www.joim.com

Journal Of Investment Management, Vol. 23, No. 1, (2025), pp. 4–21

FORECASTING AND MANAGING VOLATILITY: AN S&P 500 CASE STUDY

Wei Dai[a], *Xing Hong*[b], *Robert C. Merton*[c,d] *and Mathieu Pellerin*[e,*]

Using daily and intraday data from 1997 to 2023, we study strategies that stabilize volatility around a target by rebalancing between the S&P 500 and Treasury bills based on a broad set of volatility forecasts. Somewhat counterintuitively, lower forecasting errors do not necessarily result in more stable strategy volatility. Simple forecasts with fewer parameters can stabilize volatility as well as more complex models. In particular, combinations of implied volatility and simple estimators based on past returns exhibit good volatility control and lower turnover. On the implementation front, we show that the target volatility strategies we study are viable in the presence of realistic trading costs, delays between forecasting and rebalancing, or constraints on rebalancing frequency. Collectively, our findings can help design target volatility strategies that improve upon portfolios with constant target weights (e.g., a 60/40 portfolio) in achieving and maintaining investors' desired volatility exposures over time.

1 Introduction

Portfolios with constant target weights (e.g., a 60/40 portfolio) pervade the asset allocations of both institutional and individual investors. The target weights of different asset classes, such as equities and fixed income, are often chosen to reflect the amount of portfolio volatility an investor is willing to take. The rationale behind maintaining stable portfolio weights may stem from the goal of maintaining stable volatility risk exposure. However, this rationale is valid only if asset risk characteristics are constant over time.

Since the volatility of assets is time-varying, portfolios with static target weights can experience drastic changes in volatility. Therefore, to maintain a desired volatility exposure, a better alternative may be a *target volatility* (TV) strategy that dynamically adjusts portfolio weights to

[a]Head of Investment Research, Dimensional Fund Advisors, Austin, TX, USA. E-mail: wei.dai@dimensional.com
[b]Senior Researcher, Dimensional Fund Advisors, Austin, TX, USA. E-mail: xing.hong@dimensional.com
[c]School of Management Distinguished Professor of Finance, MIT Sloan School of Management, Cambridge, MA, USA. E-mail: rmerton@mit.edu
[d]Resident Scientist, Dimensional Fund Advisors, Austin, TX, USA.
[e]Retirement Research Director, Dimensional Fund Advisors, Austin, TX, USA. E-mail: mathieu.pellerin@dimensional.com
*Corresponding author.

stabilize volatility around a target. Such strategies are feasible because volatility is persistent and forecastable.

TV strategies may be relevant to a broad set of investors. In principle, volatility targeting can be overlaid on a wide range of investment strategies, active and passive, to help reduce deviations between desired and realized portfolio volatility. Just as diversification across many securities can lower unnecessary risks and reduce overall portfolio volatility, volatility stabilization can lower unnecessary fluctuations around a desired level of volatility. This can lead to lower volatility of volatility, providing investors with a more effective tool to pursue their investment objectives. Volatility stabilization may also be desirable when designing indices for use in structured products, such as structured notes or fixed index annuities (FIAs). In this last context, TV strategies may reduce hedging costs for the issuing entity and result in more attractive payoffs to the end investor.

We study the effectiveness of volatility stabilization for TV strategies based on a broad set of volatility forecasts. In particular, we focus on TV strategies that rebalance daily between the S&P 500 and Treasury bills. Our main contribution is threefold. First, we show that, somewhat counterintuitively, lower forecast errors do not necessarily result in TV strategies with more stable volatility. Second, we show that simple estimators can achieve volatility control that is competitive with more complex models, reducing volatility of volatility by up to 60–70%. Third, we show that combining implied volatility, which captures market expectations of future volatility, with simple forecasts that use backward-looking information, may improve volatility control and reduce turnover. Collectively, our findings can help design TV portfolios and indices that better adhere to their target.

We also explore the practical aspects of TV strategies, including turnover and trading costs, as well as rebalancing timing and frequency. Our findings provide support for the viability of TV portfolios based on liquid assets in the presence of real-world frictions.[1]

