Erasmus University Rotterdam
Erasmus School of Economics
Master Thesis Quantitative Finance

Volatility Timing using Forecasted Realized Variance

Giuseppe Gianfranceschi (619050)

Supervisor:	dr. Rasmus Lönn
Second assessor:	Pierluigi Vallarino
Date final version:	31st July 2024

Abstract

This thesis investigates the effectiveness of volatility timing using forecasted realized variance to enhance the performance of factor-based portfolios. Building upon the foundational work of Moreira and Muir (2017), which demonstrated the potential benefits of volatility-managed portfolios, this research explores the use of volatility forecasting models including GARCH, HAR-RV, ARFIMA, MIDAS, Random Forest and two forecast combination models. The study aims to determine whether these forecasting methods can improve risk-adjusted returns compared to traditional factor strategies and the original volatility-managed approach.

The empirical analysis considers returns of major factors like market (MKT), size (SMB), value (HML), profitability (RMW), investment (CMA) and momentum (MOM) spanning from 1966 to 2022. By incorporating transaction costs into the performance evaluation, this thesis assesses the practical viability of volatility-managed portfolios.

The results indicate that while volatility timing with forecasted realized variance offers theoretical benefits for some of the factors under consideration, these strategies fail to significantly improve mean variance characteristics over the traditional factor strategies when transaction costs are taken into account.

Contents

1 Introduction

Portfolio management plays a crucial role in the financial industry, as it involves the strategic allocation of assets to achieve specific investment goals while managing risk. Effective portfolio management is essential for investors seeking to maximize returns and preserve wealth in the face of market uncertainties.

Traditional factor investing, which involves constructing portfolios based on well-established factors such as value, momentum, and size, has been a popular approach in the portfolio management industry. However, these traditional factors are shown to have experienced attenuation since their discovery (Chordia et al., 2014; Linnainmaa & Roberts, 2018). This decrease in performance can be attributed, among other things, to the crowding effect as more investors adopt similar strategies (Chordia et al., 2014), and the time-varying nature of factor premia (Moreira & Muir, 2017).

In light of these challenges, the concept of volatility-managed portfolios (VMP), as introduced by Moreira and Muir (2017), has gained significant attention in the literature. Their work proposes a novel approach to portfolio construction, where the exposure to a given factor is dynamically adjusted based on the factor's realized variance (RV). By scaling factor exposures inversely proportional to their realized variance, the authors demonstrate that volatility-managed portfolios can achieve superior risk-adjusted returns compared to their unmanaged counterparts.

While the findings of Moreira and Muir (2017) have generated considerable interest, subsequent research has highlighted some limitations of their approach when applied out-of-sample. Cederburg et al. (2020) take a broader sample of factors and find no evidence for a systematic improvement in Sharpe ratios for volatility-managed portfolios. They also attribute the findings of Moreira and Muir (2017) to estimation error and argue that their results are not implementable in real-time. Barroso and Detzel (2021) analyse the net-of-costs performance of volatility-managed portfolios using a more sophisticated transaction cost model and find that after accounting for costs, none of the volatility-managed portfolios generate positive alphas and most have significantly smaller Sharpe ratios compared to their unmanaged counterparts. Thus, the performance of volatility-managed portfolios has been found to be less impressive in real-world settings, suggesting that further improvements and refinements are necessary.

This thesis aims to build upon the foundational work of (Moreira & Muir, 2017) by exploring potential enhancements to the volatility-managed portfolio framework. By considering forecasted realized variance as a scaling factor and by introducing machine learning and combiner models to forecast factor realized variance we aim to improve the performance of the volatility-managed portfolios and develop strategies that are implementable in the real world. Furthermore, examining the impact of transaction costs, this research seeks to develop a robust analysis of the practical performance of the volatility-managed strategies to see if they hold up out-of-sample.

An important consideration in portfolio management is the impact of transaction costs on the performance of trading strategies. Transaction costs can have a significant impact on the returns of a portfolio, especially since the volatility-managed portfolios under consideration are rebalanced monthly. By incorporating estimated asset-level transaction costs following Hasbrouck (2009), this research aims to provide a realistic assessment of the performance of volatility-managed portfolios in practice.

We employ a set of volatility forecasting models: GARCH, HAR-RV, ARFIMA, MIDAS, Random forest and two combiner models: mean forecast combination and stacking, to identify the most effective techniques for forecasting realized variance with the aim of producing favourable timing rules. The performance of these models will be evaluated based on their ability to improve the risk-adjusted returns of volatility-managed portfolios. Additionally, net-of-costs returns will be assessed to provide a realistic assessment of their practical viability. We select the optimal model specification for each of the model classes which result in seven different volatility forecasting models. We then use the forecasted realized variance from these models to construct timing rules using different cost-mitigation techniques like scaling by realized volatility or by introducing leverage constraints. These timing rules are then used to construct volatility-managed portfolios which are compared to their unmanaged counterparts.

Having analysed six well-known factors: market (MKT), size (SMB), value (HML), profitability (RMW), investment (CMA) and momentum (MOM) with out-of-sample period 1976-2022, we find none of the generated timing rules to produce significantly better risk-adjusted returns compared to the traditional factor strategies. In fact, for most factors, these strategies lead to significantly worse performance compared to their unmanaged counterparts. Interestingly, we also find that when one would have a 'perfect' forecast of next month's realized variance to be used as a timing signal, the Sharpe ratio of the market, size and momentum factor increases substantially while the Sharpe ratio of the value, profitability and investment factor decreases substantially.

The rest of this thesis is structured as follows: Section 2 provides an overview of the relevant literature on volatility-managed portfolios, volatility forecasting, model validation and transaction cost modelling. Section 3 describes the data used in this study and its sources. Section 4 presents a preliminary analysis of the data, including summary statistics on the performance of volatility-managed portfolios without the use of volatility forecasting. Section 5 outlines the volatility forecasting models used in this research and how they are tuned. Section 6 discusses the performance evaluation metrics used to compare different volatility timing techniques based on forecasted realized variance as well as how their significance is tested. Section 7 presents the empirical results of the volatility-managed portfolios and the results of the model selection/tuning. Finally, Section 8 concludes the thesis with a summary of the key findings and suggestions for future research.

2 Literature

2.1 Volatility-managed portfolios

The concept of volatility-managed portfolios is introduced by Moreira and Muir (2017). They construct managed portfolios based on nine factors[1] where the investment in each factor is scaled by the inverse of the factor's realized variance. They motivate the choice for scaling by realized variance from the perspective of a mean-variance investor, whose optimal portfolio weight is proportional to the risk-return trade-off. Since empirical evidence shows that volatility is highly variable, persistent and does not predict returns, they argue that taking the inverse of the conditional variance is a good proxy for the risk-return trade-off. Portfolio construction is then further simplified by using the realized variance as an approximation for the conditional variance.

They find that most of the volatility-managed portfolios outperform the unmanaged portfolios in terms of alpha and Sharpe ratio. Furthermore, they also show that the managed portfolios have lower beta's to their unmanaged counterparts during NBER recessions, thus decreasing the managed portfolios' exposure during historically volatile periods. After accounting for transactions costs and implementing various transaction cost mitigation techniques, they show that their managed market factor still produces positive alphas.

Moreira and Muir (2017) also show that the alphas produced by volatility-managed portfolios are a direct measure of the comovement between risk premium and volatility. They show that these alphas thus allow for the reconstruction of the variation in the price of risk due to volatility for individual risk drivers. They also compare volatility timed portfolio on simulated return data using 4 leading asset pricing models. From this, they find that at most only in 0.2% of the simulated samples achieve matching alphas to the alphas recovered from historic data. Thus demonstrating that their volatility-managed portfolios pose a fresh challenge to these asset pricing models.

Cederburg et al. (2020) dispute the findings of Moreira and Muir (2017). They argue that the superior performance of the managed portfolios is the result of an estimation error. This estimation error is due to the fact that the performance increase is evaluated using a spanning regression intercept. Cederburg et al. (2020) argue that this implies an optimal combination of both unmanaged and managed portfolios with weights that are not known in real-time. Therefore, the strategy is not implementable in practice. They also argue that evaluating performance gains of the managed portfolios by quantifying increase in Sharpe ratio is more appropriate.

Barroso and Detzel (2021) further dispute the findings of Moreira and Muir (2017). In their study they find that after accounting for transaction costs and despite employing six cost-mitigation techniques, the volatility-managed portfolios do not produce significant positive alphas (except for the momentum factor).

Another finding of Barroso and Detzel (2021) is that following high realizations of the (M.

[1] The market factor (MKT), size factor (SMB) and value factor (HML) (Fama & French, 1993).
The momentum factor (MOM), profitability factor (RMW) and investment factor (CMA) (Fama & French, 2015).
The investment factor (IA) and return on equity factor (ROE) (Hou et al., 2015).
The betting-against-beta (BAB) factor (Frazzini & Pedersen, 2014).

Baker & Wurgler, 2006) sentiment index[2], volatility-managed portfolio Sharpe ratios more than doubled. On the other hand, following low realizations of the index, Sharpe ratios were reduced. This finding aligns with the work of Yu and Yuan (2011), who demonstrated that investor sentiment significantly affects the mean-variance trade-off in the stock market. Specifically, Yu and Yuan (2011) found a strong positive relationship between returns and conditional volatility in low-sentiment periods, but a weak and sometimes negative relationship in high-sentiment periods.

DeMiguel et al. (2021) propose a multifactor perspective on volatility-managed portfolios. They show that a portfolio constructed from the nine factors considered by Moreira and Muir (2017) produces positive alphas, even after accounting for transaction costs. The multifactor portfolios are formed by weighting each factor by the inverse of the market volatility and the factors' exposure to this volatility. They also find that their proposed multifactor strategy performs better than the unmanaged portfolios during both high- and low-sentiment periods, in contrast to the findings of Barroso and Detzel (2021).

2.2 Volatility forecasting

Volatility forecasting is a primary focus of this thesis. There exist various methods to forecast volatility, ranging from simple autoregressive models to more complex machine learning models. The literature on volatility forecasting is vast, and we will only highlight a few key studies here.

One of the more common models used for volatility modelling is the heterogeneous autoregressive model of realized volatility (HAR-RV) (Corsi, 2009). Y. Wang et al. (2016) forecast the realized volatility of the S&P-500 using HAR-RV model and various extensions. They combine several of these extensions using dynamic model averaging (DMA) as well as Bayesian model averaging (BMA) and mean forecast combinations (MFC). They not only evaluate model performance using popular error metrics such as MSE/MAE but also evaluate portfolio performance which is constructed by considering a mean-variance utility investor. Likewise, they find that the HAR-RV model which allow for time varying parameters performs best while the model specification with constant parameters performs worst.

A lesser known method, not only for volatility forecasting but also for time series forecasting in general, is the use of fractionally integrated autoregressive moving average (ARFIMA) models. Chen et al. (2019) perform a study regarding volatility scaling portfolios where they forecast realized volatility using a handful of ARFIMA models. They find that using a fractionally integrated ARMA model to forecast realized volatility and scaling the factor returns by the forecasted volatility leads to significantly improved portfolio performance compared to using historic realized volatility to scale the factor returns.

Most models in the literature focus on forecasting one-step ahead volatility. However, in practice, it is often more useful to forecast volatility over multiple periods. Ghysels et al. (2009) conduct a comparative study of several approaches of producing multi-period forecasts of volatility. They compare iterated, direct and mixed data sampling approaches as well as the "scaling-up" method. The latter involves producing a one-step ahead forecast and scaling it

[2] The sentiment index by M. Baker and Wurgler (2006) is a composite index formed from several proxies of investor sentiment: the closed-end fund discount, NYSE share turnover, the number and average first-day returns on IPOs, the equity share in new issues, and the dividend premium (M. Baker & Wurgler, 2006).

to the desired forecast horizon. They find that mixed data sampling (MIDAS) dominates all other techniques from a horizon of 10 steps ahead and onwards. The performance dominance of MIDAS is found to be most apparent at horizons of 30 periods and longer.

Machine learning models have also been applied to the problem of volatility forecasting. Christensen et al. (2023) evaluate the performance of different machine learning models in forecasting the Dow Jones Industrial Average index. They compare these models' performance against multiple HAR models and find that the ML models are competitive with the HAR models. They also find that the ML models are able to capture the time-varying nature of volatility better than the HAR models. In particular, they find Random Forest and Neural Network models to outperform the HAR models. As regressors, they use lagged realized volatility (daily, weekly and monthly), the CBOE volatility (VIX) index, the Hang Seng stock index daily squared log-return (HSI), the Aruoba et al. (2009) business conditions (ADS) index, the US 3-month T-bill rate (US3M), and the economic policy uncertainty (EPU) index from S. R. Baker et al. (2016).

2.3 Model validation

In the literature, model validation is most often performed using cross-validation techniques. This technique however, is not widely applied for time series data. Instead, most researchers opt for the use of out-of-sample testing. Bergmeir and Benítez (2012) and Bergmeir et al. (2018) counter this choice and instead argue that cross-validation techniques can successfully be applied on time series data and models. They show that when the time series models that are used are purely autoregressive (which is the case for most machine learning techniques) and the data is stationary, cross-validation can be applied successfully. This has the benefit of being able to use more of the available data for fitting and producing more reliable estimates of the model's performance. Only when residuals are heavily serially correlated is cross-validation not recommended. The authors argue that this can easily be tested against by checking the residuals of the model. In such cases traditional out-of-sample testing is recommended.

2.4 Transaction cost modelling

Our goal is to make a real-world practical assessment of the performance of volatility-managed portfolios. This requires us to account for transaction costs. Transaction costs are an important aspect of portfolio management and can have a significant impact on the performance of a strategy.

One approach to transaction cost modelling is to estimate transaction costs using bid-ask spreads. These can be estimated using high-frequency data, usually from the TAQ database. However, this data is not always readily available and can be expensive to acquire. Additionally, for the purpose of estimating transaction costs on the factors used in this thesis, one would need to acquire high-frequency data on the whole universe of CRSP stocks. This would require acquiring and processing tremendous amounts of data and is therefore not feasible for this purpose.

Alternatively, a widely used technique is that of Hasbrouck (2009) to estimate the bid-ask spreads. This technique involves using Bayesian Gibbs sampler to estimate trading costs using daily returns. The estimated effective spreads are shown by Hasbrouck (2009) to have a 96.5% correlation with trading costs estimated from high-frequency data of 300 firms during a period spanning from 1993 to 2005. This procedure has limitations since it does not account for the price impact of large trades. Therefore, it is only effective in estimating costs faces by smaller investors. An advantage however, is that it is easy to implement for all CRSP stocks as it only requires daily returns data. Code for this procedure can be found on the author's website.

Corwin and Schultz (2012) present another way to estimate bid-ask spreads using daily data. They show that the bid-ask spread can be estimated using the high, low and close prices of a stock. Similarly to Hasbrouck (2009), this method is easy to implement and can be applied to all CRSP stocks. The authors show that their method is able to estimate the bid-ask spread with a high degree of accuracy.

While spread estimation methods provide insights into transaction costs faced by smaller investors, they fail to capture the impact of larger trades on market prices. A complementary approach to transaction cost modelling is estimating Kyle's lambda (Kyle, 1985), which measures the price impact of trading. This method, used by studies such as Novy-Marx and Velikov (2016) and Barroso and Detzel (2021), aims to model how larger trades affect market prices. This can give insight into how the attraction of large capital can affect the profitability of a strategy.

For the purposes of this thesis, we employ the Hasbrouck estimator to model transaction costs. This choice is motivated by its relative ease of implementation using readily available daily returns data and its use in the work of Barroso and Detzel (2021). By adopting the same methodology, we enable a more direct comparison of our results with their findings. We also limit ourselves to the estimation of transaction costs for smaller investors and therefore do not consider the price impact of large trades.

3 Data

Moreira and Muir (2017) do not find evidence that any particular factors are a significant driver of their findings. They argue that their volatility timing strategy can be applied to any individual asset, but only becomes economically interesting when applied to a structural risk driver. In their paper they consider the five Fama-French factors (Fama & French, 2015): excess market return (MKT), size (SMB), value (HML), profitability (RMW) and investment (CMA), the momentum factor (MOM) of Jegadeesh and Titman (1993), the Hou et al. (2015) Q-factors: investment (INV) and return on equity (ROE) and the betting-against-beta (BAB) factor of Frazzini and Pedersen (2014). In this thesis we limit the analysis to the five Fama-French factors and the momentum factor. Returns for these factors can be obtained from the Kenneth R. French data library. An overview of the factors and their coverage is presented in Table 1.

Table 1
Data sources and availability for the factors used in the construction of the volatility-managed portfolios

Factor	Data source	Data availability
MKT	Kenneth R. French data library	1963–07 to 2024–01
SMB	Kenneth R. French data library	1963–07 to 2024–01
HML	Kenneth R. French data library	1963–07 to 2024–01
RMW	Kenneth R. French data library	1963–07 to 2024–01
CMA	Kenneth R. French data library	1963–07 to 2024–01
MOM	Kenneth R. French data library	1927–01 to 2024–01

To facilitate the cost analysis, we need to know the asset weights used for the construction of each of the factors. Since these are not made available through the data sources where the factor returns are obtained from, we need to replicate them ourselves. To this end, stock level data is needed. Stock price data as well as shares outstanding is obtained from CRSP, which is used to calculate market capitalization. In addition to this, accounting data is needed to calculate factor portfolio weights for specific factors. This accounting data is obtained from the COMPUSTAT library. A list of accounting data which is needed to form the factor portfolios is presented in Table 2.

Table 2
Data sources and availability for the accounting data needed to form the factor portfolios.

Accounting data	Needed for factor	Data source
Equity	SMB, HML	COMPUSTAT: Stockholders Equity - Total (teq)
Operating profitability	RMW	COMPUSTAT: Gross Profit (Loss) (gp)
Assets	CMA	COMPUSTAT: Assets - Total (at)

COMPUSTAT refers to WRDS: Compustat Daily Updates - Fundamentals Annually. All mentioned data is available from 1961–03 to 2024–03.

Sentiment index

Barroso and Detzel (2021) found that the M. Baker and Wurgler (2006) sentiment index has an effect on the performance of the volatility-managed portfolios. Specifically, following high (low) realizations of the index, Sharpe ratios of the managed portfolios doubled (declined). Therefore, it is interesting to include this index in the analysis in order to construct trading rules based on its realizations. The sentiment index is available from 1965–12 to 2022–06 with monthly frequency from the website of Jeffrey Wurgler.

Additional covariates

Christensen et al. (2023) model realized volatility using a Random Forest model and find their method to outperform traditional methods. For the construction of this model they add additional covariates in addition to lagged realizations of realized variance. These additional covariates are the VIX index, the ADS index (Aruoba et al., 2009), the EPU index (S. R. Baker et al., 2016), the Hang Seng stock index daily returns and the 3-month T-bill rate. The data sources and availability for these covariates are presented in Table 3.

Table 3
Data sources and availability for the additional covariates used in the analysis.

Covariate	Data source	Data availability
VIX index	CBOE dataset on WRDS	1986–01 to 2024–02
ADS index	Federal Reserve Bank Philadelphia	1960–03 to 2024–02
EPU index	Economic Policy Uncertainty	1985–01 to 2024–02
Hang Seng stock index	Yahoo Finance	1987–01 to 2024–02
3-month T-bill rate	Federal Reserve Bank of St. Louis	1934–01 to 2024–02

4 Preliminary analysis

4.1 Data processing

The relevant data as described in Section 3 is collected and processed resulting in daily and monthly returns for each of the six factors. We set the time period of the data to be 1966-01 to 2022-06 in order to have the greatest coverage of all variables of interest and keep as large a sample space as possible. The VIX, HSI and EPU indices are excluded from the analysis due to data availability since they are found to not have a substantial impact on the performance of the Random Forest model and were therefore too large of a limiting factor for the overall sample size.

For the volatility scaling as devised by Moreira and Muir (2017), monthly realized variance is calculated for each factor using it's daily returns. Monthly realized variance for factor f in month τ is computed as:

$$\sigma_\tau^2(f) = \hat{RV}_\tau^{(\mathrm{m})}(f) = \sum_{i=0}^{21} \left(r_{t-i}(f) - \frac{\sum_{i=0}^{22} r_{t-i}(f)}{22} \right)^2, \tag{1}$$

where $r_t(f)$ is the daily return of factor f on day t, where t is taken to be the last day of month τ. The superscript (m) denotes that the realized variance is calculated on a monthly basis. The realized variance is calculated over the last 22 days of the month as this is the average number of trading days in a month.

In an effort to improve the performance of the volatility-managed portfolios, we forecast realized monthly variance using different time series models and use the forecasted values to form the managed portfolios. Some of these models target daily realized variance as the response variable. Since daily variance is not directly observed, we use the squared daily return as a proxy for the daily realized variance. Using intraday returns to estimate daily realized variance is a common practice in the literature, since this provides better estimated realized variances (Andersen & Bollerslev, 1998). However, since the factors in question are formed on the whole universe of US stocks, it is computationally infeasible to calculate intraday returns for each asset. Therefore, we use the daily returns as a proxy for the daily realized variance:

$$\hat{RV}_t^{(\mathrm{d})}(f) = r_t^2(f), \tag{2}$$

where $r_t(f)$ is the daily return of factor f on day t. The superscript (d) denotes that the realized variance is calculated on a daily basis.

4.2 Volatility scaling

Volatility scaling as introduced by Moreira and Muir (2017) scales exposure to factor portfolios by the previous month's realized variance of that portfolio. To follow the method used by Moreira and Muir (2017) in constructing the volatility-managed portfolios, the realized variance for factor f in month τ is computed as in Equation (1). Given the unscaled portfolio return $R_\tau(f)$ in month

τ, the scaled portfolio is then constructed as:

$$R_\tau(f^\sigma) = \frac{c}{\sigma^2_{\tau-1}(f)} R_\tau(f), \tag{3}$$

where f^σ denotes the volatility-managed factor and c is a constant and is chosen such that the managed portfolio has the same unconditional variance as the unmanaged portfolio, ensuring that the overall risk level of the strategy does not change:

$$c = \sqrt{\frac{\mathbb{V}\left(R_\tau(f)\right)}{\mathbb{V}\left(\frac{1}{\sigma^2_{\tau-1}(f)} R_\tau(f)\right)}}. \tag{4}$$

Since c is a constant, it has no effect on the Sharpe ratio of the managed portfolio. Therefore, the use of the full sample in setting c has no effect on the outcome of the analysis.

In addition to using the past realized variance for constructing the managed portfolios, we also construct managed portfolios using forecasted realized variance. Using the forecasted one-period ahead monthly realized variance $\hat{RV}_{\tau+1}(f)$, the managed portfolio is constructed as:

$$R_\tau(f^\sigma) = \frac{c}{\hat{RV}_{\tau+1}(f)} R_\tau(f), \tag{5}$$

where c is again chosen such that the managed portfolio has the same unconditional variance as the unmanaged portfolio as described by Equation (4).

