1$										
$\eta = 1/2$ SE	0.59	0.39	0.89	—	—	0.63	0.60	0.93	—	—
QLIKE	0.63	0.41	0.89	—	—	0.63	0.60	0.93	—	—
$\eta = 1$ SE	0.57	0.37	0.89	—	—	0.63	0.61	0.93	—	—
QLIKE	0.62	0.40	0.89	—	—	0.63	0.59	0.94	—	—
$\eta = \infty$ SE	**0.41**	0.32	**0.93**	—	—	**0.50**	**0.46**	**0.95**	—	—
QLIKE	**0.41**	0.32	**0.93**	—	—	**0.50**	**0.46**	**0.95**	—	—

Notes: The same table as Table 4 but with a shorter evaluation sample starting in 2015. For each loss function SE and QLIKE, the first two columns report the time-series mean and median of the loss ratio L^j_m/L^B_m. The column LR^j reports the proportion of months in which the cross-sectional loss L^j of model j is lower than the one of the 12m-RV^d benchmark forecast. $Rk^j \leq 1$ and $Rk^j \leq 4$ report the proportion of the respective model being the best or among the four best-performing models as measured by L^j_m. In Panel B and C, we report results for combined forecasts with $\delta = 0$ and $\delta = 1$, respectively. The combined forecast with equal weights ($\eta = 0$) is listed only for $\delta = 0$ because equal weights are independent of the smoothing parameter δ. In Panel A, numbers in bold highlight the lowest average and median loss ratio across models. Similarly, we highlight the highest LR^j, $Rk^j = 1$, and $Rk^j \leq 4$. We jointly apply the same highlighting to Panel B and C that are based on forecast combinations. The evaluation period is 2015:M1–2021:M12.

Appendix C Alternative weighting schemes

Table C.1: Returns of low-volatility portfolios—value-weighted instead of volatility-weighted

		Without TC			
		Ret	Std	SR	Δ_4
Post-hoc		11.69	13.30	0.88	—
Panel A: Model-based portfolios					
12m-RV^d		10.25	13.66	0.75	1.64
4y-RV^d		10.42	13.71	0.76	1.49
6m-RV^d		10.37	13.65	0.76	1.50
1m-RV^d		10.46	13.71	0.76	1.45
RM monthly, 12 months		10.22	13.65	0.75	1.66
RM monthly, 6 months		10.37	13.66	0.76	1.51
RM daily, 12 months		10.33	13.67	0.76	1.56
RM daily, 6 months		10.39	13.69	0.76	1.52
GJR-GARCH		10.20	13.71	0.74	1.72
Panel GJR-GARCH		10.41	13.59	0.77	1.43
Factor GARCH		10.23	13.68	0.75	1.67
Factor GARCH-VIX		10.22	13.70	0.75	1.69
Factor GARCH-ΔHous		10.30	13.67	0.75	1.59
Factor GARCH-TS		10.28	13.69	0.75	1.62
MEM		10.58	13.62	0.78	1.29
Panel MEM		10.61	13.66	0.78	1.28
HAR		10.51	13.60	0.77	1.34
HAR-LR		10.78**	13.60	0.79**	1.07
HAR-SPX		10.60*	13.55**	0.78*	1.22
HAR-SPX-LR		**10.89*** **	**13.48** **	**0.81*** **	**0.89**
HAR-VIX		10.68**	13.61	0.78**	1.18
HAR-VIX-LR		10.87**	13.51	0.80***	0.94
Panel HAR		10.52	13.60	0.77	1.33
Panel HAR-LR		10.72**	13.58	0.79**	1.12
MIDAS		10.76**	13.55	0.79***	1.06
Panel MIDAS		10.48	13.66	0.77	1.41
Panel B: Loss-based portfolios $\delta = 0$					
$\eta = 0$		10.46	13.58	0.77	1.38
$\eta = 1/2$	SE	10.44	13.60	0.77	1.41
	QLIKE	10.45	13.59*	0.77	1.39
$\eta = 1$	SE	10.42	13.60	0.77	1.43
	QLIKE	10.46	13.59*	0.77	1.39
$\eta = \infty$	SE	10.62*	13.63	0.78*	1.25
	QLIKE	10.37	13.69	0.76	1.53
Panel C: Loss-based portfolios $\delta = 1$					
$\eta = 0$		10.46	13.58	0.77	1.38
$\eta = 1/2$	SE	10.47	13.58*	0.77	1.37
	QLIKE	10.44	13.58	0.77	1.40
$\eta = 1$	SE	10.47	13.59*	0.77	1.38
	QLIKE	10.47	13.59	0.77	1.38
$\eta = \infty$	SE	**10.77** **	**13.55***	**0.79** **	**1.05**
	QLIKE	10.71**	13.58	0.79**	1.13