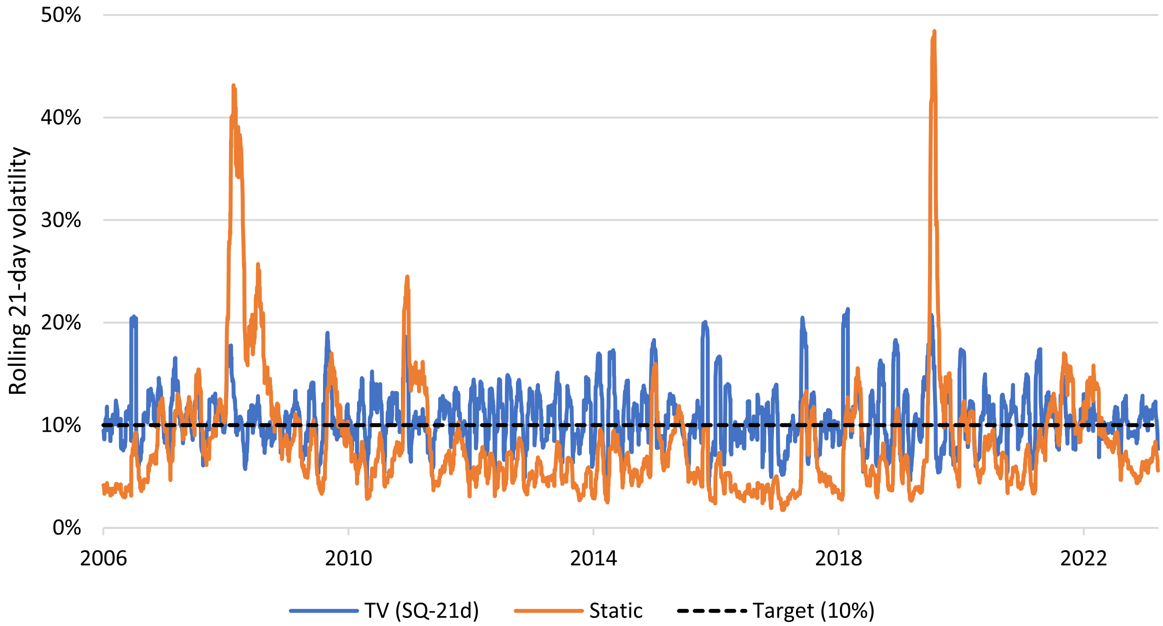
Our work connects with the vast literature on volatility estimation and forecasting, including the use of intra-day returns (Ghysels *et al.*, 2006; Corsi, 2009; Liu *et al.*, 2015; Patton and Sheppard, 2015; Bollerslev *et al.*, 2018, among many others) and implied volatility (Poon and Granger, 2003; Becker *et al.*, 2007; Busch *et al.*, 2011; Bekaert and Hoerova, 2014; Todorov and Zhang, 2022). A growing literature also examines the performance of TV portfolios (e.g. Lo, 2016; Moreira and Muir, 2017; Cederburg *et al.*, 2020), with a focus on whether TV strategies lead to utility gains or improvements in Sharpe ratios relative to conventional portfolios. Our focus is different but complementary. We seek to answer a practical question: When constructing a TV strategy, what types of volatility forecasts are likely to help the strategy adhere to its target while controlling turnover and trading costs?

Section 2 illustrates the differences between target weight and target volatility strategies through an example. Section 3 outlines our data sources and evaluation framework. Section 4 presents our three key results on forecast selection and volatility control. Section 5 explores implementation considerations, while Section 6 concludes. The Appendix contains a detailed description of our methodology and some additional results.

2 Target Weights vs. Target Volatility Strategies

Exhibit 1 shows the rolling monthly volatility of a TV strategy with a 10% volatility target and a static strategy with a fixed weight of 50% in equity and 50% in one-month Treasury bills

Exhibit 1. Rolling volatility for a 50/50 static strategy and a TV strategy with the same long-run volatility of 10%.



The data runs from August 11, 2006 to November 30, 2023. The rolling volatility for each strategy is the standard deviation of log daily returns over the rolling 21 days. Both strategies rebalance daily between the S&P 500 and one-month Treasury bills. The TV strategy targets 10% volatility and uses the average of squared daily close-to-close log returns over the past 21 trading days as the volatility forecast. See Appendix A for details about strategy construction. The static strategy has fixed weights (50% in equities and 50% in one-month Treasury bills) that yield a volatility of 10% over the sample period.

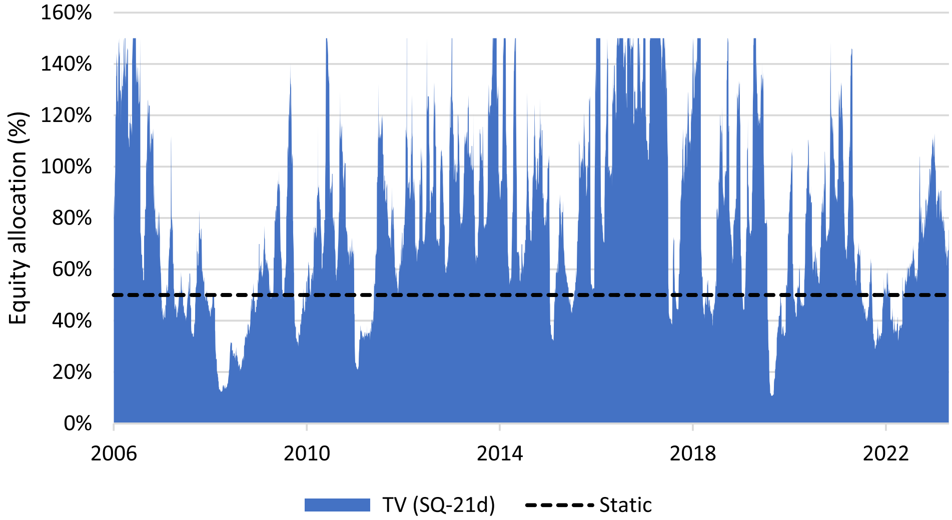
that is calibrated to have 10% volatility over the full sample period. The different objectives of both strategies (stabilizing volatility vs. stabilizing weights) lead to sharp differences in realized volatilities over time, especially around the global financial crisis (GFC) and COVID-19. The TV strategy's rolling volatility stayed relatively close to the target during these periods, while the static strategy's rolling volatility surged above 40%, more than four times its long-term value.