4.3 Transaction costs

In order to evaluate the real world performance of our managed portfolios compared to the unmanaged portfolios, we need to take transaction costs into account. As found by Barroso and Detzel (2021), forming volatility-managed portfolios using prior months realized variance can lead to a turnover increase of as much as 15 times relative to the unmanaged portfolios. This increase in turnover can lead to substantial transaction costs and therefore could have a non-negligible effect on net-of-costs performance. To this effect we estimate transaction costs for each of the formed portfolios by estimating trading spreads and determining transaction costs using the model developed by Hasbrouck (2009). We restrict ourselves to the case in which small investors are trading and therefore do not consider the price impacts of a given strategy attracting large capital. Since the estimated trading costs of the Hasbrouck (2009) procedure are an effective measure for costs faced by small investors and does not account for the price impact of large trades, this will be sufficient for our purposes.

4.3.1 Determining turnover

To determine transaction costs, we construct portfolio weights using the original authors' methodology. The procedure for constructing each factor is outlined in Appendix E. Monthly turnover for each asset in the factor strategy can then be calculated. Turnover is defined as the difference in weights at the end of the preceding month and the beginning of the current month. Turnover for asset i in factor f at the start of month τ is calculated as:

$$TO^i_{f,\tau} = \left| L_\tau(f) w^i_{f,\tau} - L_{\tau-1}(f) w^i_{f,\tau^-} \right| \quad \text{for } i = 1, \ldots, N_{f,\tau}, \tag{6}$$

where $N_{f,\tau}$ is the number of assets in the portfolio f in month τ, $L_\tau(f)$ is the leverage or scaling term in month τ as described in Equation (3)[3], $w^i_{f,\tau}$ is the weight of asset i in month τ in portfolio f after rebalancing and w^i_{f,τ^-} is the weight of asset i in month τ in portfolio f just before rebalancing:

$$w^i_{f,\tau^-} = \left(\sum_{j \in f_{S,\tau-1}} w^j_{f,\tau-1}(1 + r_{j,\tau-1}) \right)^{-1} w^i_{f,\tau-1} \cdot (1 + r_{i,\tau-1}) \quad \text{for } i \in f_{S,\tau-1}, \text{ and} \tag{7}$$

$$w^i_{f,\tau^-} = \left(\sum_{j \in f_{L,\tau-1}} w^j_{f,\tau-1}(1 + r_{j,\tau-1}) \right)^{-1} w^i_{f,\tau-1} \cdot (1 + r_{i,\tau-1}) \quad \text{for } i \in f_{L,\tau-1}, \tag{8}$$

where $r_{f,\tau-1}$ is the return of the factor in month $\tau - 1$ and $r_{i,\tau-1}$ is the return of asset i in month $\tau - 1$ and $f_{S,\tau}$ and $f_{L,\tau}$ are the short and long leg of the factor in month τ respectively.

4.3.2 Transaction cost estimation

In the same manner as Barroso and Detzel (2021) we estimate trading spreads using the model developed by Hasbrouck (2009). The procedure estimates costs using a Bayesian Gibbs sampler on a generalized Roll (1984) model of stock price dynamics. An outline of the procedure is presented in Appendix D.

The Hasbrouck procedure produces cost estimates for each stock in the CRSP database. These estimates are updated for each year. We use these cost estimates to calculate the monthly estimated incurred transaction costs for each portfolio. The monthly transaction costs for factor f in month τ is calculated as:

$$TC_{f,\tau} = \sum_{i=1}^{N_{f,\tau}} TO^i_{f,\tau} \cdot \hat{TC}^i_y, \tag{9}$$

where $\hat{TC}^i_y$ is the estimated transaction costs for asset i in year y and $TO^i_{f,\tau}$ is the turnover for asset i in factor f in month τ.

We fill in missing transaction cost estimates by taking the transaction costs of the closest asset in terms of market capitalization for the corresponding year.

4.3.3 Net-of-costs returns

Using the transaction costs estimates, we calculate the net-of-costs returns for each managed portfolio. We take the net-of-costs return of the managed portfolios by subtracting the rebalancing costs from the gross returns for each month:

$$R^*_\tau(f^\sigma) = R_\tau(f^\sigma) - TC_{f^\sigma,\tau}. \tag{10}$$

[3]For the unmanaged factor portfolios, leverage is set equal to 1 for all months.

Using the gross returns and net-of-costs returns, we can calculate both gross and net-of-costs annualized Sharpe ratios of the portfolios as:

$$SR(f^\sigma) = \frac{\mathbb{E}\left[R_\tau(f^\sigma)\right]}{\sqrt{\mathbb{V}\left[R_\tau(f^\sigma)\right]}} \cdot \sqrt{12}, \text{ and} \tag{11}$$

$$SR^*(f^\sigma) = \frac{\mathbb{E}\left[R^*_\tau(f^\sigma)\right]}{\sqrt{\mathbb{V}\left[R_\tau(f^\sigma)\right]}} \cdot \sqrt{12}. \tag{12}$$

4.4 Preliminary results

We start by analysing the performance of the unmanaged and managed factors over the sample period using the scaling technique of Moreira and Muir (2017). This gives us a baseline result on which we can evaluate the performance of the forecasted variance managed portfolios. We analyse both the gross and net-of-costs performance of the factors and managed portfolios.

4.4.1 Performance of unmanaged portfolios

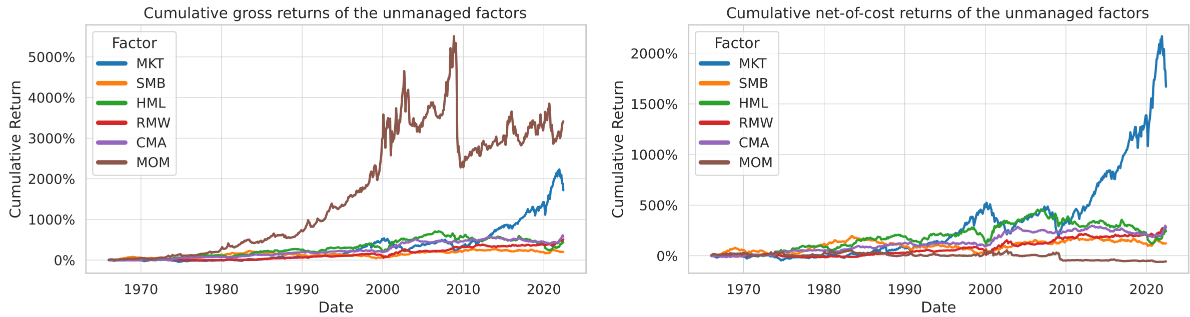
We start by evaluating the performance of the unmanaged portfolios. The unmanaged portfolios consist of the six factors that were selected in Section 3. We look at their performance over the whole sample (1966-01 to 2022-06) both gross and net of costs. The cumulative gross and net-of-costs returns are shown in Figure 1a and Figure 1b respectively. Annualized mean return and Sharpe ratios are presented in Table 4.

Table 4
Summary statistics for the unmanaged factors

Both gross and net-of-costs Sharpe ratios are reported. Transaction costs are reported as the average monthly incurred transaction costs in percentages.

	Sharpe ratio		Transaction costs (%)
	Gross	Net-of-costs	
MKT	0.406	0.403	0.004
SMB	0.238	0.187	0.045
HML	0.336	0.260	0.066
RMW	0.451	0.346	0.068
CMA	0.511	0.367	0.085
MOM	0.504	-0.021	0.647

What is apparent is that all factors produce mean positive returns over the sample period when only their gross returns are taken into account. However, when considering trading costs, the mean return of the momentum factor become negative. This is not surprising since the momentum factor is rebalanced monthly, while the other factors (except the market factor) are rebalanced annually. This shows that trading costs can have a significant influence on the performance of the factors. The transaction costs for the unmanaged portfolios are in line with those found by Novy-Marx and Velikov (2016).



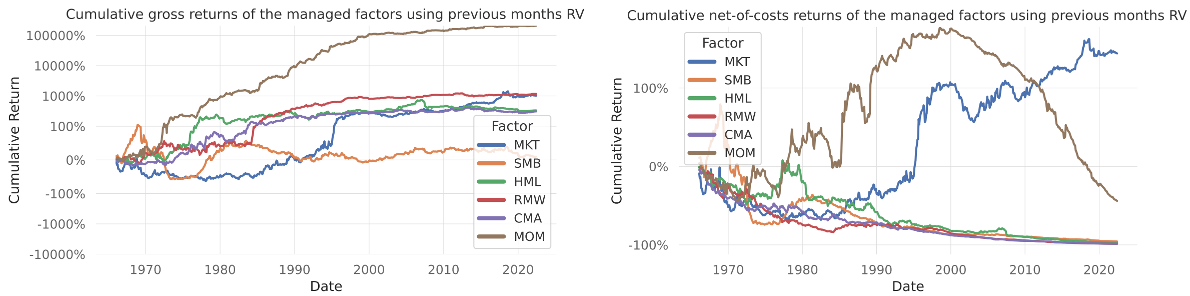
(a) Gross returns

(b) Net-of-costs returns

Figure 1
Cumulative returns of unmanaged portfolios

4.4.2 Performance of volatility-managed portfolios

Next we look at the performance of the volatility-managed portfolios as specified by Moreira and Muir (2017). We construct the managed portfolios using the realized variance of the previous month as described in Section 4.2. The cumulative gross and net-of-costs returns over time are shown in Figure 2a and Figure 2b respectively.



(a) Gross returns

(b) Net-of-costs returns

Figure 2
Cumulative returns of managed portfolios
The managed portfolios are constructed using the previous months' realized variance as the scaling factor.

In addition to the managed portfolios constructed with the previous months realized variance, we also construct managed portfolios using the next months realized variance. This signifies the case in which we would have a 'perfect' forecast of the realized variance for the coming month and would be able to leverage our portfolio in a 'perfect' response to the coming month's volatility. Annualized Sharpe ratios (gross and net-of-costs) for both the case using lagged realized variance and next months realized variance as the scaling are presented in Table 5.

We find that the gross Sharpe ratios of the volatility-managed portfolios using lagged monthly realized variance align with those found by Cederburg et al. (2020). An interesting observation is that having a 'perfect' forecast of the coming months realized variance does not universally lead to higher Sharpe ratios. For half of the factors (HML, RMW, CMA) does the 'perfect' forecast lead to a lower Sharpe ratio, with or without taking transaction costs into account. While for the other half, it leads to an increase in Sharpe ratio. The net-of-costs Sharpe ratio of the Market factor even triples. This is surprising result since one would expect that having

Table 5
Sharpe ratios for managed factors using lagged monthly realized variance and actual monthly realized variance.

Scaling by the actual monthly variance is equivalent to having a perfect forecast of the coming month's variance for construction of the volatility managed portfolios. Both gross and net-of-cost Sharpe ratios are reported. Transaction costs are reported as the average monthly incurred transaction costs in percentages.

	Lagged monthly RV			Actual monthly RV		
	Gross SR	Net-of-costs SR	Transaction costs (%)	Gross SR	Net-of-costs SR	Transaction costs (%)
MKT	0.352	0.238	0.149	1.068	0.952	0.152
SMB	0.070	-0.467	0.475	0.409	-0.179	0.521
HML	0.302	-0.562	0.751	0.192	-0.835	0.894
RMW	0.619	-0.941	1.009	0.524	-1.183	1.106
CMA	0.386	-1.017	0.831	0.177	-1.166	0.797
MOM	0.991	0.002	1.217	1.201	0.223	1.205

a 'perfect' forecast of the coming months realized variance would enable better hedging against market uncertainties and therefore lead to a higher Sharpe ratio. Why this is not the case for all factors is left to future research.

5 Volatility forecasting

We add to the literature by forming volatility-managed portfolios using forecasted realized variance. We use a variety of models to forecast realized variance and compare the performance of the managed portfolios using forecasted realized variance to the managed portfolios using past realized variance. The models used for forecasting realized variance are the GARCH(1,1) model, HAR-RV model, ARFIMA model, MIDAS model, and Random Forest. We also consider combiner models to combine the forecasts of the different models, Mean Forecast Combination and Stacking. We start by defining how model selection and hyperparameter tuning are performed, then we introduce the different model specifications.

5.1 Model selection and hyperparameter tuning

Some models in use either have multiple functional forms based on their specification or have hyperparameters that need to be tuned. In an effort to keep the comparison between managed portfolios using forecasted realized variance and managed portfolios using past realized variance uncluttered, we apply a model selection procedure to determine the optimal model specification / hyperparameters for each model class. This serves as a filtering step to select the best model for each factor within each model class and does not serve as an evaluation of the model's ability to produce meaningful timing rules. Therefore, no statistical tests on the significance of a difference in performance metrics are needed at this stage. That analysis is performed after selection of the optimal model specifications.

Due to the use of two autoregressive models (GARCH and ARFIMA), it would not be suitable to perform k-fold cross validation (Bergmeir & Benítez, 2012). Instead, we apply traditional out-of-sample evaluation. A rolling window of 5 years is formed over the dataset which is moved one month every iteration. This simulates a real-world scenario where an investor rebalances their portfolio every month using the most recently available data. This produces out-of-sample forecasts for each model specification for the realized variance of each month from 1971–01 to 2022–06 for the first five models and 1976–01 to 2022-06 for the combiner models.

Since we are not particularly interested in producing the most accurate volatility forecasts, but rather in producing volatility timing rules which maximize the mean-variance trade-off, we do not primarily evaluate the models' performance based on the accuracy of the forecasts. Instead, we evaluate the models based on their ability to produce managed portfolios with the highest Sharpe ratio. For each model specification, we produce out-of-sample forecasts of the next month's realized variance using the rolling window. We then use these forecasts to form volatility-managed portfolios as described in Equation (5). Transaction costs for implementing the managed portfolio are also calculated and net-of-cost returns are used to calculate the Sharpe ratio of the managed portfolio.

For each of the six factors, we take the net-of-costs return of the managed portfolios as in Equation (10) and calculate the annualized net-of-costs Sharpe ratio as in Equation (12). The model specification with the highest Sharpe ratio per factor is then selected as the optimal model for that model class / factor combination. The only exception being the random forest model, since it is computationally intensive to evaluate it for each of the factors, we only evaluate it for the market factor and fix the hyperparameters for the other factors. The optimal model specifications for each model class are then used to form the managed portfolios for the evaluation of the volatility timing rules.

5.2 Model specifications

5.2.1 GARCH(1,1) model

Traditionally, a popular model for modelling volatility has been the GARCH(1,1) model. It assumes that volatility can be modelled as an ARMA(1,1) model and that the conditional variance is a linear function of previous squared returns and previous conditional variance. The return process is assumed to be:

$$r_t = \mu + \epsilon_t, \text{ with} \tag{13}$$

$$\epsilon_t = \sigma_t z_t, \tag{14}$$

where μ is the mean return, ϵ_t is the return shock, σ_t is the conditional standard deviation and z_t is a standard normal random variable. The conditional variance is modelled as:

$$\sigma_t^2 = c + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{15}$$

The GARCH(1,1) model makes several assumptions about the return process and the conditional variance. It assumes that the return shocks are conditionally normally distributed with mean zero and time-varying variance. The model also assumes that the conditional variance is a positive function of the model parameters and the past squared return shocks and conditional variances. For the GARCH(1,1) process to guarantee $\sigma_t^2 \geq 0$ and be stationary, the following conditions must hold $c > 0$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$. The GARCH(1,1) model parameters are estimated using Maximum Likelihood Estimation (MLE).

We implement two GARCH(1,1) models, one for daily variance and one for monthly variance. Monthly variance forecasts are then generated directly from the monthly model and by scaling the daily one-step ahead forecast to a monthly forecast by multiplication by 22. Which version of the GARCH(1,1) model is used for forming managed portfolios is determined using the model selection procedure as outlined in Section 5.1.

One caveat of the GARCH(1,1) model is that it does not accommodate long memory in variances. The model assumes that the impact of past return shocks on the current conditional variance decays exponentially over time, with the rate of decay determined by the sum of α and β. This means that the model cannot capture the slow decay of the autocorrelation function observed in realized volatility time series (Andersen et al., 2003; Corsi, 2009), which signifies a long memory processes. The empirical autocorrelation function of realized volatility is shown to decay at a hyperbolic rate (Andersen et al., 2003), which is slower than the exponential decay implied by the GARCH(1,1) model.

5.2.2 HAR-RV model

Financial return data is known to exhibit long memory and is shown to exhibit evidence of scaling and multiscaling behaviour (Corsi, 2009). Scaling refers to the phenomena where the return distribution has a similar shape across multiple time scales and multiscaling refers to the phenomena where the distribution exhibits different scaling behaviour for different moments. Traditional volatility models such as the GARCH model are not able to capture these characteristics and appear as white noise when aggregated over longer time scales due to their short memory, thus exhibiting no scaling behaviour.

To better accommodate for these characteristics Corsi (2009) propose the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV model). This model can capture the scaling behaviour of financial returns and is shown to outperform traditional volatility models in forecasting realized volatility (Y. Wang et al., 2016).

The HAR-RV model assumes that the realized variance can be modelled as a function of previously realized variance, realized variance over the past week, and realized variance over the past month. The model is specified as:

$$RV_{t+1}^{(\mathrm{d})} = c + \beta^{(\mathrm{d})} RV_t^{(\mathrm{d})} + \beta^{(w)} RV_t^{(w)} + \beta^{(\mathrm{m})} RV_t^{(\mathrm{m})} + \epsilon_{t+1}, \text{ with} \tag{16}$$

$$RV_t^{(w)} = \frac{1}{5}\sum_{i=0}^{4} RV_{t-i}^{(\mathrm{d})}, \text{ and} \tag{17}$$

$$RV_t^{(\mathrm{m})} = \frac{1}{22}\sum_{i=0}^{21} RV_{t-i}^{(\mathrm{d})}, \tag{18}$$

where $RV_t^{(\mathrm{d})}$ is the realized variance for day t. The parameters $\beta^{(\mathrm{d})}$, $\beta^{(w)}$ and $\beta^{(\mathrm{m})}$ are estimated using OLS regression. Estimates for the one-step ahead daily realized variance are scaled by 22 to obtain monthly realized variance forecasts.

We also propose a variant of the HAR-RV model which models monthly realized variance directly and may be able to better accommodate the long memory properties of realized variance. This model is specified as:

$$RV_{\tau+1}^{(\mathrm{m})} = c + \beta^{(\mathrm{m})} RV_\tau^{(\mathrm{m})} + \beta^{(q)} RV_\tau^{(q)} + \beta^{(y)} RV_\tau^{(y)} + \epsilon_{\tau+1}, \text{ with} \tag{19}$$

$$RV_\tau^{(q)} = \frac{1}{4}\sum_{i=0}^{4} RV_{\tau-i}^{(\mathrm{m})}, \text{ and} \tag{20}$$

$$RV_\tau^{(y)} = \frac{1}{12}\sum_{i=0}^{12} RV_{\tau-i}^{(\mathrm{m})}, \tag{21}$$

where $RV_\tau^{(\mathrm{m})}$ is the realized variance for month τ. The parameters $\beta^{(\mathrm{m})}$, $\beta^{(q)}$ and $\beta^{(y)}$ are estimated using OLS regression.

Which version of the HAR-RV model is used for forming managed portfolios is determined using the model selection procedure as outlined in Section 5.1.

5.2.3 ARFIMA model

Another model that allows for long memory processes is the Fractionally Integrated ARMA (ARFIMA) model (introduced by Baillie (1996)). It is an extension of the traditional ARMA model. The model incorporates a fractional differencing parameter, denoted as d, which captures the long-term dependence structure of the time series. The ARFIMA(p, d, q) model can be expressed as:

$$(1 - L)^d \Phi(L)(X_t - \mu) = \Theta(L)\varepsilon_t, \tag{22}$$

where X_t is the variable of interest, L is the lag operator, such that $L^k X_t = X_{t-k}$ and d is the fractional differencing parameter, which can take non-integer values. The autoregressive (AR) and moving average (MA) polynomials are defined as: $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$ and $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$ respectively for orders p and q. μ is the mean of the process and ε_t is a white noise process with zero mean and constant variance.

The fractional differencing operator $(1 - L)^d$ can be expanded as follows:

$$(1 - L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k - d)}{\Gamma(k + 1)\Gamma(-d)} (L)^k, \tag{23}$$

where $\Gamma(\cdot)$ is the gamma function.

For an ARFIMA process to be stationary, the fractional differencing parameter d must satisfy $|d| < 0.5$ (Konstantinidi et al., 2008). The interpretation of the fractional integration parameter d is as follows:

- When $d = 0$, the process reduces to a standard ARMA model, indicating short memory.
- When $0 < d < 0.5$, the process exhibits long memory and the autocorrelations decay hyperbolically.
- When $-0.5 < d < 0$, the process is anti-persistent, meaning that positive values are more likely to be followed by negative values, and vice versa. The autocorrelations alternate in sign and decay hyperbolically.

The fractional differencing parameter d captures the long-term dependence structure of the time series. It allows the ARFIMA model to describe processes that exhibit long memory, where the impact of shocks persists over a long period. The ACF of an ARFIMA process, denoted as $\rho(k)$ for lag k, can be approximated as (Brailsford & Faff, 1996):

$$\rho(k) \sim ck^{2d-1} \text{ as } k \to \infty, \tag{24}$$

where c is a constant.

This approximation shows that the ACF decays at a rate proportional to k^{2d-1}. When $0 < d < 0.5$, the exponent $2d - 1$ lies between -1 and 0, resulting in a hyperbolic decay. As previously mentioned, the autocorrelation of realized variance is shown to exhibit hyperbolic decay (Andersen et al., 2003). Therefore, the ARFIMA model is expected to better capture this aspect of the data compared to the more traditional GARCH model.

Following Chen et al. (2019) we implement an ARFIMA(1,d,1) model for monthly realized variance. We also implement an ARFIMA(1,d,1) model for daily realized variance, where the one period ahead forecast is transformed to a monthly forecast by multiplication by 22. The parameters of the ARFIMA model, except d, are estimated using Maximum Likelihood Estimation (MLE). Due to the computational complexity of including the fractional differencing parameter d in the MLE, we opt to treat it as a hyperparameter of the model. We form a grid of possible values for d ranging from 0.05 to 0.45 with a step size of 0.05. The procedure for selecting the optimal value of d and the choice between modelling daily or monthly realized variance is outlined in Section 5.1.

5.2.4 MIDAS model

For the purposes of forming volatility-managed portfolios, we are interested in multiperiod forecasts of realized variance. Specifically, we need to forecast the realized variance for the next month. In the literature there exists mostly models which directly forecast volatility at a certain frequency or those which model volatility at a daily frequency. In the latter case a multi-period forecast can then be obtained by iteratively forecasting multiple one-step ahead forecasts, or by scaling the one-period ahead forecast up to a longer period.