Notes: Average annualized excess 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 post-hoc portfolio; see Equation (10). 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) and the volatility test according to Ledoit and Wolf (2011). Statistical significance at the 10%, 5%, and 1% level are indicated by *, **, and *** respectively. In Panel A, numbers in bold represent the highest return, the lowest standard deviation, the highest SR and the lowest fee Δ_4 across all models. We jointly apply the same highlighting to all forecast combinations in Panel B and C. The evaluation period is 2005:M1–2021:M12.

x

Table C.2: Returns of low-volatility portfolios—volatility-weighted but without stock exclusion

		Without TC			
		Ret	Std	SR	Δ_4
Post-hoc		12.02	14.08	0.85	—
Panel A: Model-based portfolios					
12m-RV^d		10.71	14.84	0.72	1.76
4y-RV^d		10.79	14.93**	0.72	1.74
6m-RV^d		10.75	14.73***	0.73	1.65
1m-RV^d		10.84	**14.48*** **	0.75**	1.41
RM monthly, 12 months		10.72	14.83	0.72	1.75
RM monthly, 6 months		10.75	14.73***	0.73	1.65
RM daily, 12 months		10.79	14.72**	0.73	1.60
RM daily, 6 months		10.79	14.69***	0.73	1.58
GJR-GARCH		10.75	14.64***	0.73*	1.60
Panel GJR-GARCH		10.73	14.69**	0.73	1.65
Factor GARCH		10.84	14.85	0.73	1.63
Factor GARCH-VIX		10.86*	14.90	0.73	1.64
Factor GARCH-ΔHous		10.88*	14.91*	0.73	1.64
Factor GARCH-TS		10.86*	14.92*	0.73	1.65
MEM		**11.19** **	14.57***	**0.77*** **	**1.12**
Panel MEM		11.04**	14.68***	0.75***	1.33
HAR		10.84	14.58***	0.74***	1.47
HAR-LR		10.97	14.54***	0.75**	1.32
HAR-SPX		10.88	14.65**	0.74**	1.47
HAR-SPX-LR		10.94	14.52**	0.75**	1.33
HAR-VIX		10.94*	14.64***	0.75***	1.40
HAR-VIX-LR		10.99	14.53**	0.76**	1.29
Panel HAR		11.00**	14.70***	0.75***	1.38
Panel HAR-LR		11.06**	14.65**	0.76***	1.29
MIDAS		11.01**	14.66***	0.75***	1.35
Panel MIDAS		10.93	14.70**	0.74**	1.46
Panel B: Loss-based portfolios $\delta = 0$					
$\eta = 0$		10.89**	14.76**	0.74***	1.53
$\eta = 1/2$	SE	10.89**	14.77*	0.74***	1.53
	QLIKE	10.89**	14.76**	0.74***	1.53
$\eta = 1$	SE	10.90**	14.77*	0.74***	1.53
	QLIKE	10.89**	14.76**	0.74***	1.53
$\eta = \infty$	SE	10.95**	14.68***	0.75***	1.43
	QLIKE	10.87	14.71***	0.74***	1.52
Panel C: Loss-based portfolios $\delta = 1$					
$\eta = 0$		10.89**	14.76**	0.74***	1.53
$\eta = 1/2$	SE	10.88**	14.76**	0.74***	1.54
	QLIKE	10.89**	14.76**	0.74***	1.53
$\eta = 1$	SE	10.88**	14.77**	0.74***	1.54
	QLIKE	10.89**	14.75**	0.74***	1.53
$\eta = \infty$	SE	**11.07** **	**14.64*** **	**0.76*** **	**1.28**
	QLIKE	11.05**	14.65**	0.75***	1.31

Notes: Average annualized excess 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 post-hoc portfolio; see Equation (10). 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) and the volatility test according to Ledoit and Wolf (2011). Statistical significance at the 10%, 5%, and 1% level are indicated by *, **, and *** respectively. In Panel A, numbers in bold represent the highest return, the lowest standard deviation, the highest SR and the lowest fee Δ_4 across all models. We jointly apply the same highlighting to all forecast combinations in Panel B and C. The evaluation period is 2005:M1–2021:M12.