A third method is to apply Mixed-Data Sampling (MIDAS) (proposed by Ghysels et al., 2005, 2006). This method aims to find a middle ground between these former approaches by directly producing multi-period forecasts using daily returns.

The MIDAS specification allows for the use of high-frequency data to forecast low-frequency data. The model is based on the idea that the relationship between high-frequency and low-frequency data can be captured by a lag polynomial with a certain functional form. The model is specified as follows:

$$V_{t+1}^{k} = \mu_k + \phi_k \sum_{j=0}^{j_{\max}} b_k(j,\theta) r_{t-j/k}^2 + \varepsilon_{k,t}, \tag{25}$$

where V_{t+1}^{k} is a measure of the k-period realized variance at time $t+1$, μ_k is the intercept term, ϕ_k is a scaling parameter. The polynomial lag parameters $b_k(j,\theta)$ are parametrized to be a function of θ. The estimation of μ_k, ϕ_k and θ is done using QMLE. We take $k = 22$ to forecast monthly realized variance and vary $j_{\max} \in [20, 25, \ldots, 55, 60]$. This results in the following model specification for the realized variance forecast for month $\tau+1$, where we set t to be the last day of month τ:

$$RV_{\tau+1}^{(\mathrm{m})} = \mu + \phi \sum_{j=0}^{j_{\max}} b(j,\theta) r_{t-j}^2 + \varepsilon_{\tau+1}. \tag{26}$$

Ghysels et al. (2009) propose several functional forms for the lag polynomial $b(j,\theta)$ to model the interactions between the daily returns and the monthly realized variance. The functional forms are chosen to ensure that the weights are positive and sum to one. The functional forms are as follows[4]:

1. Exponential:
$$b_k(j, \theta_1, \theta_2) = \frac{\exp\{\theta_1 j + \theta_2 j^2\}}{\sum_{i=1}^{j_{\max}} \exp\{\theta_1 i + \theta_2 i^2\}}. \tag{27}$$
This functional form guarantees positive weights for the lags and makes sure that the weights sum to one. Positive weights are necessary to guarantee that the forecasted variance is also positive. When $\theta_2 < 0$, weights decay with lag length.

2. Beta:
$$b_k(j, \theta_1, \theta_2) = \frac{f(j/j_{\max}, \theta_1, \theta_2)}{\sum_{i=1}^{j_{\max}} f(i/j_{\max}, \theta_1, \theta_2)}, \tag{28}$$
where: $f(z, a, b) = z^{a-1}(1-z)^{b-1}/B(a, b)$ and $B(a, b)$ is based on the Gamma function, or $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. The Gamma function is defined as: $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\, dt$, for $\mathbb{R}(z) > 0$.
For $\theta_1 = 1$ and $\theta_2 > 1$ this specification has a slowly decaying pattern typical of volatility processes.

Optimal choices for the functional form of the lag parameters as well as $j_{\max}$ need to be found. This is done using the procedure described in Section 5.1, where the best performing combination is chosen for each factor.

5.2.5 Random Forest

Using parametric models, some restriction in the functional form of the model is imposed. This can limit the model's ability to capture complex relationships in the data which could not be captured by parametric models. To relax this limitation, a non-parametric model can instead be used. Random Forest (RF) (introduced by Ho (1995)) is a popular non-parametric model that can be used for regression tasks. It has the potential to capture complex, non-linear relationships in the data.

RF is an ensemble learning method that constructs a range of decision trees during training and outputs the mean prediction of the individual trees as the estimate. The key idea is to combine many decision trees in order to reduce the risk of overfitting.

Given p available covariates, the algorithm for random forests is as follows:

1. For $b \in (1, \ldots, B)$ form a regression tree T_b as follows:
   (a) Draw a bootstrap sample X^* of size $N_{\mathcal{B}}$ from the training data using i.i.d. bootstrapping[5]. Where $N_{\mathcal{B}} = \max(1, \text{round}(N * n_{\mathcal{B}}))$ and $n_{\mathcal{B}}$ is a hyperparameter of the model and can take values between 0 and 1.

[4]Ghysels et al. (2009) also propose the step function, which we do not consider in this thesis. The reason for this being the similarity with the HAR-RV model (The HAR-RV model could be specified as a special case of the MIDAS model with a step function as the lag polynomial). Additionally, the exact implementation of the step function is not clearly defined in the literature and choices need to be made regarding the number and locations of the steps. This would add a complicated extra dimension to the estimation of this model specification.

[5]Instead of i.i.d. bootstrapping, block bootstrapping could also be appropriate due to the serial correlation of variance. However, Christensen et al. (2023) did not find a discernable difference in forecasting performance between block bootstrapping and i.i.d. bootstrapping. Therefore, we implement i.i.d. bootstrapping to reduce implementation complexity.

(b) Grow a regression tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size $n_{\min}$ is reached where $n_{\min}$ is a hyperparameter of the model.

    i. Select M covariates at random from the p available covariates. Where $M = \max(1, \lfloor m * p \rfloor)$ and m is a hyperparameter of the model and can take values between 0 and 1.

    ii. Pick the best variable/split-point among the m by minimizing sum of squared errors.

    iii. Split the node into two daughter nodes.

2. Output the ensemble of trees $\{T_1, \ldots, T_B\}$.

To make a prediction at a new point x:

$$\hat{f}_{RF}(X) = \frac{1}{B} \sum_{b=1}^{B} T_b(X), \tag{29}$$

where $(B, n_{\mathcal{B}}, m, n_{\min})$ are hyperparameters of the model and need to be tuned. B is the number of trees in the forest, $n_{\mathcal{B}}$ is the percentage of samples in the bootstrap sample, m is the percentage of variables to consider at each split, and $n_{\min}$ is the minimum number of samples required to split a node.

The individual trees are assumed to be i.i.d. due to the bootstrapping of the data in the construction of each tree. RF is also robust to overfitting, as the ensemble of trees helps to reduce the variance of the model. Additionally, RF can handle high-dimensional data and is not sensitive to outliers or missing values.

The target variable of the RF model is set to the monthly realized variance. As covariates we set the realized variance of the previous 22 days as well as the previous weekly realized variance and previous monthly realized variance. Additionally, the ADS index, 3-month T-bill rate and the Sentiment Index are included as regressors in the model. $(B, N_{\mathcal{B}}, m, n_{\min})$ are hyperparameters that need to be tuned. This is done using the procedure as described in Section 5.1, where the hyperparameters are selected using Tree-structured Parzen Estimation (TPE) (Bergstra et al., 2011)[6].

TPE is a Bayesian optimization method that suggests the next set of hyperparameters to evaluate. The tree-structured part refers to the fact that the model is able to handle tree-structured hyperparameters, however, this aspect of the model is not used in this case. For each parameter, a series of random draws are made, and the objective function is evaluated for each of these draws. The sets of parameter draws are then divided into two sets based on their objective values, split around the 10% quantile. For both sets of draws a Parzen estimator is used to estimate the marginal probability distributions of the parameters, resulting in $l(x)$ and $g(x)$, where $l(x)$ is the probability distribution of the parameter set which is below the 10%

[6]Any other hyperparameter optimization procedure would be applicable as well, such as performing a grid search. TPE was chosen out of convenience since it is available for use through the *Optuna* python package. Since the focus of this thesis does not lie in finding the best hyperparameter optimization method, we will not go further into hyperparameter optimization intricacies.

quantile of the objective values and $g(x)$ is the probability density of the parameter set above the 10% quantile of the objective values. The optimal next set of parameters to evaluate is then chosen by maximizing $l(x)/g(x)$. The proof for this is found in Bergstra et al. (2011).

The hyperparameter ranges for the Random Forest model are as follows:

- $B : \{B \in \mathbb{Z} \mid 50 \leq B \leq 500\}$.
- $n_{\mathcal{B}} : \{n_{\mathcal{B}} \in \mathbb{R} \mid 0.05 \leq n_{\mathcal{B}} \leq 1.00\}$.
- $m : \{m \in \mathbb{R} \mid 0.05 \leq m \leq 1.00\}$.
- $n_{\min} : \{n_{\min} \in \mathbb{Z} \mid 1 \leq n_{\min} \leq 50\}$.

5.3 Model combinations

Since different models make use of different ranges of data and make different assumptions about relationships in the data, selecting a single 'best' model exposes the analysis to data uncertainty and model uncertainty (X. Wang et al., 2022). One way to combat this is to introduce combiner models to combine the forecasts of multiple models into a single combined forecast. By combining the forecasts of different models, combiner models can reduce the risk of overfitting and capture different aspects of the data. There is a large range of combiner models available in the literature, but we will restrict ourselves to two combiner models: Mean Forecast Combination (MFC) and Stacking.

5.3.1 Mean Forecast Combination (MFC)

A popular combiner model is Mean Forecast Combination (MFC). It is a simple forecast combination technique that combines the forecasts of multiple models by taking the equally weighted average of the individual forecasts. One of the advantages of this approach are its ease of implementation since no parameter estimation is needed. Additionally, this approach can reduce variance and bias through averaging out individual model bias (Palm & Zellner, 1992), thus increasing out-of-sample performance.

The MFC forecast is calculated as:

$$\hat{y}_t = \frac{1}{N_\mathrm{m}} \sum_{i=1}^{N_\mathrm{m}} \hat{y}_{t,i}, \tag{30}$$

where $\hat{y}_t$ is the combined forecast, $\hat{y}_{t,i}$ is the forecast of model i, and N_m is the number of models.

MFC is easy to implement since the models can be trained individually, and their forecasts combined by taking the average. No hyperparameter tuning or model selection is needed for MFC.

5.3.2 Stacking

Instead of taking the simple average of the individual forecasts, one can give more weight to the forecasts of the better performing models. One way to determine the weights is by using the Stacking (Wolpert, 1992) model. It is a slightly more advanced forecast combination technique that combines the forecasts of multiple models using different weights for each model. We restrict the combiner model to be a linear combination of the individual forecasts. In this form, Stacking (like MFC) can reduce the risk of overfitting and reduce the variance and bias of the models (Granger & Ramanathan, 1984), with the additional relaxation of the constraint of equal weights. Thus Stacking adds a layer of flexibility to the forecast combination process when compared to MFC.

The stacking procedure defines so called level-one data, which is obtained by training the individual models on the training set and using the out-of-sample forecasts as samples in the stacking models. The stacking models then combine the forecasts of multiple models by fitting a combiner model on the level-one data.

For N_{m} individual models, we define the combiner model as:

$$\hat{y}_t = \sum_{i=1}^{N_{\mathrm{m}}} \beta_i \hat{y}_{t,i}, \tag{31}$$

where $\hat{y}_t$ is the combined forecast, $\hat{y}_{t,i}$ is the forecast of model i, N_{m} is the number of models and β_i are taken to minimize by means of OLS:

$$\min_{\beta_i} \sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{N_{\mathrm{m}}} \beta_i \hat{y}_{t,i} \right)^2, \tag{32}$$

where the constraint $\beta_i \geq 0$ is imposed to ensure that the forecasted variance is always positive. Additionally, this restriction reduces the risk of overfitting (Breiman, 1996).

In addition to using ordinary least squares (OLS) to estimate the combiner coefficients β_i, regularization techniques such as Lasso, Ridge, and Elastic Net can be employed to potentially improve the stacked regression model's performance and reduce risk of overfitting.

Lasso regularization (Tibshirani, 1996) adds an ℓ_1 penalty term to the optimization objective, which encourages sparsity in the coefficients:

$$\min_{\beta_i} \sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{N_{\mathrm{m}}} \beta_i \hat{y}_{t,i} \right)^2 + \lambda \sum_{i=1}^{N_{\mathrm{m}}} |\beta_i|. \tag{33}$$

Ridge regularization (Hoerl & Kennard, 1970) incorporates an ℓ_2 penalty term, which shrinks the coefficients towards zero:

$$\min_{\beta_i} \sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{N_{\mathrm{m}}} \beta_i \hat{y}_{t,i} \right)^2 + \lambda \sum_{i=1}^{N_{\mathrm{m}}} \beta_i^2. \tag{34}$$

Elastic Net regularization (Zou & Hastie, 2005) combines both ℓ_1 and ℓ_2 penalties, offering a balance between Lasso and Ridge:

$$\min_{\beta_i} \sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{N_{\mathrm{m}}} \beta_i \hat{y}_{t,i} \right)^2 + (\alpha)\lambda \sum_{i=1}^{N_{\mathrm{m}}} |\beta_i| + (1-\alpha)\lambda \sum_{i=1}^{N_{\mathrm{m}}} \beta_i^2, \tag{35}$$

where $0 < \alpha < 1$.

In these regularization methods, λ and α are tuning parameters that control the strength of the regularization. Lasso has the benefit of performing variable selection by setting some coefficients to exactly zero, which can be useful when dealing with many individual models. Elastic Net strikes a balance between Lasso and Ridge, offering both sparsity and coefficient shrinkage.

λ and α are hyperparameters that need to be tuned. This is done using the procedure as described in Section 5.1, where the hyperparameters are selected using Tree-structured Parzen Estimation (TPE) (Bergstra et al., 2011). The predictors are scaled up by 1×10^4. This is done to ensure that the coefficients of the combiner model are not too small, which can lead to convergence issues. The predictions are then scaled back down by 1×10^{-4} to obtain the final forecast. The hyperparameter ranges are as follows:

- $\lambda : \{\lambda \in \mathbb{R} \mid 0 \leq \lambda \leq 1 \times 10^3\}$.
- $\alpha : \{\alpha \in \mathbb{R} \mid 0 \leq \alpha \leq 1\}$.

6 Portfolio performance evaluation

After having selected the optimal model for each model class, we evaluate the performance of the volatility-managed portfolios whose scaling is a function of forecasted realized variance. Using the optimal models for each model class, we produce out-of-sample forecasts of realized variance over the period 1976-01 to 2022-06 using a rolling window of 5 years. We compare the performance of the managed portfolios produced using forecasted realized variance to the performance of the managed portfolios produced using past realized variance. We also consider a range of transaction cost mitigation techniques to produce a more complete comparison of the performance of using forecasted realized variance over past realized variance.

All the metrics covered are calculated for each unmanaged factor and each of the managed portfolios over the period ranging from 1976-01 to 2022-06. This is to ensure that the sample period is equal for all unmanaged and managed portfolios and that the performance metrics are comparable. This is in contrast to the performance evaluation of Section 4.4 where the sample period was the full sample period 1966-01 to 2022-06.

6.1 Evaluating portfolio performance

To compare portfolio performance, we again look at the net-of-costs return of the managed portfolios. The net-of-costs return of the managed portfolios is calculated using Equation (10).

We then calculate the net-of-costs certainty equivalent rate (CER) and net-of-costs Sharpe ratio (SR) of the managed portfolios. Where Sharpe ratios are calculated using Equation (12) and the CER is calculated as:

$$CER = \mathbb{E}\left[R_{\tau}^{\sigma *}(f)\right] - \frac{1}{2}\lambda \cdot \mathbb{V}\left[R_{\tau}^{\sigma}(f)\right], \tag{36}$$

where λ is the risk aversion parameter and is set $\lambda = 3$ following Barroso and Detzel (2021).

Additionally, Mean Squared Forecast Errors (MSFE) of the realized variance models are presented in Appendix C. These are not used in the evaluation of the managed portfolios but are presented for reference.

6.2 Transaction cost mitigation

To produce a more complete comparison of the performance of using forecasted realized variance over past realized variance, we also consider a range of transaction cost mitigation techniques. We then compare the performance of the managed portfolios using forecasted RV to the managed portfolios using historic RV with these transaction cost mitigation techniques.

Barroso and Detzel (2021) find that scaling the factors by the prior month's variance increases monthly turnover by as much as 15 times relative to the unmanaged factors. To mitigate this sharp increase in transaction costs we look at additional ways to construct volatility-managed portfolios that reduce turnover. Three cost mitigation techniques are introduced by Moreira and Muir (2017) in their paper. The first technique is to scale the factors by the realized volatility of the factors instead of the realized variance. The second technique is to scale by expected variance instead of realized variance. Since scaling by forecasted realized variance is equivalent to scaling by expected variance, this will not be considered as a cost-mitigation technique. The third technique is to cap leverage at 1.5.

A fourth technique which is mentioned by Barroso and Detzel (2021) and follows the methodology of Barroso and Santa-Clara (2015) is to scale by realized volatility which is estimated over the prior six-month period. Barroso and Detzel (2021) also consider a fifth cost-mitigation technique by Novy-Marx and Velikov (2016) in which they exclude small-cap stocks from the formation of the factor portfolios. They argue that small-cap stocks are associated with higher transaction costs and that excluding them from the factor portfolios can reduce turnover.

As an additional strategy we will also evaluate managed portfolios conditional on the sentiment index. As found by Barroso and Detzel (2021), the Sharpe ratio of managed portfolios doubled after high realizations of the index. Conversely, they were reduced after low realizations of the same index. This suggests that the sentiment index can be used as a timing signal for the managed portfolios. We will construct managed portfolios which practice volatility timing during high realizations of the sentiment index (above the median value) and which do not practice volatility timing during low realizations of the sentiment index (below the median value).

This results in the volatility scaling technique used by Moreira and Muir (2017), five cost-mitigation techniques and timing conditional on the Sentiment index to consider:

1. Scaling by realized variance.
2. Scaling by realized volatility[7].
3. Scaling by realized volatility over the prior six-month period.
4. Capping leverage at 1.5 (weight can not exceed 1.5).
5. Excluding small-cap stocks from factor portfolios.
6. Conditional on Sentiment index.

The last three cost-mitigation techniques require some additional explanation.

- By capping leverage at 1.5, what is meant is that the scaling of the factor returns as a result of the volatility timing can not exceed 1.5. That is, when the managed factor is constructed as follows:

$$R_\tau(f^\sigma) = \frac{c}{\sigma^2_{\tau-1}(f)} R_\tau(f) = L_\tau(f) R_\tau(f), \tag{37}$$

  where we impose the constraint that $L_\tau(f) \leq 1.5$. This is implemented by setting $L_\tau(f) = \min(1.5, L_\tau(f))$.

[7]Something to note for the construction of the managed portfolios scaled by realized volatility is that the square root of the estimated realized variance is used as an estimate of the realized volatility. This introduces bias in the estimation of the realized volatility. This is due to the well-known Jensen's inequality, which states that $\mathbb{E}[X^2]$ is generally not equal to $\mathbb{E}[X]^2$. Molnár (2012) discussed this issue in detail and shows for a multitude of realized volatility estimators that this bias is equal to some constant specific to the used estimator. Since introducing a constant in the scaling factor of Equation (3) does not affect the Sharpe ratio, this bias is of no concern for the evaluation of the portfolio performances.

- Excluding small-cap stocks from the factor portfolios is done by constructing the factor portfolios using only stocks which have a market capitalization above the NYSE median. The factor portfolios are then constructed using these stocks. The realized variances of the original factors are still used as the scaling factor for these managed portfolios.

- When constructing the managed portfolios conditional on the sentiment index, the sentiment index is used as a timing signal. When the sentiment index is above its median value in an expanding window, the managed portfolios are constructed using the volatility timing strategy. Conversely, if the sentiment index is below the median value, the leverage term of the managed portfolio is set equal to the leverage of the preceding month:

$$L_\tau(f) = \begin{cases} \frac{c}{\sigma^2_{\tau-1}(f)} & SENT_{\tau-1} > \text{median}(SENT_1, \dots, SENT_{\tau-1}) \\ L_{\tau-1}(f) & SENT_{\tau-1} \leq \text{median}(SENT_1, \dots, SENT_{\tau-1}) \end{cases}, \tag{38}$$

  where $SENT_\tau$ is the value of the sentiment index in month τ.

6.3 Inference on portfolio performance

6.3.1 Sharpe ratio difference test

In order to assess the significance of Sharpe ratio improvements, we perform a statistical significance test on the difference between two Sharpe ratios. We follow the same procedure as Ledoit and Wolf (2008). Testing the null hypothesis that the difference in Sharpe ratios is zero. Given two investment strategies i and j whose excess returns at time t are $r_{i,t}$ and $r_{j,t}$. It is assumed that these returns constitute a strictly stationary time series. We test $H_0 : \Delta_{\text{SR}} = 0$ against $H_1 : \Delta_{\text{SR}} \neq 0$, with a significance level of $\alpha = 0.05$.

Let $\mu_i = \mathbb{E}(r_{it})$, $\mu_j = \mathbb{E}(r_{jt})$, $\gamma_i = \mathbb{E}(r_{it}^2)$ and $\gamma_j = \mathbb{E}(r_{nt}^2)$. Furthermore, let $v = (\mu_i, \mu_j, \gamma_i, \gamma_j)'$ and $\hat{v} = (\hat{\mu}_i, \hat{\mu}_j, \hat{\gamma}_i, \hat{\gamma}_n)'$ be the vector of the first two moments of both return series and its estimate respectively. Then the difference between Sharpe ratios can be written as:

$$\Delta_{\text{SR}} = f(v) \quad \text{and} \quad \hat{\Delta}_{\text{SR}} = f(\hat{v}), \tag{39}$$

with

$$f(v) = f(a, b, c, d) = \frac{a}{\sqrt{c - a^2}} - \frac{b}{\sqrt{d - b^2}}. \tag{40}$$

Assuming that $\sqrt{T}(\hat{v} - v) \xrightarrow{d} N(0, \Psi)$, the delta method then implies:

$$\sqrt{T}(\hat{\Delta}_{\text{SR}} - \Delta_{\text{SR}}) \xrightarrow{d} N(0, \nabla' f(v) \Psi \nabla f(v)), \tag{41}$$

with

$$\nabla' f(v) = \left(\frac{c}{(c - a^2)^{1.5}}, -\frac{d}{(d - b^2)^{1.5}}, -\frac{1}{2}\frac{a}{(c - a^2)^{1.5}}, \frac{1}{2}\frac{b}{(d - b^2)^{1.5}} \right). \tag{42}$$

With a consistent estimator $\hat{\Psi}$ of Ψ we can estimate the standard error of $\hat{\Delta}_{\text{SR}}$ as:

$$s(\hat{\Delta}_{\mathrm{SR}}) = \sqrt{\frac{\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v})}{T}}, \tag{43}$$

where $\hat{\Psi}$ is obtained via Parzen kernel estimation.

Inference based on asymptotic normality has been shown to provide less accurate results when compared to the bootstrap method (Ledoit & Wolf, 2008). Therefore, we apply circular block bootstrapping[8] (Politis & Romano, 1991) to improve inference accuracy.

Using circular block bootstrapping we generate $M = 5000$ bootstrapped time series realizations, where we set the window size of the circular block bootstrap to 12 months[9]. We can estimate the standard error of the mth bootstrap realization $\hat{\Delta}^{*,m}_{\mathrm{SR}}$ as follows:

$$s(\hat{\Delta}^{*,m}_{\mathrm{SR}}) = \sqrt{\frac{\nabla' f(\hat{v}^{*,m})\hat{\Psi}^{*,m}\nabla f(\hat{v}^{*,m})}{T}}, \tag{44}$$

where $\hat{v}^{*,m} = (\hat{\mu}^{*,m}_i, \hat{\mu}^{*,m}_j, \hat{\gamma}^{*,m}_i, \hat{\gamma}^{*,m}_n)'$ denotes the moment estimates of the mth bootstrap realization and $\hat{\Psi}^{*,m}$ is the regular sample covariance matrix of the mth bootstrap realization.

We are interested in the two-sided p-value of the null hypothesis of equal Sharpe ratios. Therefore, we calculate the p-value in the following way. Denote the 'original' test statistic by d as follows:

$$d_{\mathrm{SR}} = \frac{|\hat{\Delta}_{\mathrm{SR}}|}{s(\hat{\Delta}_{\mathrm{SR}})} \tag{45}$$

and let the centred bootstrap statistic from the mth bootstrap realization be denoted by $d^{*,m}$ for $m = 1, \ldots, M$:

$$d^{*,m}_{\mathrm{SR}} = \frac{|\hat{\Delta}^{*,m}_{\mathrm{SR}} - \hat{\Delta}_{\mathrm{SR}}|}{s(\hat{\Delta}^{*,m}_{\mathrm{SR}})}. \tag{46}$$

Then the two-sided p-value for the null of equal Sharpe ratios is given by (Ledoit & Wolf, 2008):

$$p = \frac{\#\{d^{*,m}_{\mathrm{SR}} \geq d_{\mathrm{SR}}\} + 1}{M + 1}. \tag{47}$$

6.3.2 Certainty equivalent rate difference test

Similarly to the Sharpe ratio difference test, we can perform a statistical significance test on the difference in certainty equivalent rates. We follow the same procedure as Ledoit and Wolf (2008), where we test the null hypothesis of equal certainty equivalent rates with a significance level of $\alpha = 0.05$.

Again, let $\mu_i = \mathbb{E}(r_{it})$, $\mu_j = \mathbb{E}(r_{jt})$, $\gamma_i = \mathbb{E}(r^2_{it})$, $\gamma_j = \mathbb{E}(r^2_{nt})$, $v = (\mu_i, \mu_j, \gamma_i, \gamma_j)'$ and $\hat{v} = (\hat{\mu}_i, \hat{\mu}_j, \hat{\gamma}_i, \hat{\gamma}_j)'$ be the vector of the first two moments of both return series and its estimate

[8] The full sample $(X_1, \ldots, X_n)$ is wrapped around in a 'circle'. Meaning that for $i > n$; $X_i \equiv X_{i_n}$, with $i_n = i \bmod n$. The wrapped sample is then divided into n overlapping blocks B_i of length b, where $B_i = (X_i, \ldots, X_{i+b-1})$. k blocks are then sampled with replacement from $(B_1, \ldots, B_n)$, where $k \sim n/b$ and stitched together in the order in which they were sampled.

[9] As a robustness check the same procedure is performed with a window size set to 6 months. These results are presented in Appendix B but show no notable differences in outcomes.

respectively. Then the difference between certainty equivalent rates can be written as:

$$\Delta_{\text{CER}} = f(v) \quad \text{and} \quad \hat{\Delta}_{\text{CER}} = f(\hat{v}), \tag{48}$$

with

$$f(v) = f(a,b,c,d) = a - \frac{1}{2}\lambda(c - a^2) - b + \frac{1}{2}\lambda(d - b^2). \tag{49}$$

Similarly, assuming that $\sqrt{T}(\hat{v} - v) \xrightarrow{d} N(0, \Psi)$, the delta method then implies:

$$\sqrt{T}(\hat{\Delta}_{\text{CER}} - \Delta_{\text{CER}}) \xrightarrow{d} N(0, \nabla' f(v)\Psi\nabla f(v)), \tag{50}$$

with

$$\nabla' f(v) = \left(1 + \lambda a, -(1 + \lambda b), -\frac{1}{2}\lambda, \frac{1}{2}\lambda\right). \tag{51}$$

With a consistent estimator $\hat{\Psi}$ of Ψ we can estimate the standard error of $\hat{\Delta}_{\text{CER}}$ as:

$$s(\hat{\Delta}_{\text{CER}}) = \sqrt{\frac{\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v})}{T}}, \tag{52}$$

where $\hat{\Psi}$ is obtained via the Parzen kernel estimator.

For the same reasons as with the test on the Sharpe ratio difference, we apply circular block bootstrapping. Using circular block bootstrapping we generate $M = 5000$ bootstrapped time series realizations, where we set the window size of the circular block bootstrap to 12 months[10]. We can estimate the standard error of the mth bootstrap realization $\hat{\Delta}^{*,m}_{\text{CER}}$ as follows:

$$s(\hat{\Delta}^{*,m}_{\text{CER}}) = \sqrt{\frac{\nabla' f(\hat{v}^{*,m})\hat{\Psi}^{*,m}\nabla f(\hat{v}^{*,m})}{T}}, \tag{53}$$

where $\hat{v}^{*,m} = (\hat{\mu}_i^{*,m}, \hat{\mu}_j^{*,m}, \hat{\gamma}_i^{*,m}, \hat{\gamma}_n^{*,m})'$ denotes the moment estimates of the mth bootstrap realization and $\hat{\Psi}^{*,m}$ is the regular sample covariance matrix of the mth bootstrap realization.

Again, we are interested in the two-sided p-value of the null hypothesis of equal Sharpe ratios. Therefore, we calculate the p-value in the following way. Denote the 'original' test statistic by d as follows:

$$d_{\text{CER}} = \frac{|\hat{\Delta}_{\text{CER}}|}{s(\hat{\Delta}_{\text{CER}})} \tag{54}$$

and let the centred bootstrap statistic from the mth bootstrap realization be denoted by $d^{*,m}$ for $m = 1, \ldots, M$:

$$d^{*,m}_{\text{CER}} = \frac{|\hat{\Delta}^{*,m}_{\text{CER}} - \hat{\Delta}_{\text{CER}}|}{s(\hat{\Delta}^{*,m}_{\text{CER}})}. \tag{55}$$

Then the two-sided p-value for the null of equal certainty equivalent rates is given by (Ledoit & Wolf, 2008):

$$p = \frac{\#\{d^{*,m}_{\text{CER}} \geq d_{\text{CER}}\} + 1}{M + 1}. \tag{56}$$

[10] As a robustness check the same procedure is performed with a window size set to 6 months. These results are presented in Appendix B but show no notable differences in outcomes.

7 Results

7.1 Performance evaluation

Historic volatility measures

We first look at the performance of the managed portfolios using past realized variance and the cost-mitigation techniques. Based on the different timing rules, we form managed portfolios for each factor. We calculate the net-of-costs Sharpe ratio and net-of-costs certainty equivalent rate for each managed portfolio. We also test for equality of the Sharpe ratios and certainty equivalent rates of the managed portfolios with the unmanaged portfolios following the procedure as outlined in Section 6.3. The resulting performance measures and inference results are presented in Table 6. We also present the transaction costs and turnover for the managed portfolios based on the past realized variance in Figure 3.

From Figure 3 we can see that the cost mitigation techniques do have an effect on the average turnover and transaction costs. However, in Table 6 we see that only for the momentum factor does a timing rule produce a higher Sharpe ratio compared to its unmanaged counterpart. For all the other factors, employing one of the timing rules based on past realized variance leads to a decrease in the Sharpe ratio. For the momentum factor, the highest increase in Sharpe ratio is observed when scaling by the 6-month moving realized volatility. This is in line with the findings of Barroso and Detzel (2021), they also found none of the timing rules based on historic volatility to produce significant alpha's except for the case of the momentum factor using the 6-month moving realized volatility timing rule. The certainty equivalent rates show a similar pattern. For the momentum factor, the highest increase in certainty equivalent rate is observed when scaling by the 6-month moving realized volatility. For the other factors, the certainty equivalent rates decrease when employing one of the timing rules based on historic volatility.

When we consider the inference results, the story becomes even less encouraging. Even though the momentum factor produces a higher Sharpe ratio and certainty equivalent rate when using the 6-month moving realized volatility as a timing rule, we are unable to reject the null hypothesis of equal Sharpe ratios and certainty equivalent rates at a 5% confidence level. The only timing rules which produce a significant difference in performance measures are the ones which experience a sharp decline in Sharpe ratio and certainty equivalent rate compared to the unmanaged factor.

Table 6
Net-of-cost performance metrics for managed factors which are constructed using different scaling factors based on historic volatlity measures.

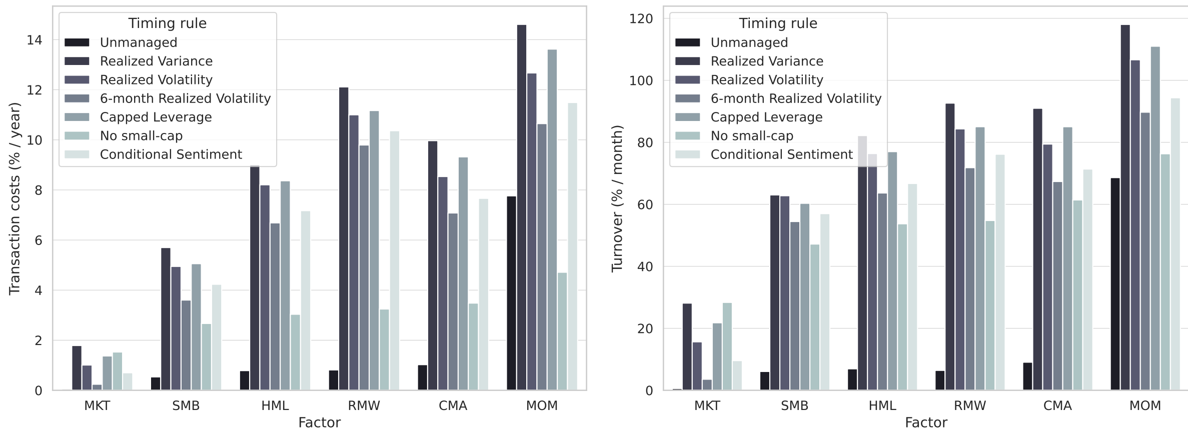
The scaling factors are: the lagged monthly realized variance, the lagged monthly realized volatility, the lagged 6-month realized volatility, the capped leverage of the lagged monthly realized variance, the lagged monthly realized variance without small-cap stocks, and the conditional sentiment scaling factor. We also report performance metric for the unmanaged factors as reference. The performance metrics are the net-of-cost Sharpe ratio and the net-of-cost certainty equivalent rate. The certainty equivalent rate is calculated with an investor risk aversion of 3. The performance metrics are calculated for the period 1976-01 to 2022-06 to cover the same sample period as that of the forecasted volatility models. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Net-of-cost Sharpe ratios for managed factors which are constructed using different scaling factors based on historic volatlity measures. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.519** (−)	**0.202** (−)	**0.220** (−)	**0.420** (−)	**0.337** (−)	-0.039 (−)
Realized Variance	0.424 (0.53)	-0.369 (0.01)	-0.615 (0.00)	-0.962 (0.00)	-1.107 (0.00)	-0.001 (0.88)
Realized Volatility	0.512 (0.94)	-0.259 (0.00)	-0.483 (0.00)	-0.798 (0.00)	-0.825 (0.00)	0.060 (0.46)
6m Realized Volatility	0.513 (0.94)	-0.134 (0.00)	-0.379 (0.00)	-0.685 (0.00)	-0.589 (0.00)	**0.157** (0.18)
Capped Leverage	0.480 (0.72)	-0.204 (0.01)	-0.522 (0.00)	-0.786 (0.00)	-0.930 (0.00)	-0.053 (0.95)
No small-cap	0.439 (0.62)	-0.676 (0.00)	-0.389 (0.04)	-0.561 (0.00)	-0.666 (0.00)	-0.085 (0.79)
Conditional Sentiment	0.515 (0.95)	-0.220 (0.00)	-0.458 (0.00)	-0.736 (0.00)	-0.703 (0.00)	0.040 (0.58)

Panel B: Net-of-cost certainty equivalent rates for managed factors which are constructed using different scaling factors based on historic volatlity measures. Investor risk aversion is set equal to 3. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.37%** (−)	**0.04%** (−)	**0.05%** (−)	**0.20%** (−)	**0.13%** (−)	-0.33% (−)
Realized Variance	0.25% (0.53)	-0.44% (0.00)	-0.69% (0.00)	-0.73% (0.00)	-0.69% (0.00)	-0.28% (0.85)
Realized Volatility	0.36% (0.94)	-0.35% (0.00)	-0.57% (0.00)	-0.62% (0.00)	-0.53% (0.00)	-0.21% (0.49)
6m Realized Volatility	0.36% (0.94)	-0.24% (0.00)	-0.48% (0.00)	-0.54% (0.00)	-0.40% (0.00)	**-0.09%** (0.21)
Capped Leverage	0.32% (0.72)	-0.30% (0.00)	-0.60% (0.00)	-0.61% (0.00)	-0.59% (0.00)	-0.35% (0.92)
No small-cap	0.27% (0.60)	-1.13% (0.00)	-0.22% (0.11)	-0.21% (0.00)	-0.28% (0.00)	-0.14% (0.35)
Conditional Sentiment	0.37% (0.94)	-0.31% (0.00)	-0.55% (0.00)	-0.57% (0.00)	-0.46% (0.00)	-0.23% (0.52)



(a) Average yearly transaction costs in %.

(b) Average monthly turnover in %.

Figure 3
Transaction costs and turnover for the managed portfolios using past realized variance and cost-mitigation techniques.

Forecasted volatility measures

We continue with forming managed portfolios using the forecasted realized variance from each of the optimal models. The net-of-costs Sharpe ratios and certainty equivalent rates of the managed portfolios as well as the p-values for the test on equality of performance measures between unmanaged and managed portfolios are presented in Table 7. We also present the results for the managed portfolios formed using forecasted realized volatility as a timing rule, where realized volatility is calculated as the square root of the forecasted realized variance. These performance metrics and their inference results are presented in Table 8.

Table 7
Net-of-cost performance metrics for VMP using forecasted realized variance.

We report net-of-cost Sharpe ratios and the net-of-cost certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.519** (−)	**0.202** (−)	**0.220** (−)	**0.420** (−)	**0.337** (−)	-0.039 (−)
Realized Variance	0.424 (0.55)	-0.369 (0.00)	-0.615 (0.00)	-0.962 (0.00)	-1.107 (0.00)	-0.001 (0.88)
GARCH	0.451 (0.58)	-0.244 (0.00)	-0.519 (0.00)	-0.693 (0.00)	-0.826 (0.00)	0.155 (0.25)
HAR-RV	0.473 (0.66)	-0.128 (0.03)	-0.460 (0.00)	-0.762 (0.00)	-0.749 (0.00)	0.110 (0.38)
ARFIMA	0.121 (0.39)	-0.246 (0.07)	0.043 (0.52)	-0.091 (0.03)	-0.190 (0.11)	-0.210 (0.24)
MIDAS	0.401 (0.20)	-0.156 (0.00)	-0.407 (0.00)	-0.780 (0.00)	-0.635 (0.00)	0.164 (0.21)
Random Forest	0.467 (0.65)	-0.213 (0.06)	-0.542 (0.00)	-0.838 (0.00)	-0.812 (0.00)	0.107 (0.48)
MFC	0.495 (0.82)	-0.207 (0.00)	-0.470 (0.00)	-0.755 (0.00)	-0.809 (0.00)	0.151 (0.24)
Stacking	0.483 (0.66)	-0.097 (0.00)	-0.323 (0.13)	-0.700 (0.00)	-0.562 (0.00)	**0.177** (0.25)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.37%** (−)	**0.04%** (−)	**0.05%** (−)	**0.20%** (−)	**0.13%** (−)	-0.33% (−)
Realized Variance	0.25% (0.53)	-0.44% (0.00)	-0.69% (0.00)	-0.73% (0.00)	-0.69% (0.00)	-0.28% (0.84)
GARCH	0.28% (0.59)	-0.33% (0.00)	-0.60% (0.00)	-0.55% (0.00)	-0.53% (0.00)	-0.09% (0.23)
HAR-RV	0.31% (0.66)	-0.24% (0.00)	-0.55% (0.00)	-0.59% (0.00)	-0.49% (0.00)	-0.14% (0.39)
ARFIMA	-0.14% (0.04)	-0.33% (0.19)	-0.10% (0.45)	-0.14% (0.01)	-0.17% (0.02)	-0.55% (0.22)
MIDAS	0.22% (0.19)	-0.26% (0.00)	-0.50% (0.00)	-0.60% (0.00)	-0.42% (0.00)	-0.08% (0.23)
Random Forest	0.30% (0.65)	-0.31% (0.00)	-0.62% (0.00)	-0.64% (0.00)	-0.52% (0.00)	-0.15% (0.43)
MFC	0.34% (0.82)	-0.30% (0.00)	-0.56% (0.00)	-0.59% (0.00)	-0.52% (0.00)	-0.09% (0.26)
Stacking	0.32% (0.66)	-0.21% (0.00)	-0.43% (0.25)	-0.55% (0.00)	-0.38% (0.00)	**-0.06%** (0.24)

Table 8
Net-of-cost performance metrics for VMP using forecasted realized volatility.

We report net-of-cost Sharpe ratios and the net-of-cost certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized volatility. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized volatility. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.519 (−)	**0.202** (−)	**0.220** (−)	**0.420** (−)	**0.337** (−)	-0.039 (−)
Realized Volatility	0.512 (0.94)	-0.259 (0.00)	-0.483 (0.00)	-0.798 (0.00)	-0.825 (0.00)	0.060 (0.47)
GARCH	0.505 (0.83)	-0.174 (0.00)	-0.417 (0.00)	-0.609 (0.00)	-0.665 (0.00)	0.081 (0.22)
HAR-RV	0.521 (0.96)	-0.115 (0.00)	-0.377 (0.00)	-0.676 (0.00)	-0.627 (0.00)	0.088 (0.28)
ARFIMA	0.427 (0.31)	-0.308 (0.01)	-0.215 (0.20)	-0.254 (0.01)	-0.371 (0.12)	-0.156 (0.35)
MIDAS	0.471 (0.43)	-0.123 (0.00)	-0.343 (0.00)	-0.659 (0.00)	-0.553 (0.00)	0.102 (0.16)
Random Forest	0.507 (0.85)	-0.165 (0.00)	-0.417 (0.00)	-0.673 (0.00)	-0.635 (0.00)	0.082 (0.29)
MFC	**0.528** (0.89)	-0.143 (0.00)	-0.379 (0.00)	-0.698 (0.00)	-0.622 (0.00)	0.105 (0.20)
Stacking	0.507 (0.80)	-0.088 (0.00)	-0.479 (0.00)	-0.562 (0.00)	-0.502 (0.00)	**0.124** (0.19)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized volatility. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.37% (−)	**0.04%** (−)	**0.05%** (−)	**0.20%** (−)	**0.13%** (−)	-0.33% (−)
Realized Volatility	0.36% (0.93)	-0.35% (0.00)	-0.57% (0.00)	-0.62% (0.00)	-0.53% (0.00)	-0.21% (0.50)
GARCH	0.35% (0.83)	-0.27% (0.00)	-0.51% (0.00)	-0.49% (0.00)	-0.44% (0.00)	-0.18% (0.28)
HAR-RV	0.37% (0.97)	-0.22% (0.00)	-0.48% (0.00)	-0.53% (0.00)	-0.42% (0.00)	-0.17% (0.37)
ARFIMA	0.25% (0.29)	-0.39% (0.05)	-0.33% (0.00)	-0.25% (0.00)	-0.27% (0.00)	-0.48% (0.32)
MIDAS	0.31% (0.42)	-0.23% (0.00)	-0.45% (0.00)	-0.52% (0.00)	-0.38% (0.00)	-0.15% (0.26)
Random Forest	0.36% (0.84)	-0.27% (0.00)	-0.51% (0.00)	-0.53% (0.00)	-0.42% (0.00)	-0.18% (0.37)
MFC	**0.38%** (0.89)	-0.25% (0.00)	-0.48% (0.00)	-0.55% (0.00)	-0.42% (0.00)	-0.15% (0.30)
Stacking	0.35% (0.79)	-0.20% (0.00)	-0.57% (0.00)	-0.46% (0.00)	-0.35% (0.00)	**-0.13%** (0.28)

Table 7 and Table 8 show a similar pattern to the managed portfolios using historic volatility measures. For most factors, the managed portfolios using forecasted realized variance and realized volatility produce lower Sharpe ratios and certainty equivalent rates compared to their unmanaged counterparts. The only exception being the market and momentum factors. The MFC combiner model is able to produce timing rules based on realized volatility which result in slightly higher performance measure for the market factor, but these increases are statistically insignificant with p-values of 0.89. Conversely, for the momentum factor, the Stacking model produces a timing rule based on realized variance and realized volatility which results in a higher Sharpe ratio and certainty equivalent rate compared to the unmanaged factor. However, this increase is statistically insignificant with p-values of 0.25 and 0.24 respectively.

The underperformance of the other factors is not entirely surprising since we already showed that having a 'perfect' forecast of the realized variance does not necessarily lead to and increase in Sharpe ratio. The results for the managed portfolios using forecasted realized variance and volatility support this finding.

For reference, we also report gross performance metrics for the volatility-managed portfolios using forecasted realized variance in Table 9. This shows that applying volatility timing based on past realized variance or forecasted realized variance is able to increase gross Sharpe ratios significantly compared to the unmanaged strategy for the momentum factor. Thus, underlining the fact that due to the transactions costs associated with these strategies, potential gains are eroded. Gross performance metrics for past realized variance based timing rules and the other cost mitigation techniques are presented in Appendix A but not shown here since they show similar results.

Furthermore, we employ the same cost-mitigation techniques as with the historic volatility measures where we limit leverage to be no higher than 1.5, exclude small-cap stocks, and practice conditional sentiment scaling. Our results show that these cost mitigation techniques generally do not alter the main conclusions drawn from the unmitigated strategies. The managed portfolios using these techniques still do not consistently outperform their unmanaged counterparts in terms of net-of-costs performance metrics and show a very similar pattern to the results without cost mitigation. For brevity, we present the results of these additional analyses in Appendix A. But we can conclude that even when limiting leverage, excluding small-cap stocks, and practising conditional sentiment scaling, the managed portfolios using forecasted realized variance still to not perform significantly better than their unmanaged counterparts in terms of net-of-costs Sharpe ratios and net-of-costs certainty equivalent rates.

This leads us to conclude that we are unable to produce timing rules which result in significant improvement in the risk-return characteristics of the factors under consideration over the studied sample period.

Table 9
Gross performance metrics for VMP using forecasted realized variance.

We report gross Sharpe ratios and the gross certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance. Only gross sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.521 (−)	0.254 (−)	0.294 (−)	0.520 (−)	**0.488** (−)	0.484 (−)
Realized Variance	**0.547** (0.87)	**0.286** (0.81)	0.267 (0.89)	**0.696** (0.31)	0.391 (0.37)	**1.058** (<u>0.00</u>)
GARCH	0.481 (0.74)	0.141 (0.23)	**0.301** (0.97)	0.591 (0.64)	0.182 (<u>0.02</u>)	0.822 (0.06)
HAR-RV	0.541 (0.85)	0.265 (0.92)	0.205 (0.62)	0.678 (0.30)	0.397 (0.45)	0.867 (<u>0.04</u>)
ARFIMA	0.161 (0.38)	0.001 (0.39)	0.200 (0.69)	0.155 (0.11)	0.120 (0.12)	-0.081 (<u>0.00</u>)
MIDAS	0.472 (0.58)	0.261 (0.92)	0.274 (0.90)	0.604 (0.48)	0.395 (0.41)	0.896 (<u>0.02</u>)
Random Forest	0.515 (0.95)	0.245 (0.96)	0.186 (0.48)	0.675 (0.24)	0.365 (0.25)	1.015 (<u>0.01</u>)
MFC	0.533 (0.92)	0.201 (0.50)	0.234 (0.70)	0.393 (0.64)	0.338 (0.22)	0.952 (<u>0.01</u>)
Stacking	0.494 (0.73)	0.235 (0.76)	-0.092 (0.32)	0.565 (0.68)	0.353 (0.23)	1.001 (<u>0.01</u>)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance. Investor risk aversion is set equal to 3. Only gross certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.37% (−)	0.09% (−)	0.12% (−)	0.27% (−)	**0.22%** (−)	0.32% (−)
Realized Variance	**0.41%** (0.86)	**0.11%** (0.81)	0.09% (0.89)	**0.39%** (0.34)	0.16% (0.48)	**1.04%** (<u>0.00</u>)
GARCH	0.32% (0.74)	-0.01% (0.21)	**0.13%** (0.96)	0.32% (0.63)	0.05% (<u>0.04</u>)	0.75% (<u>0.04</u>)
HAR-RV	0.40% (0.84)	0.10% (0.91)	0.04% (0.63)	0.37% (0.37)	0.17% (0.56)	0.80% (<u>0.03</u>)
ARFIMA	-0.09% (0.06)	-0.13% (0.42)	0.04% (0.67)	0.02% (0.07)	0.01% (0.11)	-0.38% (<u>0.00</u>)
MIDAS	0.31% (0.59)	0.09% (0.92)	0.10% (0.90)	0.32% (0.53)	0.17% (0.48)	0.84% (<u>0.03</u>)
Random Forest	0.37% (0.96)	0.08% (0.95)	0.02% (0.51)	0.37% (0.33)	0.15% (0.37)	0.99% (<u>0.01</u>)
MFC	0.39% (0.91)	0.04% (0.49)	0.07% (0.70)	0.18% (0.53)	0.13% (0.30)	0.91% (<u>0.01</u>)
Stacking	0.34% (0.73)	0.07% (0.75)	-0.22% (0.36)	0.30% (0.73)	0.14% (0.29)	0.97% (<u>0.01</u>)

7.2 Model selection

In this section we present the results from the model selection / hyperparameter tuning procedure for each of the model classes.

7.2.1 GARCH model selection

As described in Section 5.1, the optimal GARCH(1,1) model (daily/monthly) is selected based on the highest Sharpe ratio of the managed portfolios formed using the forecasted realized variance per factor. The resulting net-of-costs Sharpe ratios as well as the optimal model per factor are presented in Table 10.

Table 10
Model selection criteria for the GARCH models.

Both a daily and monthly GARCH model are estimated using a rolling window with window length of 5 years. The daily forecasts are rescaled to a monthly frequency by multiplying by 22. Managed portfolios are constructed using the forecasted realized variance as the scaling factor. Net-of-costs Sharpe ratios are calculated for each of the managed portfolios. For each row, the highest Sharpe ratio is highlighted.

	Daily	Monthly	Optimal
MKT	0.360	**0.379**	Monthly
SMB	-0.480	**-0.275**	Monthly
HML	**-0.478**	-0.522	Daily
RMW	-0.956	**-0.743**	Monthly
CMA	-0.813	**-0.774**	Monthly
MOM	0.053	**0.109**	Monthly

For most factors, the monthly GARCH model produces higher net-of-costs Sharpe ratios than the daily model. This is in line with Ghysels et al. (2019) who find that using a daily model and rescaling forecasts to a longer horizon produces worse forecasts than directly modelling the longer horizon. However, for the value (HML) factor, the daily GARCH model produces the highest net-of-costs Sharpe ratio. Therefore, we select the daily GARCH model the optimal GARCH model for the value factor.

7.2.2 HAR-RV model selection

Similarly, as with the GARCH model, the HAR-RV model also produces higher Sharpe ratios when using monthly realized variance forecasts compared to daily realized variance forecasts, except for the market factor. The results are presented in Table 11.

We therefore select the daily HAR-RV model as the optimal HAR-RV model for the market factor and the monthly HAR-RV model as the optimal HAR-RV model for the other factors.

7.2.3 ARFIMA model selection

The results of the rolling window out-of-sample evaluation for the ARFIMA model are presented in Table 12. When comparing the specifications across the different values for the fractional differencing parameter d, we notice that both for the daily and monthly there is no clear pattern to be discerned in the net of costs Sharpe ratios. For almost all the factors does a different fractional differencing parameter produce the highest net-of-costs Sharpe ratio. However, when

Table 11
Model selection criteria for the HAR-RV models.

Both a daily and monthly HAR-RV model are estimated using a rolling window with window length of 5 years. The daily forecasts are rescaled to a monthly frequency by multiplying by 22. Managed portfolios are constructed using the forecasted realized variance as the scaling factor. Net-of-costs Sharpe ratios are calculated for each of the managed portfolios. For each row, the highest Sharpe ratio is highlighted.

	Daily	Monthly	Optimal
MKT	**0.387**	0.343	Daily
SMB	-0.420	**-0.304**	Monthly
HML	-0.515	**-0.414**	Monthly
RMW	-1.019	**-0.844**	Monthly
CMA	-0.849	**-0.672**	Monthly
MOM	0.052	**0.137**	Monthly

comparing the daily model with the monthly model, we see that for all factors except the value (HML) factor that the daily model achieves higher net-of-costs Sharpe ratios. The optimal ARFIMA model specifications for each factor are presented in Table 13.

Table 12
Model selection criteria for the ARFIMA models.

Both a daily and monthly ARFIMA model are estimated using a rolling window with window length of 5 years. The daily forecasts are rescaled to a monthly frequency by multiplying by 22. Managed portfolios are constructed using the forecasted realized variance as the scaling factor. Net-of-costs Sharpe ratios are calculated for each of the managed portfolios. For each row, the highest Sharpe ratio is highlighted.

Panel A: Daily ARFIMA(1,d,1) model Sharpe ratios for different d values.

$d =$	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
MKT	0.420	0.318	0.406	**0.421**	0.207	0.125	0.195	0.286	0.120
SMB	-0.247	**-0.036**	-0.334	-0.442	-0.350	-0.057	-0.462	-0.324	-0.299
HML	-0.454	-0.441	-0.547	-0.471	-0.359	-0.313	-0.523	-0.430	**-0.037**
RMW	-0.800	-0.929	-0.808	-0.224	-0.423	-0.588	**0.046**	-0.628	-0.276
CMA	-0.838	-0.865	-0.857	-0.831	-0.801	-0.783	-0.902	-0.672	**-0.012**
MOM	0.069	0.052	0.116	-0.064	0.035	-0.035	0.102	**0.191**	-0.216

Panel B: Monthly ARFIMA(1,d,1) model Sharpe ratios for different d values.

$d =$	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
MKT	0.397	**0.398**	0.261	-0.034	0.137	0.283	0.143	0.030	0.088
SMB	**-0.321**	-0.397	-0.389	-0.401	-0.402	-0.331	-0.391	-0.449	-0.375
HML	-0.427	-0.485	-0.469	-0.472	-0.457	-0.524	-0.445	-0.273	**0.042**
RMW	-0.882	**-0.762**	-0.932	-0.969	-0.999	-1.043	-1.078	-1.109	-1.129
CMA	-0.743	-0.767	-0.777	-0.772	-0.795	-0.831	-0.857	-0.762	**-0.537**
MOM	0.064	0.068	0.053	0.056	**0.082**	0.021	-0.022	-0.075	-0.052

7.2.4 MIDAS model selection

The results of the rolling window out-of-sample evaluation for the MIDAS model are presented in Table 14. When comparing the model with the exponential lag polynomial against the model with the beta lag polynomial, we see that for all factors except the profitability (RMW) factor that the exponential model produces more favourable timing rules. For the exponential lag polynomial models, there is not a substantial difference across the different values for j_{max} with regard to the Sharpe ratios. However, it does seem that higher values tend to produce slightly

Table 13
Optimal ARFIMA model specifications for each factor.

	Frequency	d	ARFIMA(p,d,q)
MKT	Daily	0.2	ARFIMA(1,0.2,1)
SMB	Daily	0.1	ARFIMA(1,0.1,1)
HML	Monthly	0.45	ARFIMA(1,0.45,1)
RMW	Daily	0.35	ARFIMA(1,0.35,1)
CMA	Daily	0.45	ARFIMA(1,0.45,1)
MOM	Daily	0.4	ARFIMA(1,0.4,1)

more favourable timing rules. The optimal MIDAS model specifications for each factor are presented in Table 15.

Table 14
Model selection criteria for the MIDAS models.

Both an exponential and beta MIDAS model are estimated using a rolling window with window length of 5 years. Managed portfolios are constructed using the forecasted realized variance as the scaling factor. Net-of-costs Sharpe ratios are calculated for each of the managed portfolios. For each row, the highest Sharpe ratio is highlighted.

Panel A: Beta MIDAS model net-of-costs Sharpe ratios for different j_{max} values.

$j_{max} =$	20	25	30	35	40	45	50	55	60
MKT	0.287	**0.301**	0.263	0.283	0.297	0.294	0.289	0.293	0.283
SMB	-0.297	-0.305	-0.290	-0.275	-0.250	-0.262	**-0.247**	-0.272	-0.251
HML	-0.473	-0.461	**-0.436**	-0.480	-0.453	-0.449	-0.456	-0.458	-0.465
RMW	-0.888	-0.895	-0.884	-0.874	-0.873	-0.872	-0.878	**-0.865**	-0.878
CMA	-0.660	-0.680	-0.700	-0.734	-0.689	-0.652	**-0.638**	-0.699	-0.670
MOM	0.079	0.031	0.036	0.002	0.056	0.082	0.064	0.072	**0.153**

Panel B: Exponential MIDAS model net-of-costs Sharpe ratios for different j_{max} values.

$j_{max} =$	20	25	30	35	40	45	50	55	60
MKT	0.261	0.293	0.279	0.269	0.258	0.274	0.267	0.289	**0.314**
SMB	-0.254	-0.253	-0.225	-0.231	-0.243	-0.232	-0.225	**-0.221**	-0.230
HML	-0.373	-0.365	-0.369	-0.369	-0.384	-0.365	-0.360	-0.361	**-0.356**
RMW	-0.874	-0.902	-0.876	-0.883	-0.884	-0.880	-0.880	-0.883	**-0.873**
CMA	-0.629	-0.611	-0.616	-0.621	-0.628	-0.625	**-0.611**	-0.618	-0.631
MOM	0.116	0.139	0.093	0.079	**0.169**	0.166	0.157	0.148	0.147

Table 15
Optimal MIDAS model specifications for each factor.

	Lag polynomial	j_{max}
MKT	Exponential	60
SMB	Exponential	55
HML	Exponential	60
RMW	Beta	55
CMA	Exponential	50
MOM	Exponential	45

7.2.5 Random Forest model selection

For the selection of the random forest model, we tune the hyperparameters $(B, n_{\mathcal{B}}, m, n_{\min})$ using Tree-structured Parzen Estimation (Watanabe, 2023) to find the optimal hyperparameters. We only perform hyperparameter tuning for the market factor due to the computational load of this procedure. These hyperparameters are then used to form the Random Forest models for the other factors as well. The results of the tuning procedure are presented in Table 16. The results show that the random forest model with $B = 440$, $n_{\mathcal{B}} = 0.64$, $m = 0.36$, and $n_{\min} = 3$ produces the highest net-of-costs Sharpe ratio for the market factor. We use these hyperparameters to form the Random Forest models for the market factor and the other factors as well.

Table 16
Optimal model parameters for the random forest model.

The optimal parameters are selected based on the maximisation of the net-of-costs Sharpe ratio for the managed market factor. Forecasted realized variance using a rolling window with a window length of 5 years is used for construction of the managed portfolio. The optimal parameters are selected using Tree-structured Parzen Estimation.

	No. of trees B	Subsample ratio $n_{\mathcal{B}}$	Feature ratio m	Min. node size $n_{\min}$
Range	50-500	0.05-1	0.05-1	1-50
Optimal value	440	0.64	0.36	3

7.2.6 Stacking model selection

For the stacking model selection we again apply Tree-structured Parzen Estimation (Watanabe, 2023) for tuning the penalty terms for the Ridge, Lasso and Elastic Net models. The tuning is performed for each factor separately. The results of the tuning procedure are presented in Table 17. These optimal model parameters are then used to forecast the realized variance and produce timing rules. The results of the rolling window out-of-sample evaluation for the stacking model as well as the optimal model per factor are presented in Table 18.

Table 17
Optimal model parameters for the stacking models.

The optimal parameters are selected based on the maximisation of the net-of-costs Sharpe ratio for each factor. Forecasted realized variance using a rolling window with a window length of 5 years is used for construction of the managed portfolios. The optimal parameters are selected using Tree-structured Parzen Estimation.

	Ridge	Lasso		Elastic net
	λ_{opt}	λ_{opt}	α_{opt}	λ_{opt}
MKT	0.137	999.529	0.950	953.820
SMB	0.162	30.100	0.073	475.333
HML	0.226	945.784	0.440	773.005
RMW	999.821	947.099	0.968	426.149
CMA	999.780	438.741	0.545	310.894
MOM	999.517	0.155	0.207	0.265

Table 18
Model selection criteria for the Stacking models.

All models are estimated using a rolling window with window length of 5 years. Managed portfolios are constructed using the forecasted realized variance as the scaling factor. Net-of-costs Sharpe ratios are calculated for each of the managed portfolios. For each row, the highest Sharpe ratio is highlighted.

	Ols	Ridge	Lasso	Elastic net	Optimal
MKT	0.479	0.172	0.483	**0.483**	Elastic net
SMB	-0.213	-0.228	-0.100	**-0.097**	Elastic net
HML	-0.487	**-0.181**	-0.375	-0.379	Ridge
RMW	-0.722	-0.734	**-0.700**	-0.704	Lasso
CMA	-0.727	-0.662	**-0.562**	-0.562	Lasso
MOM	0.179	**0.183**	0.073	0.099	Ols

8 Conclusion

Volatility-managed portfolios seem to offer superior risk reward characteristics compared to traditional buy and hold strategies. However, our findings show that while volatility-managed portfolios based on past realized variance can offer attractive risk reward characteristics, these gains can be easily offset by transaction costs. When we consider a range of cost mitigation techniques in the spirit of Novy-Marx and Velikov (2016), we still find that net-of-costs performance metrics are unable to outperform traditional factor strategies. These findings are in line with the results of Barroso and Detzel (2021), who demonstrate transaction costs can substantially diminish the performance of these strategies.

By forecasting realized variance to use as a timing signal instead of past realized variance we aim to build upon the work of Moreira and Muir (2017) and to improve the risk return trade-off of these strategies. We incorporate advanced volatility forecasting models, such as GARCH, HAR-RV, ARFIMA, MIDAS and Random Forest, as well as combiner models like mean forecast combination and stacking. However, after introducing forecasted realized variance as a timing signal, we are still unable to produce significantly better risk return characteristics compared to the traditional unmanaged factor strategies. For most factors these managed strategies actually lead to significantly worse performance compared to their unmanaged counterparts. Leading us to conclude that the forecasting of factor realized variance is not a viable strategy for improving the risk reward characteristics of any factor strategy under consideration.

Furthermore, we find that, when one would have a 'perfect' forecast of next months' realized variance to be used as a timing signal instead of past realized variance, the Sharpe ratio of the market (MKT), size (SMB) and momentum (MOM) factor experience an increase while the Sharpe ratio of the value (HML), profitability (RMW) and investment (CMA) factor decreases. An explanation for this phenomenon is outside the scope of this research and is left for future research. However, it does make the framework of Moreira and Muir (2017) more puzzling. It would make for an interesting research question to see where this discrepancy comes from.

We find that most of the constructed timing rules lead to significantly worse performance compared to the traditional unmanaged factor strategies. In some cases this leads to substantially negative Sharpe ratios. We restricted ourselves solely to the application of volatility timing as a means of improving mean variance characteristics of the factor strategies. However, it could be interesting to explore whether shorting these heavily underperforming strategies could lead to interesting trading strategies.

Also, in our analysis we only considered volatility-managed portfolios consisting of a single managed factor. Another approach would be to follow the work of DeMiguel et al. (2021) and consider portfolios consisting of multiple factors. In their analysis, DeMiguel et al. (2021) form multifactor volatility-managed portfolios in which the individual factor weights are varied over time as a function of market variance. They find that this approach to forming a multifactor volatility-managed portfolio leads to improved risk reward characteristics compared to its unmanaged counterpart. Extending this research by considering forecasted realized variance as a timing signal could be an interesting topic for future research.

As shown by Barroso and Detzel (2021) and DeMiguel et al. (2021) as well as our analysis, transaction costs can have a substantial impact on the performance of volatility-managed portfolios. Therefore, to improve the performance of these strategies, it is important to consider more advanced cost mitigation techniques. Rebalancing the portfolios at lower frequencies could help in further reducing transaction costs at the expense of reacting to changes in market volatility. Furthermore, introducing hysteresis in the rebalancing rules as proposed by Novy-Marx and Velikov (2019) and in the process generate timing rules which require less frequent rebalancing could also help in reducing transaction costs.

In summary, our research contributes to the growing body of literature on volatility-managed portfolios. While we did not find that using forecasted realized variance significantly enhances factor strategy performance, this study provides insights into the practical limitations of implementing such strategies, and highlights the difficulties in consistently outperforming traditional factor strategies through the use of volatility timing.

References

Andersen, T. G., & Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review*, *39*(4), 885–905. https://doi.org/10.2307/2527343

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and Forecasting Realized Volatility. *Econometrica*, *71*(2), 579–625. https://doi.org/10.1111/1468-0262.00418

Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-Time Measurement of Business Conditions. *Journal of Business & Economic Statistics*, *27*(4), 417–427. Retrieved April 23, 2024, from https://www.jstor.org/stable/27799095

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, *73*(1), 5–59. https://doi.org/10.1016/0304-4076(95)01732-1

Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, *61*(4), 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty*. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Barroso, P., & Detzel, A. (2021). Do limits to arbitrage explain the benefits of volatility-managed portfolios? *Journal of Financial Economics*, *140*(3), 744–767. https://doi.org/10.1016/j.jfineco.2021.02.009

Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, *116*(1), 111–120. https://doi.org/10.1016/j.jfineco.2014.11.010

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70–83. https://doi.org/10.1016/j.csda.2017.11.003

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. *Advances in Neural Information Processing Systems*, *24*. Retrieved July 13, 2024, from https://papers.nips.cc/paper_files/paper/2011/hash/86e8f7ab32cfd12577bc2619bc635690-Abstract.html

Brailsford, T. J., & Faff, R. W. (1996). An evaluation of volatility forecasting techniques. *Journal of Banking & Finance*, *20*(3), 419–438. https://doi.org/10.1016/0378-4266(95)00015-1

Breiman, L. (1996). Stacked regressions. *Machine Learning*, *24*(1), 49–64. https://doi.org/10.1007/BF00117832

Cederburg, S., O'Doherty, M. S., Wang, F., & Yan, X. ((2020). On the performance of volatility-managed portfolios. *Journal of Financial Economics*, *138*(1), 95–117. https://doi.org/10.1016/j.jfineco.2020.04.015

Chen, A.-S., Chang, H.-C., & Cheng, L.-Y. (2019). Time-varying Variance Scaling: Application of the Fractionally Integrated ARMA Model. *The North American Journal of Economics and Finance*, *47*, 1–12. https://doi.org/10.1016/j.najef.2018.11.007

Chordia, T., Subrahmanyam, A., & Tong, Q. (2014). Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics*, *58*(1), 41–58. https://doi.org/10.1016/j.jacceco.2014.06.001

Christensen, K., Siggaard, M., & Veliyev, B. (2023). A Machine Learning Approach to Volatility Forecasting*. *Journal of Financial Econometrics*, *21*(5), 1680–1727. https://doi.org/10.1093/jjfinec/nbac020

Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, *7*(2), 174–196. https://doi.org/10.1093/jjfinec/nbp001

Corwin, S. A., & Schultz, P. (2012). A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *The Journal of Finance*, *67*(2), 719–759. Retrieved July 14, 2024, from https://www.jstor.org.tudelft.idm.oclc.org/stable/41419709

DeMiguel, V., Martin-Utrera, A., & Uppal, R. (2021, December 10). *A Multifactor Perspective on Volatility-Managed Portfolios*. https://doi.org/10.2139/ssrn.3982504

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25. https://doi.org/10.1016/j.jfineco.2013.10.005

Ghysels, E., Plazzi, A., Valkanov, R., Rubia, A., & Dossani, A. (2019). Direct Versus Iterated Multiperiod Volatility Forecasts. *Annual Review of Financial Economics*, *11*, 173–195. https://doi.org/10.1146/annurev-financial-110217-022808

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, *76*(3), 509–548. https://doi.org/10.1016/j.jfineco.2004.03.008

Ghysels, E., Sinko, A., & Valkanov, R. I. (2006, February 1). *Midas Regressions: Further Results and New Directions*. https://doi.org/10.2139/ssrn.885683

Ghysels, E., Valkanov, R. I., & Serrano, A. R. (2009, February 16). *Multi-Period Forecasts of Volatility: Direct, Iterated, and Mixed-Data Approaches*. https://doi.org/10.2139/ssrn.1344742

Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, *3*(2), 197–204. https://doi.org/10.1002/for.3980030207

Harris, L. (1990). Statistical Properties of the Roll Serial Covariance Bid/Ask Spread Estimator. *The Journal of Finance*, *45*(2), 579–590. https://doi.org/10.2307/2328671

Hasbrouck, J. (2009). Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data. *The Journal of Finance*, *64*(3), 1445–1477. https://doi.org/10.1111/j.1540-6261.2009.01469.x

Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, *1*, 278–282 vol.1. https://doi.org/10.1109/ICDAR.1995.598994

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.2307/1267351

Hou, K., Xue, C., & Zhang, L. (2015). Digesting Anomalies: An Investment Approach. *The Review of Financial Studies*, *28*(3), 650–705. https://doi.org/10.1093/rfs/hhu068

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, *48*(1), 65–91. https://doi.org/10.2307/2328882

Konstantinidi, E., Skiadopoulos, G., & Tzagkaraki, E. (2008). Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *Journal of Banking & Finance*, *32*(11), 2401–2411. https://doi.org/10.1016/j.jbankfin.2008.02.003

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica*, *53*(6), 1315–1335. https://doi.org/10.2307/1913210

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859. https://doi.org/10.1016/j.jempfin.2008.03.002

Linnainmaa, J. T., & Roberts, M. R. (2018). The History of the Cross-Section of Stock Returns. *The Review of Financial Studies*, *31*(7), 2606–2649. https://doi.org/10.1093/rfs/hhy030

Molnár, P. (2012). Properties of range-based volatility estimators. *International Review of Financial Analysis*, *23*, 20–29. https://doi.org/10.1016/j.irfa.2011.06.012

Moreira, A., & Muir, T. (2017). Volatility-Managed Portfolios. *The Journal of Finance*, *72*(4), 1611–1643. Retrieved March 1, 2024, from https://www.jstor.org/stable/26652549

Novy-Marx, R., & Velikov, M. (2016). A Taxonomy of Anomalies and Their Trading Costs. *The Review of Financial Studies*, *29*(1), 104–147. https://doi.org/10.1093/rfs/hhv063

Novy-Marx, R., & Velikov, M. (2019). Comparing Cost-Mitigation Techniques. *Financial Analysts Journal*, *75*(1), 85–102. https://doi.org/10.1080/0015198X.2018.1547057

Palm, F. C., & Zellner, A. (1992). To Combine or Not to Combine? Issues of Combining Forecasts. *Journal of Forecasting*, *11*(8), 687. Retrieved July 15, 2024, from https://www.proquest.com/docview/219158280/abstract/E458B52E4CF24610PQ/1

Politis, D. N., & Romano, J. P. (1991). *A Circular Block-resampling Procedure for Stationary Data*. Purdue University. Department of Statistics.

Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *The Journal of Finance*, *39*(4), 1127–1139. https://doi.org/10.1111/j.1540-6261.1984.tb03897.x

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. Retrieved May 27, 2024, from https://www.jstor.org/stable/2346178

Wang, X., Hyndman, R. J., Li, F., & Kang, Y. (2022, September 23). *Forecast combinations: An over 50-year review*. arXiv: 2205.04216 [stat]. Retrieved July 15, 2024, from http://arxiv.org/abs/2205.04216

Wang, Y., Ma, F., Wei, Y., & Wu, C. (2016). Forecasting realized volatility in a changing world: A dynamic model averaging approach. *Journal of Banking & Finance*, *64*, 136–149. https://doi.org/10.1016/j.jbankfin.2015.12.010

Watanabe, S. (2023, May 26). *Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance*. arXiv: 2304.11127 [cs]. https://doi.org/10.48550/arXiv.2304.11127

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Yu, J., & Yuan, Y. (2011). Investor sentiment and the mean–variance relation. *Journal of Financial Economics*, *100*(2), 367–381. https://doi.org/10.1016/j.jfineco.2010.10.011

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

A Additional results

Table 19
Gross performance metrics for managed factors which are constructed using different scaling factors based on historic volatlity measures.

The scaling factors are: the lagged monthly realized variance, the lagged monthly realized volatility, the lagged 6-month realized volatility, the capped leverage of the lagged monthly realized variance, the lagged monthly realized variance without small-cap stocks, and the conditional sentiment scaling factor. We also report performance metric for the unmanaged factors as reference. The performance metrics are the net-of-cost Sharpe ratio and the net-of-cost certainty equivalent rate. The certainty equivalent rate is calculated with an investor risk aversion of 3. The performance metrics are calculated for the period 1976-01 to 2022-06 to cover the same sample period as that of the forecasted volatility models. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Gross Sharpe ratios for managed factors which are constructed using different scaling factors based on historic volatlity measures. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.521 (−)	0.254 (−)	0.294 (−)	0.520 (−)	**0.488** (−)	0.484 (−)
Realized Variance	0.547 (0.87)	0.286 (0.81)	0.267 (0.88)	0.696 (0.32)	0.391 (0.39)	**1.058** (<u>0.00</u>)
Realized Volatility	**0.581** (0.50)	0.259 (0.94)	**0.320** (0.82)	0.687 (0.12)	0.452 (0.60)	0.969 (<u>0.00</u>)
6m Realized Volatility	0.528 (0.93)	0.241 (0.88)	0.282 (0.92)	0.631 (0.18)	0.474 (0.84)	0.913 (<u>0.01</u>)
Capped Leverage	0.571 (0.65)	**0.333** (0.45)	0.286 (0.95)	0.708 (0.09)	0.470 (0.78)	0.916 (<u>0.00</u>)
No small-cap	0.543 (0.90)	-0.487 (<u>0.00</u>)	0.173 (0.57)	0.327 (0.34)	0.135 (<u>0.03</u>)	0.559 (0.69)
Conditional Sentiment	0.566 (0.44)	0.247 (0.90)	0.259 (0.73)	**0.708** (0.14)	0.458 (0.69)	0.877 (<u>0.00</u>)

Panel B: Gross certainty equivalent rates for managed factors which are constructed using different scaling factors based on historic volatlity measures. Investor risk aversion is set equal to 3. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.37% (−)	0.09% (−)	0.12% (−)	0.27% (−)	**0.22%** (−)	0.32% (−)
Realized Variance	0.41% (0.86)	0.11% (0.81)	0.09% (0.89)	0.39% (0.35)	0.16% (0.49)	**1.04%** (<u>0.00</u>)
Realized Volatility	**0.45%** (0.48)	0.09% (0.94)	**0.14%** (0.82)	0.38% (0.13)	0.20% (0.70)	0.93% (<u>0.00</u>)
6m Realized Volatility	0.38% (0.93)	0.08% (0.87)	0.11% (0.92)	0.34% (0.29)	0.21% (0.88)	0.86% (<u>0.01</u>)
Capped Leverage	0.44% (0.64)	**0.15%** (0.44)	0.11% (0.95)	0.39% (0.12)	0.21% (0.83)	0.86% (<u>0.00</u>)
No small-cap	0.40% (0.90)	-0.89% (<u>0.00</u>)	0.04% (0.65)	0.09% (0.12)	0.03% (<u>0.03</u>)	0.29% (0.88)
Conditional Sentiment	0.43% (0.45)	0.08% (0.89)	0.09% (0.71)	**0.39%** (0.10)	0.20% (0.73)	0.82% (<u>0.00</u>)

Table 20
Gross performance metrics for VMP using forecasted realized volatility.

We report gross Sharpe ratios and the gross certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized volatility. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized volatility. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.521 (–)	0.254 (–)	0.294 (–)	0.520 (–)	**0.488** (–)	0.484 (–)
Realized Volatility	**0.581** (0.51)	0.259 (0.94)	0.320 (0.82)	**0.687** (0.11)	0.452 (0.60)	**0.969** (0.00)
GARCH	0.522 (0.99)	0.204 (0.30)	**0.327** (0.73)	0.632 (0.19)	0.340 (0.02)	0.775 (0.02)
HAR-RV	0.561 (0.54)	0.259 (0.94)	0.280 (0.89)	0.647 (0.13)	0.447 (0.54)	0.812 (0.04)
ARFIMA	0.485 (0.74)	0.135 (0.54)	0.284 (0.95)	0.310 (0.25)	0.213 (0.16)	0.101 (0.02)
MIDAS	0.508 (0.83)	**0.263** (0.86)	0.304 (0.91)	0.590 (0.30)	0.446 (0.45)	0.801 (0.03)
Random Forest	0.535 (0.83)	0.238 (0.80)	0.258 (0.70)	0.626 (0.15)	0.429 (0.30)	0.853 (0.02)
MFC	0.550 (0.66)	0.237 (0.73)	0.285 (0.93)	0.586 (0.40)	0.413 (0.21)	0.836 (0.03)
Stacking	0.515 (0.89)	0.248 (0.86)	0.175 (0.47)	0.561 (0.58)	0.427 (0.26)	0.872 (0.02)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized volatility. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.37% (–)	0.09% (–)	0.12% (–)	0.27% (–)	**0.22%** (–)	0.32% (–)
Realized Volatility	**0.45%** (0.49)	0.09% (0.95)	0.14% (0.82)	**0.38%** (0.13)	0.20% (0.68)	**0.93%** (0.00)
GARCH	0.37% (0.99)	0.04% (0.32)	**0.15%** (0.73)	0.34% (0.19)	0.14% (0.04)	0.69% (0.01)
HAR-RV	0.42% (0.55)	0.09% (0.93)	0.11% (0.89)	0.35% (0.23)	0.20% (0.64)	0.73% (0.02)
ARFIMA	0.33% (0.66)	-0.01% (0.54)	0.11% (0.94)	0.13% (0.17)	0.06% (0.16)	-0.16% (0.02)
MIDAS	0.36% (0.82)	**0.09%** (0.86)	0.13% (0.92)	0.31% (0.38)	0.20% (0.52)	0.72% (0.02)
Random Forest	0.39% (0.83)	0.07% (0.78)	0.09% (0.69)	0.34% (0.27)	0.19% (0.40)	0.79% (0.01)
MFC	0.41% (0.64)	0.07% (0.74)	0.11% (0.92)	0.31% (0.40)	0.18% (0.29)	0.76% (0.01)
Stacking	0.36% (0.89)	0.08% (0.87)	0.01% (0.47)	0.30% (0.64)	0.19% (0.36)	0.81% (0.01)

Table 21
Net-of-cost performance metrics for VMP using forecasted realized variance, where the leverage has been capped by 1.5.

We report net-of-cost Sharpe ratios and the net-of-cost certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance, where the leverage has been capped by 1.5. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance with capped leverage of 1.5. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.519** (−)	**0.202** (−)	**0.220** (−)	**0.420** (−)	**0.337** (−)	-0.039 (−)
Realized Variance	0.405 (0.01)	-0.188 (0.00)	-0.252 (0.00)	-0.271 (0.03)	-0.374 (0.00)	-0.132 (0.18)
GARCH	0.474 (0.39)	-0.022 (0.00)	-0.262 (0.00)	-0.248 (0.05)	-0.230 (0.00)	-0.108 (0.26)
HAR-RV	0.441 (0.09)	-0.103 (0.00)	-0.196 (0.00)	-0.190 (0.03)	-0.259 (0.00)	-0.102 (0.40)
ARFIMA	0.514 (0.36)	-0.079 (0.00)	-0.241 (0.00)	-0.364 (0.01)	-0.412 (0.00)	**-0.021** (0.33)
MIDAS	0.464 (0.19)	-0.108 (0.00)	-0.219 (0.00)	-0.134 (0.04)	-0.262 (0.01)	-0.136 (0.27)
Random Forest	0.427 (0.07)	-0.104 (0.00)	-0.176 (0.00)	-0.177 (0.04)	-0.259 (0.00)	-0.107 (0.39)
MFC	0.463 (0.29)	-0.050 (0.00)	-0.196 (0.00)	-0.209 (0.03)	-0.224 (0.01)	-0.107 (0.37)
Stacking	0.489 (0.57)	-0.038 (0.00)	-0.225 (0.00)	-0.107 (0.06)	-0.204 (0.01)	-0.128 (0.37)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance with capped leverage of 1.5. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.37%** (−)	**0.04%** (−)	**0.05%** (−)	**0.20%** (−)	**0.13%** (−)	-0.33% (−)
Realized Variance	0.22% (0.01)	-0.29% (0.00)	-0.36% (0.00)	-0.26% (0.00)	-0.27% (0.00)	-0.45% (0.06)
GARCH	0.31% (0.38)	-0.15% (0.00)	-0.37% (0.00)	-0.25% (0.00)	-0.19% (0.00)	-0.42% (0.16)
HAR-RV	0.27% (0.08)	-0.21% (0.00)	-0.32% (0.00)	-0.21% (0.00)	-0.21% (0.00)	-0.41% (0.23)
ARFIMA	0.36% (0.38)	-0.19% (0.00)	-0.36% (0.00)	-0.32% (0.00)	-0.30% (0.00)	**-0.31%** (0.42)
MIDAS	0.30% (0.19)	-0.22% (0.00)	-0.34% (0.00)	-0.17% (0.00)	-0.21% (0.00)	-0.45% (0.15)
Random Forest	0.25% (0.07)	-0.21% (0.00)	-0.30% (0.00)	-0.20% (0.00)	-0.21% (0.00)	-0.42% (0.26)
MFC	0.30% (0.28)	-0.17% (0.00)	-0.31% (0.00)	-0.22% (0.00)	-0.19% (0.00)	-0.42% (0.29)
Stacking	0.33% (0.55)	-0.16% (0.00)	-0.34% (0.00)	-0.15% (0.00)	-0.18% (0.00)	-0.44% (0.22)

Table 22
Gross performance metrics for VMP using forecasted realized variance, where the leverage has been capped by 1.5.

We report gross Sharpe ratios and the gross certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance, where the leverage has been capped by 1.5. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance with capped leverage of 1.5. Only gross sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.521** (−)	0.254 (−)	0.294 (−)	**0.520** (−)	0.488 (−)	0.484 (−)
Realized Variance	0.449 (0.10)	0.224 (0.57)	0.249 (0.37)	0.384 (0.04)	0.443 (0.27)	0.270 (0.01)
GARCH	0.487 (0.50)	**0.323** (0.15)	0.224 (0.23)	0.387 (0.04)	**0.547** (0.31)	0.315 (0.00)
HAR-RV	0.476 (0.30)	0.229 (0.64)	0.247 (0.44)	0.362 (0.04)	0.452 (0.47)	0.279 (0.00)
ARFIMA	0.518 (0.27)	0.256 (0.75)	0.282 (0.28)	0.505 (0.46)	0.475 (0.75)	**0.507** (0.25)
MIDAS	0.497 (0.55)	0.238 (0.71)	0.227 (0.32)	0.396 (0.06)	0.452 (0.53)	0.266 (0.01)
Random Forest	0.453 (0.18)	0.255 (0.99)	0.274 (0.77)	0.366 (0.04)	0.455 (0.55)	0.258 (0.01)
MFC	0.485 (0.48)	0.289 (0.50)	0.246 (0.48)	0.393 (0.02)	0.484 (0.95)	0.260 (0.00)
Stacking	0.496 (0.61)	0.264 (0.86)	**0.297** (0.50)	0.441 (0.34)	0.482 (0.91)	0.235 (0.01)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance with capped leverage of 1.5. Investor risk aversion is set equal to 3. Only gross certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.37%** (−)	0.09% (−)	0.12% (−)	**0.27%** (−)	0.22% (−)	0.32% (−)
Realized Variance	0.28% (0.10)	0.06% (0.56)	0.08% (0.37)	0.18% (0.03)	0.19% (0.30)	0.06% (0.00)
GARCH	0.33% (0.49)	**0.14%** (0.12)	0.06% (0.22)	0.18% (0.03)	**0.25%** (0.31)	0.11% (0.00)
HAR-RV	0.31% (0.30)	0.07% (0.62)	0.08% (0.43)	0.16% (0.01)	0.20% (0.49)	0.07% (0.00)
ARFIMA	0.37% (0.45)	0.09% (0.75)	0.11% (0.36)	0.26% (0.49)	0.21% (0.75)	**0.35%** (0.29)
MIDAS	0.34% (0.56)	0.07% (0.71)	0.06% (0.29)	0.18% (0.03)	0.20% (0.56)	0.05% (0.00)
Random Forest	0.29% (0.18)	0.09% (0.99)	0.10% (0.77)	0.16% (0.01)	0.20% (0.58)	0.04% (0.00)
MFC	0.33% (0.47)	0.12% (0.48)	0.08% (0.47)	0.18% (0.01)	0.22% (0.96)	0.04% (0.00)
Stacking	0.34% (0.60)	0.09% (0.86)	**0.12%** (0.49)	0.21% (0.27)	0.22% (0.92)	0.01% (0.00)

Table 23
Net-of-cost performance metrics for VMP using forecasted realized variance, where small cap stocks are excluded from portfolio formation.

We report net-of-cost Sharpe ratios and the net-of-cost certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance, where small cap stocks are excluded from portfolio formation. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance with no small cap stocks. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.519** (–)	**0.202** (–)	**0.220** (–)	**0.420** (–)	**0.337** (–)	-0.039 (–)
Realized Variance	0.439 (0.61)	-0.676 (0.00)	-0.389 (0.05)	-0.561 (0.00)	-0.666 (0.00)	-0.085 (0.79)
GARCH	0.471 (0.70)	-0.691 (0.00)	-0.393 (0.01)	-0.283 (0.02)	-0.797 (0.00)	0.064 (0.55)
HAR-RV	0.486 (0.78)	-0.651 (0.00)	-0.466 (0.00)	-0.536 (0.00)	-0.679 (0.00)	-0.017 (0.90)
ARFIMA	0.129 (0.37)	-0.299 (0.03)	0.075 (0.63)	-0.161 (0.03)	-0.021 (0.13)	-0.227 (0.23)
MIDAS	0.412 (0.26)	-0.684 (0.00)	-0.523 (0.00)	-0.545 (0.00)	-0.590 (0.00)	**0.069** (0.48)
Random Forest	0.480 (0.75)	-0.656 (0.00)	-0.541 (0.00)	-0.548 (0.00)	-0.689 (0.00)	-0.005 (0.84)
MFC	0.508 (0.92)	-0.700 (0.00)	-0.504 (0.00)	-0.504 (0.00)	-0.736 (0.00)	0.017 (0.74)
Stacking	0.492 (0.75)	-0.697 (0.00)	-0.704 (0.03)	-0.503 (0.00)	-0.582 (0.00)	0.048 (0.60)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance with no small cap stocks. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.37%** (–)	**0.04%** (–)	**0.05%** (–)	**0.20%** (–)	**0.13%** (–)	-0.33% (–)
Realized Variance	0.27% (0.60)	-1.13% (0.00)	-0.22% (0.11)	-0.21% (0.00)	-0.28% (0.00)	-0.14% (0.36)
GARCH	0.31% (0.69)	-1.15% (0.00)	-0.22% (0.08)	-0.11% (0.01)	-0.33% (0.00)	-0.04% (0.17)
HAR-RV	0.33% (0.75)	-1.10% (0.00)	-0.25% (0.05)	-0.20% (0.00)	-0.29% (0.00)	-0.09% (0.29)
ARFIMA	-0.13% (0.05)	-0.66% (0.13)	-0.00% (0.79)	-0.07% (0.06)	-0.03% (0.10)	-0.23% (0.51)
MIDAS	0.23% (0.26)	-1.14% (0.00)	-0.28% (0.01)	-0.20% (0.00)	-0.25% (0.00)	**-0.03%** (0.17)
Random Forest	0.32% (0.74)	-1.10% (0.00)	-0.29% (0.03)	-0.20% (0.00)	-0.29% (0.00)	-0.08% (0.24)
MFC	0.35% (0.92)	-1.16% (0.00)	-0.27% (0.04)	-0.19% (0.00)	-0.31% (0.00)	-0.07% (0.22)
Stacking	0.33% (0.74)	-1.16% (0.00)	-0.36% (0.04)	-0.19% (0.00)	-0.25% (0.00)	-0.05% (0.18)

Table 24
Gross performance metrics for VMP using forecasted realized variance, where small cap stocks are excluded from portfolio formation.

We report gross Sharpe ratios and the gross certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance, where small cap stocks are excluded from portfolio formation. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance with no small cap stocks. Only gross sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.521 (−)	**0.254** (−)	**0.294** (−)	**0.520** (−)	**0.488** (−)	0.484 (−)
Realized Variance	0.543 (0.89)	-0.487 (<u>0.00</u>)	0.173 (0.57)	0.327 (0.34)	0.135 (<u>0.03</u>)	0.559 (0.66)
GARCH	0.498 (0.84)	-0.530 (<u>0.00</u>)	0.176 (0.52)	0.502 (0.93)	-0.073 (<u>0.00</u>)	0.538 (0.77)
HAR-RV	**0.546** (0.83)	-0.469 (<u>0.00</u>)	0.062 (0.16)	0.340 (0.39)	0.126 (<u>0.01</u>)	0.471 (0.94)
ARFIMA	0.158 (0.38)	-0.278 (<u>0.01</u>)	0.177 (0.62)	0.004 (0.07)	0.143 (0.09)	-0.103 (<u>0.00</u>)
MIDAS	0.476 (0.62)	-0.508 (<u>0.00</u>)	0.069 (0.10)	0.298 (0.28)	0.147 (<u>0.01</u>)	0.524 (0.83)
Random Forest	0.523 (0.99)	-0.466 (<u>0.00</u>)	0.046 (0.16)	0.338 (0.36)	0.108 (<u>0.01</u>)	0.566 (0.65)
MFC	0.542 (0.85)	-0.519 (<u>0.00</u>)	0.077 (0.20)	0.159 (0.15)	0.073 (<u>0.00</u>)	0.547 (0.73)
Stacking	0.502 (0.81)	-0.513 (<u>0.00</u>)	-0.026 (0.27)	0.321 (0.31)	0.120 (<u>0.00</u>)	**0.583** (0.60)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance with no small cap stocks. Investor risk aversion is set equal to 3. Only gross certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.37% (−)	**0.09%** (−)	**0.12%** (−)	**0.27%** (−)	**0.22%** (−)	**0.32%** (−)
Realized Variance	0.40% (0.89)	-0.89% (<u>0.00</u>)	0.04% (0.65)	0.09% (0.12)	0.03% (<u>0.04</u>)	0.29% (0.89)
GARCH	0.34% (0.85)	-0.95% (<u>0.00</u>)	0.04% (0.63)	0.15% (0.32)	-0.06% (<u>0.00</u>)	0.28% (0.84)
HAR-RV	**0.40%** (0.85)	-0.87% (<u>0.00</u>)	-0.01% (0.42)	0.09% (0.14)	0.02% (<u>0.03</u>)	0.23% (0.68)
ARFIMA	-0.09% (0.06)	-0.63% (0.12)	0.04% (0.72)	-0.02% (<u>0.04</u>)	0.03% (0.07)	-0.15% (<u>0.00</u>)
MIDAS	0.31% (0.62)	-0.92% (<u>0.00</u>)	-0.01% (0.34)	0.08% (0.10)	0.03% (<u>0.02</u>)	0.27% (0.79)
Random Forest	0.37% (1.00)	-0.87% (<u>0.00</u>)	-0.02% (0.39)	0.09% (0.12)	0.01% (<u>0.02</u>)	0.30% (0.90)
MFC	0.40% (0.85)	-0.93% (<u>0.00</u>)	-0.00% (0.42)	0.03% (<u>0.04</u>)	0.00% (<u>0.01</u>)	0.29% (0.85)
Stacking	0.35% (0.81)	-0.93% (<u>0.00</u>)	-0.05% (0.40)	0.09% (0.09)	0.02% (<u>0.01</u>)	0.31% (0.95)

Table 25
Net-of-cost performance metrics for VMP using forecasted realized variance, conditional on the sentiment index.

We report net-of-cost Sharpe ratios and the net-of-cost certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance, where voltatility timing takes place only when the sentiment index is above its median value in an expanding window. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance conditional on the sentiment index. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.519 (−)	**0.202** (−)	**0.220** (−)	**0.420** (−)	**0.337** (−)	-0.039 (−)
Realized Variance	0.431 (0.46)	-0.131 (0.06)	-0.528 (0.00)	-0.826 (0.00)	-0.911 (0.00)	0.136 (0.38)
GARCH	0.516 (0.98)	-0.303 (0.00)	-0.472 (0.00)	-0.573 (0.00)	-0.667 (0.00)	0.133 (0.28)
HAR-RV	0.460 (0.50)	-0.083 (0.01)	-0.416 (0.00)	-0.666 (0.00)	-0.678 (0.00)	0.052 (0.56)
ARFIMA	0.122 (0.37)	-0.300 (0.36)	0.040 (0.51)	-0.060 (0.03)	-0.143 (0.12)	-0.272 (0.05)
MIDAS	0.465 (0.49)	-0.027 (0.06)	-0.353 (0.00)	-0.754 (0.00)	-0.475 (0.00)	0.130 (0.27)
Random Forest	**0.527** (0.93)	-0.113 (0.01)	-0.494 (0.00)	-0.761 (0.00)	-0.697 (0.00)	**0.151** (0.26)
MFC	0.526 (0.93)	-0.172 (0.00)	-0.424 (0.00)	-0.738 (0.00)	-0.685 (0.00)	0.119 (0.29)
Stacking	0.495 (0.69)	-0.319 (0.00)	-0.576 (0.29)	-0.637 (0.00)	-0.420 (0.00)	0.037 (0.72)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance conditional on the sentiment index. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.37% (−)	**0.04%** (−)	**0.05%** (−)	**0.20%** (−)	**0.13%** (−)	-0.33% (−)
Realized Variance	0.26% (0.46)	-0.24% (0.01)	-0.61% (0.00)	-0.64% (0.00)	-0.58% (0.00)	-0.11% (0.34)
GARCH	0.37% (0.98)	-0.38% (0.00)	-0.56% (0.00)	-0.47% (0.00)	-0.44% (0.00)	-0.12% (0.26)
HAR-RV	0.29% (0.47)	-0.20% (0.00)	-0.51% (0.00)	-0.53% (0.00)	-0.45% (0.00)	-0.22% (0.58)
ARFIMA	-0.14% (0.04)	-0.38% (0.31)	-0.11% (0.44)	-0.12% (0.01)	-0.14% (0.03)	-0.62% (0.28)
MIDAS	0.30% (0.48)	-0.15% (0.01)	-0.45% (0.00)	-0.59% (0.00)	-0.33% (0.00)	-0.12% (0.34)
Random Forest	**0.38%** (0.93)	-0.22% (0.00)	-0.58% (0.00)	-0.59% (0.00)	-0.46% (0.00)	**-0.09%** (0.27)
MFC	0.38% (0.92)	-0.27% (0.00)	-0.52% (0.00)	-0.58% (0.00)	-0.45% (0.00)	-0.13% (0.31)
Stacking	0.34% (0.70)	-0.40% (0.00)	-0.65% (0.00)	-0.51% (0.00)	-0.30% (0.00)	-0.24% (0.70)

Table 26
Gross performance metrics for VMP using forecasted realized variance, conditional on the sentiment index.

We report gross Sharpe ratios and the gross certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance, where voltatility timing takes place only when the sentiment index is above its median value in an expanding window. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 12 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance conditional on the sentiment index. Only gross sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.521 (−)	0.254 (−)	**0.294** (−)	0.520 (−)	**0.488** (−)	0.484 (−)
Realized Variance	0.502 (0.87)	0.316 (0.58)	0.147 (0.47)	**0.725** (0.25)	0.257 (0.08)	**0.991** (0.00)
GARCH	0.535 (0.89)	0.057 (0.02)	0.205 (0.63)	0.649 (0.36)	0.197 (0.03)	0.776 (0.06)
HAR-RV	0.496 (0.77)	0.273 (0.81)	0.183 (0.52)	0.717 (0.22)	0.290 (0.14)	0.789 (0.10)
ARFIMA	0.162 (0.33)	-0.130 (0.28)	0.195 (0.69)	0.148 (0.10)	0.109 (0.11)	-0.146 (0.00)
MIDAS	0.515 (0.93)	**0.329** (0.37)	0.247 (0.73)	0.622 (0.46)	0.478 (0.94)	0.847 (0.05)
Random Forest	**0.555** (0.71)	0.269 (0.87)	0.172 (0.40)	0.688 (0.22)	0.331 (0.19)	0.960 (0.01)
MFC	0.550 (0.71)	0.206 (0.47)	0.193 (0.54)	0.670 (0.27)	0.342 (0.24)	0.862 (0.03)
Stacking	0.503 (0.76)	0.011 (0.03)	0.225 (0.74)	0.507 (0.92)	0.416 (0.55)	0.754 (0.17)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance conditional on the sentiment index. Investor risk aversion is set equal to 3. Only gross certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.37% (−)	0.09% (−)	**0.12%** (−)	0.27% (−)	**0.22%** (−)	0.32% (−)
Realized Variance	0.35% (0.87)	0.14% (0.57)	-0.01% (0.50)	**0.41%** (0.27)	0.09% (0.12)	**0.96%** (0.01)
GARCH	0.39% (0.87)	-0.08% (0.03)	0.04% (0.63)	0.35% (0.41)	0.05% (0.05)	0.69% (0.05)
HAR-RV	0.34% (0.76)	0.10% (0.80)	0.02% (0.52)	0.40% (0.32)	0.11% (0.19)	0.71% (0.06)
ARFIMA	-0.09% (0.07)	-0.24% (0.40)	0.03% (0.66)	0.02% (0.06)	0.00% (0.09)	-0.47% (0.00)
MIDAS	0.37% (0.93)	**0.15%** (0.38)	0.08% (0.71)	0.34% (0.48)	0.21% (0.95)	0.78% (0.04)
Random Forest	**0.42%** (0.72)	0.10% (0.86)	0.01% (0.39)	0.38% (0.31)	0.13% (0.27)	0.92% (0.01)
MFC	0.41% (0.71)	0.05% (0.47)	0.03% (0.53)	0.37% (0.28)	0.14% (0.30)	0.80% (0.02)
Stacking	0.35% (0.76)	-0.12% (0.02)	0.06% (0.74)	0.26% (0.93)	0.18% (0.60)	0.66% (0.16)

B Inference robustness check

Here we present the same results as in Section 7 and Appendix A. The only difference being a window size of 6 months used for inference on the Sharpe ratio and CER differences as opposed to 12 months. To limit the number of tables, we only present the robustness results of the net-of-costs analyses.

Table 27
Net-of-cost performance metrics for managed factors which are constructed using different scaling factors based on historic volatlity measures.

The scaling factors are: the lagged monthly realized variance, the lagged monthly realized volatility, the lagged 6-month realized volatility, the capped leverage of the lagged monthly realized variance, the lagged monthly realized variance without small-cap stocks, and the conditional sentiment scaling factor. We also report performance metric for the unmanaged factors as reference. The performance metrics are the net-of-cost Sharpe ratio and the net-of-cost certainty equivalent rate. The certainty equivalent rate is calculated with an investor risk aversion of 3. The performance metrics are calculated for the period 1976-01 to 2022-06 to cover the same sample period as that of the forecasted volatility models. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 6 months. P-values that are less than 0.05 are underlined.

Panel A: Net-of-cost Sharpe ratios for managed factors which are constructed using different scaling factors based on historic volatlity measures. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.519** (–)	**0.202** (–)	**0.220** (–)	**0.420** (–)	**0.337** (–)	-0.039 (–)
Realized Variance	0.424 (0.52)	-0.369 (0.00)	-0.615 (0.00)	-0.962 (0.00)	-1.107 (0.00)	-0.001 (0.87)
Realized Volatility	0.512 (0.94)	-0.259 (0.00)	-0.483 (0.00)	-0.798 (0.00)	-0.825 (0.00)	0.060 (0.45)
6m Realized Volatility	0.513 (0.94)	-0.134 (0.00)	-0.379 (0.00)	-0.685 (0.00)	-0.589 (0.00)	**0.157** (0.16)
Capped Leverage	0.480 (0.70)	-0.204 (0.00)	-0.522 (0.00)	-0.786 (0.00)	-0.930 (0.00)	-0.053 (0.94)
No small-cap	0.439 (0.61)	-0.676 (0.00)	-0.389 (0.02)	-0.561 (0.00)	-0.666 (0.00)	-0.085 (0.79)
Conditional Sentiment	0.515 (0.95)	-0.220 (0.00)	-0.458 (0.00)	-0.736 (0.00)	-0.703 (0.00)	0.040 (0.55)

Panel B: Net-of-cost certainty equivalent rates for managed factors which are constructed using different scaling factors based on historic volatlity measures. Investor risk aversion is set equal to 3. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.37%** (–)	**0.04%** (–)	**0.05%** (–)	**0.20%** (–)	**0.13%** (–)	-0.33% (–)
Realized Variance	0.25% (0.51)	-0.44% (0.00)	-0.69% (0.00)	-0.73% (0.00)	-0.69% (0.00)	-0.28% (0.84)
Realized Volatility	0.36% (0.94)	-0.35% (0.00)	-0.57% (0.00)	-0.62% (0.00)	-0.53% (0.00)	-0.21% (0.51)
6m Realized Volatility	0.36% (0.95)	-0.24% (0.00)	-0.48% (0.00)	-0.54% (0.00)	-0.40% (0.00)	**-0.09%** (0.20)
Capped Leverage	0.32% (0.70)	-0.30% (0.00)	-0.60% (0.00)	-0.61% (0.00)	-0.59% (0.00)	-0.35% (0.92)
No small-cap	0.27% (0.60)	-1.13% (0.00)	-0.22% (0.10)	-0.21% (0.00)	-0.28% (0.00)	-0.14% (0.36)
Conditional Sentiment	0.37% (0.95)	-0.31% (0.00)	-0.55% (0.00)	-0.57% (0.00)	-0.46% (0.00)	-0.23% (0.52)

Table 28
Net-of-cost performance metrics for VMP using forecasted realized variance.

We report net-of-cost Sharpe ratios and the net-of-cost certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 6 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.519** (–)	**0.202** (–)	**0.220** (–)	**0.420** (–)	**0.337** (–)	-0.039 (–)
Realized Variance	0.424 (0.51)	-0.369 (0.00)	-0.615 (0.00)	-0.962 (0.00)	-1.107 (0.00)	-0.001 (0.88)
GARCH	0.451 (0.59)	-0.244 (0.00)	-0.519 (0.00)	-0.693 (0.00)	-0.826 (0.00)	0.155 (0.25)
HAR-RV	0.473 (0.68)	-0.128 (0.02)	-0.460 (0.00)	-0.762 (0.00)	-0.749 (0.00)	0.110 (0.37)
ARFIMA	0.121 (0.38)	-0.246 (0.07)	0.043 (0.50)	-0.091 (0.04)	-0.190 (0.11)	-0.210 (0.23)
MIDAS	0.401 (0.23)	-0.156 (0.00)	-0.407 (0.00)	-0.780 (0.00)	-0.635 (0.00)	0.164 (0.20)
Random Forest	0.467 (0.62)	-0.213 (0.02)	-0.542 (0.00)	-0.838 (0.00)	-0.812 (0.00)	0.107 (0.45)
MFC	0.495 (0.82)	-0.207 (0.00)	-0.470 (0.00)	-0.755 (0.00)	-0.809 (0.00)	0.151 (0.22)
Stacking	0.483 (0.66)	-0.097 (0.00)	-0.323 (0.15)	-0.700 (0.00)	-0.562 (0.00)	**0.177** (0.21)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.37%** (–)	**0.04%** (–)	**0.05%** (–)	**0.20%** (–)	**0.13%** (–)	-0.33% (–)
Realized Variance	0.25% (0.51)	-0.44% (0.00)	-0.69% (0.00)	-0.73% (0.00)	-0.69% (0.00)	-0.28% (0.85)
GARCH	0.28% (0.58)	-0.33% (0.00)	-0.60% (0.00)	-0.55% (0.00)	-0.53% (0.00)	-0.09% (0.26)
HAR-RV	0.31% (0.67)	-0.24% (0.00)	-0.55% (0.00)	-0.59% (0.00)	-0.49% (0.00)	-0.14% (0.39)
ARFIMA	-0.14% (0.03)	-0.33% (0.18)	-0.10% (0.42)	-0.14% (0.01)	-0.17% (0.01)	-0.55% (0.25)
MIDAS	0.22% (0.23)	-0.26% (0.00)	-0.50% (0.00)	-0.60% (0.00)	-0.42% (0.00)	-0.08% (0.23)
Random Forest	0.30% (0.63)	-0.31% (0.00)	-0.62% (0.00)	-0.64% (0.00)	-0.52% (0.00)	-0.15% (0.43)
MFC	0.34% (0.82)	-0.30% (0.00)	-0.56% (0.00)	-0.59% (0.00)	-0.52% (0.00)	-0.09% (0.27)
Stacking	0.32% (0.64)	-0.21% (0.00)	-0.43% (0.26)	-0.55% (0.00)	-0.38% (0.00)	**-0.06%** (0.23)

Table 29
Net-of-cost performance metrics for VMP using forecasted realized volatility.

We report net-of-cost Sharpe ratios and the net-of-cost certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized volatility. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 6 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized volatility. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.519 (–)	**0.202** (–)	**0.220** (–)	**0.420** (–)	**0.337** (–)	-0.039 (–)
Realized Volatility	0.512 (0.94)	-0.259 (<u>0.00</u>)	-0.483 (<u>0.00</u>)	-0.798 (<u>0.00</u>)	-0.825 (<u>0.00</u>)	0.060 (0.45)
GARCH	0.505 (0.83)	-0.174 (<u>0.00</u>)	-0.417 (<u>0.00</u>)	-0.609 (<u>0.00</u>)	-0.665 (<u>0.00</u>)	0.081 (0.21)
HAR-RV	0.521 (0.97)	-0.115 (<u>0.00</u>)	-0.377 (<u>0.00</u>)	-0.676 (<u>0.00</u>)	-0.627 (<u>0.00</u>)	0.088 (0.29)
ARFIMA	0.427 (0.33)	-0.308 (<u>0.01</u>)	-0.215 (0.20)	-0.254 (<u>0.01</u>)	-0.371 (0.11)	-0.156 (0.34)
MIDAS	0.471 (0.46)	-0.123 (<u>0.00</u>)	-0.343 (<u>0.00</u>)	-0.659 (<u>0.00</u>)	-0.553 (<u>0.00</u>)	0.102 (0.15)
Random Forest	0.507 (0.85)	-0.165 (<u>0.00</u>)	-0.417 (<u>0.00</u>)	-0.673 (<u>0.00</u>)	-0.635 (<u>0.00</u>)	0.082 (0.27)
MFC	**0.528** (0.89)	-0.143 (<u>0.00</u>)	-0.379 (<u>0.00</u>)	-0.698 (<u>0.00</u>)	-0.622 (<u>0.00</u>)	0.105 (0.18)
Stacking	0.507 (0.80)	-0.088 (<u>0.00</u>)	-0.479 (<u>0.00</u>)	-0.562 (<u>0.00</u>)	-0.502 (<u>0.00</u>)	**0.124** (0.18)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized volatility. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.37% (–)	**0.04%** (–)	**0.05%** (–)	**0.20%** (–)	**0.13%** (–)	-0.33% (–)
Realized Volatility	0.36% (0.93)	-0.35% (<u>0.00</u>)	-0.57% (<u>0.00</u>)	-0.62% (<u>0.00</u>)	-0.53% (<u>0.00</u>)	-0.21% (0.48)
GARCH	0.35% (0.82)	-0.27% (<u>0.00</u>)	-0.51% (<u>0.00</u>)	-0.49% (<u>0.00</u>)	-0.44% (<u>0.00</u>)	-0.18% (0.29)
HAR-RV	0.37% (0.97)	-0.22% (<u>0.00</u>)	-0.48% (<u>0.00</u>)	-0.53% (<u>0.00</u>)	-0.42% (<u>0.00</u>)	-0.17% (0.38)
ARFIMA	0.25% (0.29)	-0.39% (<u>0.04</u>)	-0.33% (<u>0.00</u>)	-0.25% (<u>0.00</u>)	-0.27% (<u>0.00</u>)	-0.48% (0.35)
MIDAS	0.31% (0.44)	-0.23% (<u>0.00</u>)	-0.45% (<u>0.00</u>)	-0.52% (<u>0.00</u>)	-0.38% (<u>0.00</u>)	-0.15% (0.27)
Random Forest	0.36% (0.85)	-0.27% (<u>0.00</u>)	-0.51% (<u>0.00</u>)	-0.53% (<u>0.00</u>)	-0.42% (<u>0.00</u>)	-0.18% (0.37)
MFC	**0.38%** (0.89)	-0.25% (<u>0.00</u>)	-0.48% (<u>0.00</u>)	-0.55% (<u>0.00</u>)	-0.42% (<u>0.00</u>)	-0.15% (0.30)
Stacking	0.35% (0.80)	-0.20% (<u>0.00</u>)	-0.57% (<u>0.00</u>)	-0.46% (<u>0.00</u>)	-0.35% (<u>0.00</u>)	**-0.13%** (0.27)

Table 30
Net-of-cost performance metrics for VMP using forecasted realized variance, where the leverage has been capped by 1.5.

We report net-of-cost Sharpe ratios and the net-of-cost certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance, where the leverage has been capped by 1.5. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 6 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance with capped leverage of 1.5. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.519** (−)	**0.202** (−)	**0.220** (−)	**0.420** (−)	**0.337** (−)	-0.039 (−)
Realized Variance	0.405 (0.01)	-0.188 (0.00)	-0.252 (0.00)	-0.271 (0.01)	-0.374 (0.00)	-0.132 (0.13)
GARCH	0.474 (0.33)	-0.022 (0.00)	-0.262 (0.00)	-0.248 (0.01)	-0.230 (0.00)	-0.108 (0.22)
HAR-RV	0.441 (0.09)	-0.103 (0.00)	-0.196 (0.00)	-0.190 (0.01)	-0.259 (0.00)	-0.102 (0.36)
ARFIMA	0.514 (0.38)	-0.079 (0.00)	-0.241 (0.00)	-0.364 (0.00)	-0.412 (0.00)	**-0.021** (0.33)
MIDAS	0.464 (0.20)	-0.108 (0.00)	-0.219 (0.00)	-0.134 (0.01)	-0.262 (0.00)	-0.136 (0.21)
Random Forest	0.427 (0.06)	-0.104 (0.00)	-0.176 (0.00)	-0.177 (0.01)	-0.259 (0.00)	-0.107 (0.32)
MFC	0.463 (0.27)	-0.050 (0.00)	-0.196 (0.00)	-0.209 (0.00)	-0.224 (0.00)	-0.107 (0.35)
Stacking	0.489 (0.55)	-0.038 (0.00)	-0.225 (0.00)	-0.107 (0.02)	-0.204 (0.00)	-0.128 (0.31)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance with capped leverage of 1.5. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.37%** (−)	**0.04%** (−)	**0.05%** (−)	**0.20%** (−)	**0.13%** (−)	-0.33% (−)
Realized Variance	0.22% (0.01)	-0.29% (0.00)	-0.36% (0.00)	-0.26% (0.00)	-0.27% (0.00)	-0.45% (0.05)
GARCH	0.31% (0.32)	-0.15% (0.00)	-0.37% (0.00)	-0.25% (0.00)	-0.19% (0.00)	-0.42% (0.16)
HAR-RV	0.27% (0.09)	-0.21% (0.00)	-0.32% (0.00)	-0.21% (0.00)	-0.21% (0.00)	-0.41% (0.23)
ARFIMA	0.36% (0.39)	-0.19% (0.00)	-0.36% (0.00)	-0.32% (0.00)	-0.30% (0.00)	**-0.31%** (0.42)
MIDAS	0.30% (0.20)	-0.22% (0.00)	-0.34% (0.00)	-0.17% (0.00)	-0.21% (0.00)	-0.45% (0.14)
Random Forest	0.25% (0.06)	-0.21% (0.00)	-0.30% (0.00)	-0.20% (0.00)	-0.21% (0.00)	-0.42% (0.24)
MFC	0.30% (0.27)	-0.17% (0.00)	-0.31% (0.00)	-0.22% (0.00)	-0.19% (0.00)	-0.42% (0.28)
Stacking	0.33% (0.55)	-0.16% (0.00)	-0.34% (0.00)	-0.15% (0.00)	-0.18% (0.00)	-0.44% (0.20)

Table 31
Net-of-cost performance metrics for VMP using forecasted realized variance, where small cap stocks are excluded from portfolio formation.

We report net-of-cost Sharpe ratios and the net-of-cost certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance, where small cap stocks are excluded from portfolio formation. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 6 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance with no small cap stocks. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.519** (−)	**0.202** (−)	**0.220** (−)	**0.420** (−)	**0.337** (−)	-0.039 (−)
Realized Variance	0.439 (0.61)	-0.676 (0.00)	-0.389 (0.02)	-0.561 (0.00)	-0.666 (0.00)	-0.085 (0.78)
GARCH	0.471 (0.69)	-0.691 (0.00)	-0.393 (0.00)	-0.283 (0.02)	-0.797 (0.00)	0.064 (0.56)
HAR-RV	0.486 (0.78)	-0.651 (0.00)	-0.466 (0.00)	-0.536 (0.00)	-0.679 (0.00)	-0.017 (0.89)
ARFIMA	0.129 (0.37)	-0.299 (0.02)	0.075 (0.63)	-0.161 (0.02)	-0.021 (0.13)	-0.227 (0.21)
MIDAS	0.412 (0.28)	-0.684 (0.00)	-0.523 (0.00)	-0.545 (0.00)	-0.590 (0.00)	**0.069** (0.48)
Random Forest	0.480 (0.73)	-0.656 (0.00)	-0.541 (0.00)	-0.548 (0.00)	-0.689 (0.00)	-0.005 (0.85)
MFC	0.508 (0.92)	-0.700 (0.00)	-0.504 (0.00)	-0.504 (0.00)	-0.736 (0.00)	0.017 (0.73)
Stacking	0.492 (0.75)	-0.697 (0.00)	-0.704 (0.01)	-0.503 (0.00)	-0.582 (0.00)	0.048 (0.59)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance with no small cap stocks. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	**0.37%** (−)	**0.04%** (−)	**0.05%** (−)	**0.20%** (−)	**0.13%** (−)	-0.33% (−)
Realized Variance	0.27% (0.59)	-1.13% (0.00)	-0.22% (0.09)	-0.21% (0.00)	-0.28% (0.00)	-0.14% (0.35)
GARCH	0.31% (0.69)	-1.15% (0.00)	-0.22% (0.07)	-0.11% (0.01)	-0.33% (0.00)	-0.04% (0.17)
HAR-RV	0.33% (0.76)	-1.10% (0.00)	-0.25% (0.04)	-0.20% (0.00)	-0.29% (0.00)	-0.09% (0.27)
ARFIMA	-0.13% (0.04)	-0.66% (0.12)	-0.00% (0.77)	-0.07% (0.04)	-0.03% (0.09)	-0.23% (0.50)
MIDAS	0.23% (0.28)	-1.14% (0.00)	-0.28% (0.01)	-0.20% (0.00)	-0.25% (0.00)	**-0.03%** (0.15)
Random Forest	0.32% (0.72)	-1.10% (0.00)	-0.29% (0.02)	-0.20% (0.00)	-0.29% (0.00)	-0.08% (0.24)
MFC	0.35% (0.91)	-1.16% (0.00)	-0.27% (0.02)	-0.19% (0.00)	-0.31% (0.00)	-0.07% (0.21)
Stacking	0.33% (0.73)	-1.16% (0.00)	-0.36% (0.03)	-0.19% (0.00)	-0.25% (0.00)	-0.05% (0.18)

Table 32
Net-of-cost performance metrics for VMP using forecasted realized variance, conditional on the sentiment index.

We report net-of-cost Sharpe ratios and the net-of-cost certainty equivalent rates for volatility-managed factors which are constructed using different scaling factors based on forecasted realized variance, where voltatility timing takes place only when the sentiment index is above its median value in an expanding window. The certainty equivalent rate is calculated with an investor risk aversion of 3. All performance metrics are calculated over the period from 1976-01 to 2022-06 so that all scaling methods have the same number of observations. Performance metrics for the unmanaged factors and the managed factors based on past realized variance are reported as reference. The p-values for the null hypothesis of equal performance metrics between the unmanaged and managed factors are reported in parentheses. The p-values are calculated using a window size for the circular block bootstrap of 6 months. P-values that are less than 0.05 are underlined.

Panel A: Sharpe ratios for managed factors which are constructed using different forecasting models for realized variance conditional on the sentiment index. Only net-of-cost sharpe ratios are reported. The highest Sharpe ratio per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.519 (−)	**0.202** (−)	**0.220** (−)	**0.420** (−)	**0.337** (−)	-0.039 (−)
Realized Variance	0.431 (0.47)	-0.131 (0.03)	-0.528 (0.00)	-0.826 (0.00)	-0.911 (0.00)	0.136 (0.34)
GARCH	0.516 (0.98)	-0.303 (0.00)	-0.472 (0.00)	-0.573 (0.00)	-0.667 (0.00)	0.133 (0.26)
HAR-RV	0.460 (0.51)	-0.083 (0.00)	-0.416 (0.00)	-0.666 (0.00)	-0.678 (0.00)	0.052 (0.56)
ARFIMA	0.122 (0.38)	-0.300 (0.35)	0.040 (0.51)	-0.060 (0.05)	-0.143 (0.12)	-0.272 (0.06)
MIDAS	0.465 (0.50)	-0.027 (0.03)	-0.353 (0.00)	-0.754 (0.00)	-0.475 (0.00)	0.130 (0.26)
Random Forest	**0.527** (0.92)	-0.113 (0.01)	-0.494 (0.00)	-0.761 (0.00)	-0.697 (0.00)	**0.151** (0.24)
MFC	0.526 (0.93)	-0.172 (0.00)	-0.424 (0.00)	-0.738 (0.00)	-0.685 (0.00)	0.119 (0.27)
Stacking	0.495 (0.70)	-0.319 (0.00)	-0.576 (0.31)	-0.637 (0.00)	-0.420 (0.00)	0.037 (0.70)

Panel B: Certainty equivalents for managed factors which are constructed using different forecasting models for realized variance conditional on the sentiment index. Investor risk aversion is set equal to 3. Only net-of-cost certainty equivalents are reported. The highest CER per factor is highlighted in bold.

	MKT	SMB	HML	RMW	CMA	MOM
Unmanaged	0.37% (−)	**0.04%** (−)	**0.05%** (−)	**0.20%** (−)	**0.13%** (−)	-0.33% (−)
Realized Variance	0.26% (0.46)	-0.24% (0.00)	-0.61% (0.00)	-0.64% (0.00)	-0.58% (0.00)	-0.11% (0.33)
GARCH	0.37% (0.98)	-0.38% (0.00)	-0.56% (0.00)	-0.47% (0.00)	-0.44% (0.00)	-0.12% (0.25)
HAR-RV	0.29% (0.52)	-0.20% (0.00)	-0.51% (0.00)	-0.53% (0.00)	-0.45% (0.00)	-0.22% (0.58)
ARFIMA	-0.14% (0.04)	-0.38% (0.30)	-0.11% (0.39)	-0.12% (0.01)	-0.14% (0.02)	-0.62% (0.28)
MIDAS	0.30% (0.50)	-0.15% (0.01)	-0.45% (0.00)	-0.59% (0.00)	-0.33% (0.00)	-0.12% (0.33)
Random Forest	**0.38%** (0.92)	-0.22% (0.00)	-0.58% (0.00)	-0.59% (0.00)	-0.46% (0.00)	**-0.09%** (0.24)
MFC	0.38% (0.93)	-0.27% (0.00)	-0.52% (0.00)	-0.58% (0.00)	-0.45% (0.00)	-0.13% (0.30)
Stacking	0.34% (0.70)	-0.40% (0.00)	-0.65% (0.00)	-0.51% (0.00)	-0.30% (0.00)	-0.24% (0.70)

C Forecast accuracy

To quantify the forecast performance, we look at a commonly used error metric: Mean Squared Forecast Error (MSFE). The MSFE is defined as:

$$\text{MSFE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2, \tag{57}$$

where $\hat{y}_i$ is the predicted value, y_i is the actual value, and n is the number of samples.

Table 33
Forecast accuracy metrics for the GARCH models.

Both a daily and monthly GARCH model are estimated using a rolling window with window length of 5 years. The daily forecasts are rescaled to a monthly frequency by multiplying by 22. Mean squared forecast errors are calculated for each factor time series and each model. For each row, the lowest MSFE is highlighted.

	Daily	Monthly
MKT	**4.9E-06**	2.2E-05
SMB	**1.1E-07**	2.7E-06
HML	**2.8E-07**	1.8E-06
RMW	**3.8E-08**	2.6E-06
CMA	**3.6E-08**	1.8E-07
MOM	**9.4E-07**	1.5E-05

Table 34
Forecast accuracy metrics for the HAR-RV models.

Both a daily and monthly HAR-RV model are estimated using a rolling window with window length of 5 years. The daily forecasts are rescaled to a monthly frequency by multiplying by 22. Mean squared forecast errors are calculated for each factor time series and each model. For each row, the lowest MSFE is highlighted.

	Daily	Monthly
MKT	**4.4E-03**	5.0E-03
SMB	1.1E-03	**1.1E-03**
HML	1.2E-03	**1.2E-03**
RMW	**5.2E-04**	6.0E-04
CMA	3.9E-04	**3.8E-04**
MOM	**2.7E-03**	2.7E-03

Table 35
Forecast accuracy metrics for the ARFIMA models.

Both a daily and monthly ARFIMA model are estimated using a rolling window with window length of 5 years. The daily forecasts are rescaled to a monthly frequency by multiplying by 22. Mean squared forecast errors are calculated for each factor time series and each model. For each row, the lowest MSFE is highlighted.

Panel A: Daily ARFIMA(1,d,1) model MSFE for different d values.

$d =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
MKT	1.8E-05	**1.8E-05**	1.8E-05	1.8E-05	1.9E-05	2.0E-05	2.3E-05	2.4E-05	2.4E-05
SMB	**1.3E-06**	1.3E-06	1.3E-06	1.4E-06	1.5E-06	1.5E-06	1.5E-06	1.6E-06	1.6E-06
HML	**1.2E-06**	1.2E-06	1.4E-06	1.5E-06	1.5E-06	1.6E-06	2.0E-06	2.0E-06	2.1E-06
RMW	**2.3E-07**	2.4E-07	2.5E-07	2.7E-07	2.8E-07	3.0E-07	3.2E-07	3.5E-07	3.4E-07
CMA	**1.3E-07**	1.5E-07	1.5E-07	1.6E-07	1.8E-07	2.0E-07	2.2E-07	2.4E-07	2.5E-07
MOM	**6.8E-06**	7.4E-06	7.8E-06	8.6E-06	9.3E-06	1.0E-05	1.1E-05	1.2E-05	1.3E-05

Panel B: Monthly ARFIMA(1,d,1) model MSFE for different d values.

$d =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
MKT	3.2E-05	**2.5E-05**	2.6E-05	2.7E-05	2.9E-05	3.0E-05	3.1E-05	3.1E-05	3.2E-05
SMB	2.0E-06	1.9E-06	1.9E-06	1.9E-06	1.9E-06	1.9E-06	1.9E-06	1.9E-06	**1.8E-06**
HML	1.3E-06	1.3E-06	1.3E-06	**1.3E-06**	1.3E-06	1.3E-06	1.3E-06	1.3E-06	1.3E-06
RMW	3.9E-07	3.9E-07	3.9E-07	3.9E-07	3.8E-07	3.7E-07	**3.7E-07**	3.7E-07	3.8E-07
CMA	**1.4E-07**	1.4E-07	1.5E-07	1.5E-07	1.5E-07	1.6E-07	1.6E-07	1.6E-07	1.7E-07
MOM	9.1E-06	7.9E-06	7.8E-06	7.3E-06	**7.1E-06**	7.2E-06	7.9E-06	8.0E-06	8.1E-06

Table 36
Forecast accuracy metrics for the MIDAS models.

Both an exponential and beta MIDAS model are estimated using a rolling window with window length of 5 years. Mean squared forecast errors are calculated for each factor time series and each model. For each row, the lowest MSFE is highlighted..

Panel A: Beta MIDAS model MSFE for different j_{max} values.

$j_{max} =$	20	25	30	35	40	45	50	55	60
MKT	3.1E-05	**3.1E-05**	3.1E-05	3.2E-05	3.1E-05	3.1E-05	3.2E-05	3.1E-05	3.1E-05
SMB	1.5E-06	1.6E-06	1.5E-06	1.6E-06	1.5E-06	1.5E-06	**1.3E-06**	1.5E-06	1.3E-06
HML	1.3E-06	1.4E-06	1.3E-06	**1.3E-06**	1.3E-06	1.4E-06	1.3E-06	1.3E-06	1.3E-06
RMW	3.3E-07	3.2E-07	3.2E-07	3.3E-07	3.2E-07	**3.2E-07**	3.2E-07	3.2E-07	3.2E-07
CMA	1.5E-07	1.4E-07	1.2E-07	**1.2E-07**	1.2E-07	1.2E-07	1.3E-07	1.3E-07	1.3E-07
MOM	8.3E-06	**8.1E-06**	8.4E-06	8.3E-06	8.1E-06	8.4E-06	8.5E-06	8.5E-06	8.6E-06

Panel B: Exponential MIDAS model MSFE for different j_{max} values.

$j_{max} =$	20	25	30	35	40	45	50	55	60
MKT	**2.8E-05**	3.1E-05	3.1E-05	3.1E-05	3.1E-05	3.1E-05	3.1E-05	3.0E-05	3.1E-05
SMB	1.4E-06	1.3E-06	**1.3E-06**	1.3E-06	1.3E-06	1.4E-06	1.4E-06	1.3E-06	1.3E-06
HML	1.4E-06	1.3E-06	1.4E-06	1.4E-06	1.4E-06	1.3E-06	**1.3E-06**	1.4E-06	1.5E-06
RMW	**3.4E-07**	3.6E-07	3.5E-07	3.5E-07	3.5E-07	3.5E-07	3.5E-07	3.5E-07	3.5E-07
CMA	1.6E-07	1.5E-07	**1.4E-07**	1.5E-07	1.6E-07	1.5E-07	1.5E-07	1.8E-07	1.4E-07
MOM	8.3E-06	8.7E-06	**8.2E-06**	8.3E-06	8.2E-06	8.2E-06	8.2E-06	8.4E-06	8.4E-06

Table 37
Forecast accuracy metrics for the Stacking models.

All models are estimated using a rolling window with window length of 5 years. Mean squared forecast errors are calculated for each factor time series and each model. For each row, the lowest MSFE is highlighted.

	Ols	Ridge	Lasso	Elastic net
MKT	3.0E-05	1.0E-04	**2.7E-05**	2.7E-05
SMB	1.4E-06	2.1E-06	**1.4E-06**	1.4E-06
HML	**2.3E-06**	2.7E-06	2.6E-06	2.4E-06
RMW	6.4E-07	5.0E-07	5.1E-07	**5.0E-07**
CMA	**1.7E-07**	1.7E-07	2.5E-07	2.5E-07
MOM	**1.0E-05**	1.1E-05	1.3E-05	1.3E-05

Table 38
Forecast accuracy metrics for the different optimal models.

The lowest MSFE per row is highlighted.

	MKT	SMB	HML	RMW	CMA	MOM
Realized Variance	2.8E-05	1.7E-06	1.2E-06	2.8E-07	1.6E-07	7.3E-06
GARCH	2.5E-05	2.6E-06	1.2E-06	2.2E-06	1.9E-07	1.8E-05
HAR-RV	1.9E-05	1.3E-06	1.3E-06	3.7E-07	1.4E-07	7.2E-06
ARFIMA	**1.8E-05**	1.3E-06	1.3E-06	3.1E-07	2.4E-07	1.2E-05
MIDAS	3.1E-05	1.3E-06	1.5E-06	3.2E-07	1.5E-07	8.2E-06
Random Forest	2.5E-05	1.2E-06	1.4E-06	2.9E-07	**1.4E-07**	**6.3E-06**
MFC	1.9E-05	**1.2E-06**	**1.1E-06**	**2.3E-07**	1.4E-07	6.6E-06
Stacking	2.7E-05	1.4E-06	2.7E-06	5.1E-07	2.5E-07	1.0E-05

D Hasbrouck spread estimator

Roll (1984) models stock prices as:

$$p_t = m_t + cq_t, \text{ with} \tag{58}$$

$$m_t = m_{t-1} + u_t, \tag{59}$$

where m_t is the "true" price of the asset, p_t is the observed trade price, q_t takes value 1 if an ask order is filled and -1 if a bid order is filled, u_t is a random innovation modelling public information, and c is the effective cost of trading. Equations 59 and 58 imply that:

$$\Delta p_t = c\Delta q_t + u_t.$$

Taking the autocovariance of price changes:

$$\text{Cov}(\Delta p_t, \Delta p_{t+1}) = \text{Cov}(u_t + c\Delta q_t, u_{t+1} + c\Delta q_{t+1}).$$

Assuming that u_t and Δq_t are uncorrelated and that u_t is serially uncorrelated, we have:

$$\text{Cov}(\Delta p_t, \Delta p_{t+1}) = c^2 \text{Cov}(\Delta q_t, \Delta q_{t+1}).$$

Roll (1984) assumes an informationally efficient market, therefore both bid and ask move to different levels when information arrives. Leading to the bid-ask average fluctuating but the spread remaining constant. Given that both bid and ask orders are filled with equal probability, the covariance of Δq_t and Δq_{t+1} is -1. Therefore, the covariance of price changes is $-c^2$.

Since q_t can only take values 1 and -1, we have that:

$$\mathbb{P}(\Delta q_t = 2, \Delta q_{t+1} = 2) = \mathbb{P}(\Delta q_t = -2, \Delta q_{t+1} = -2) = 0, \text{ and} \tag{60}$$

$$\mathbb{P}(\Delta q_t = 2, \Delta q_{t+1} = -2) = \mathbb{P}(\Delta q_t = -2, \Delta q_{t+1} = 2) = 1/8, \text{ thus} \tag{61}$$

$$\mathbb{E}[\Delta q_t] = 0 \tag{62}$$

Where we omit the probabilities of $\Delta q_t = 0$ and $\Delta q_{t+1} = 0$ as they do not contribute to the covariance. We can calculate the covariance as follows:

$$\begin{aligned} \text{Cov}(\Delta q_t, \Delta q_{t+1}) &= \mathbb{E}[\Delta q_t \Delta q_{t+1}] - \mathbb{E}[\Delta q_t]\mathbb{E}[\Delta q_{t+1}] \\ &= \frac{1}{8} \cdot -4 + \frac{1}{8} \cdot -4 - 0 = -1. \end{aligned}$$

Therefore:

$$\text{Cov}(\Delta p_t, \Delta p_{t+1}) = -c^2.$$

Solving for c, we get:

$$c = \sqrt{-\text{Cov}(\Delta p_t, \Delta p_{t+1})}.$$

This expression allows us to estimate the effective cost of trading (c) using the covariance of consecutive price changes.

Earlier empirical studies use the sample autocovariances of daily price changes to estimate transaction costs, but, as noted by Hasbrouck (2009) and Harris (1990), such an estimation is infeasible due to the relatively high proportion of positive autocovariances between daily changes in stock prices in the data. Hasbrouck (2009) instead advocates a Bayesian approach to estimating the cost measure. He generalizes the previous equation to include a market return factor:

$$\Delta p_t = c\Delta q_t + \beta_m r_m + u_t,$$

and assumes $u_t \sim \mathcal{N}(0, \sigma^2)$. Then, given the history of price data and additional assumptions about initial values and prior distributions for the unknowns $\{c, \sigma^2, q_1, ..., q_T\}$, he sequentially draws the parameter estimates using a Gibbs sampler to characterize the posterior densities. The average of the draws for c is then used as the estimate of the effective cost of trading.

E Factor construction

We briefly describe the construction of the factors used in the volatility-managed portfolios. In our analysis we considered the five Fama-French factors (Fama & French, 2015): excess market return (MKT), size (SMB), value (HML), profitability (RMW) and investment (CMA), the momentum factor (MOM) of Jegadeesh and Titman (1993).

E.1 Market factor (MKT)

The market factor is defined as the return on a region's value-weight market portfolio minus the U.S. one month T-bill rate.

Included stocks are: All CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t, good shares and price data at the beginning of t, and good return data for t.

E.2 Size (SMB) and value (HML) factors

For the construction of the SMB and HML factors, six portfolios are formed based on Market Equity (ME) and Book-to-Market Equity (B/M) at the end of June each year. The ME breakpoint is the median NYSE market equity, and the B/M breakpoints are the 30th and 70th percentiles of B/M. Six portfolios are formed based on the intersection of the two ME and three B/M groups:

	Median ME	
	Small Value	Big Value
70th BE/ME percentile		
	Small Neutral	Big Neutral
30th BE/ME percentile		
	Small Growth	Big Growth

SMB is constructed by the equal-weight average of the returns on the three small stock portfolios minus the average of the returns on the three big stock portfolios:

$$\begin{aligned} \text{SMB} = \; & \frac{1}{3}(\text{Small Value} + \text{Small Neutral} + \text{Small Growth}) \\ & -\frac{1}{3}(\text{Big Value} + \text{Big Neutral} + \text{Big Growth}). \end{aligned} \tag{63}$$

HML is constructed by the equal-weight average of the returns for the two high B/M portfolios minus the average of the returns for the two low B/M portfolios,

$$\begin{aligned} \text{HML} = \; & \frac{1}{2}(\text{Small Value} + \text{Big Value}) \\ & -\frac{1}{2}(\text{Small Growth} + \text{Big Growth}). \end{aligned} \tag{64}$$

E.3 Profitability (RMW) factor

For the construction of the RMW factor, six portfolios are formed based on Market Equity (ME) and Operating Profitability (OP) at the end of June each year. The ME breakpoint is the median NYSE market equity, and the OP breakpoints are the 30th and 70th percentiles of OP.

Six portfolios are formed based on the intersection of the two ME and three OP groups: RMW

	Median ME	
	Small Robust	Big Robust
70th OP percentile		
	Small Neutral	Big Neutral
30th OP percentile		
	Small Weak	Big Weak

is defined as the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios:

$$\begin{aligned} \text{RMW} = \quad & \frac{1}{2}(\text{Small Robust} + \text{Big Robust}) \\ & -\frac{1}{2}(\text{Small Weak} + \text{Big Weak}). \end{aligned} \tag{65}$$

E.4 Investment (CMA) factor

For the construction of the CMA factor, six portfolios are formed based on Market Equity (ME) and Investment-to-assets ratio (IA) at the end of June each year. The ME breakpoint is the median NYSE market equity, and the IA breakpoints are the 30th and 70th percentiles of IA. Investment-to-assets ratio (IA) is the change in total assets from the fiscal year ending in year $t-2$ to the fiscal year ending in $t-1$, divided by $t-2$ total assets at the end of each June. Six portfolios are formed based on the intersection of the two ME and three IA groups: CMA is

	Median ME	
	Small Aggressive	Big Aggressive
70th IA percentile		
	Small Neutral	Big Neutral
30th IA percentile		
	Small Conservative	Big Conservative

defined as the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios:

$$\begin{aligned} \text{CMA} = \quad & \frac{1}{2}(\text{Small Conservative} + \text{Big Conservative}) \\ & -\frac{1}{2}(\text{Small Aggressive} + \text{Big Aggressive}). \end{aligned} \tag{66}$$

E.5 Momentum factor (MOM)

To construct the momentum factor we use six value-weighted portfolios formed on size and prior (2-12) returns to construct Mom. The portfolios, which are formed monthly, are the intersections of 2 portfolios formed on size (market equity, ME) and 3 portfolios formed on prior (2-12) return. The monthly size breakpoint is the median NYSE market equity. The monthly prior (2-12) return breakpoints are the 30th and 70th NYSE percentiles.

The momentum factor is the average return of the two high prior return portfolios minus the average return on the two low prior return portfolios.

	Median ME	
	Small High	Big High
70th return percentile		
	Small Neutral	Big Neutral
30th return percentile		
	Small Low	Big Low

$$\text{MOM} = \frac{1}{2}(\text{Small High} + \text{Big High}) - \frac{1}{2}(\text{Small Low} + \text{Big Low}). \tag{67}$$

The six portfolios used to construct MOM each month include NYSE, AMEX, and NASDAQ stocks with prior return data. To be included in a portfolio for month t (formed at the end of month $t-1$), a stock must have a price for the end of month $t-13$ and a good return for $t-2$. In addition, any missing returns from $t-12$ to $t-3$ must be -99.0, CRSP's code for a missing price. Each included stock also must have ME for the end of month $t-1$.