In addition to having more volatility spikes, the static strategy also undershoots its long-run volatility over long periods; the rolling 21-day volatility is as low as 3% in the 2010s. This is because the allocation to equities does not increase when volatility is low. By contrast, the TV strategy had higher weights allocated to equities in those periods, as seen in Exhibit 2, which substantially reduces undershooting relative to the target. Even with a target volatility of 10%, half the historical average for stocks, the TV strategy leverages up (equity allocation above 100%) in a significant proportion of the sample.

Overall, the volatility stabilization did not result in lower equity exposure or lower returns. In fact, the average equity allocation over the whole sample is 79% for the TV strategy, compared to 50% for the static strategy. The annualized compound return over the period, from August 11, 2006 to November 30, 2023, is 8.5% for the TV strategy and 6.0% for the static strategy.

The results are not predicated on complex forecasting techniques. The volatility forecast used is simply the average of squared daily, close-to-close log returns over the past month, similar to Lo (2016) and Moreira and Muir (2017). That said, TV strategies based on better forecasts

Exhibit 2. Allocation to equities for a 10% TV strategy.



The data runs from August 11, 2006 to November 30, 2023. Both strategies rebalance daily between the S&P 500 and one-month Treasury bills. The TV strategy targets 10% volatility and uses the average of squared daily close-to-close log returns over the past 21 trading days as the volatility forecast. See Appendix A for details about strategy construction. The static strategy has fixed weights (50% in equities and 50% in one-month Treasury bills) that yield a volatility of 10% over the sample period.

can meaningfully improve volatility control in our sample. Our key aim is to study the link between forecast selection and volatility control in target volatility strategies. We next present our methodology and evaluation framework.

3 Methodology Summary

We consider a variety of volatility forecasts that fall in two broad categories: standalone forecasts and forecasts based on regression models. This section briefly summarizes how we construct and evaluate these forecasts and TV strategies built upon them. Our full methodology is presented in Appendix A. The standalone forecasts based on daily returns include the equally-weighted average of squared daily close-to-close log returns (labeled SQ-D) over the past week (five trading days) or month (21 trading days), and the exponentially smoothed (ES) average with smoothing parameter $\lambda = 0.94$. We also consider standalone forecasts based on the intraday realized variance (RV), calculated using five-minute log returns from open to close, as well as the full-day RV (FRV), which adds to RV the squared overnight log return from close to next open, over 1-, 5-, and 21-day lookback windows.

Regression-based forecasts include the heterogeneous autoregressive (HAR) model of Corsi (2009), the mixed data sampling (MIDAS) model of Ghysels *et al.* (2006), and the heterogeneous exponential (HExp) model of Bollerslev *et al.* (2018). Regressions are estimated daily on an expanding window. The initial window has eight years' worth of trading days (8×252) ranging from January 12, 1998 to August 10, 2006, the second window ends on August 11, 2006, and so on. Our sample begins in the late 1990s due to the availability of intraday data, which is used to construct RV/FRV and regression-based forecasts.[2] Lastly, we use the daily close of the VIX index as a standalone predictor or averaged with other predictors. The averaging is motivated by the additional information contained in the forward-looking VIX relative to forecasts based on historical information.

Some standalone forecasts can be expected to undershoot or overshoot volatility on conceptual grounds. For example, intraday RV would systematically underestimate because it does not account for overnight price movements. On the other hand, the VIX may overstate expected volatility because it also contains a variance risk premium. Accordingly, we follow Hansen and Lunde (2005) and scale standalone forecasts by the ratio between the running average of volatility and the running average of the forecast of interest, where running averages are computed on the same expanding window as regressions and volatility is measured as either full-day RV or squared close-to-close returns. The regression-based forecasts we consider use full-day RV as the response variable, so the forecasts are naturally on the scale of full-day RV. We also present regression forecasts scaled to squared returns.

We use out-of-sample R^2 to evaluate forecasting accuracy. Out-of-sample R^2 measures the reduction in mean squared error (MSE) of a forecast of interest relative to the MSE of a naïve forecast (typically the average of past observations). The MSE is based on one-step-ahead forecasts of full-day RV over one day.

We construct target volatility portfolios that rebalance daily between one-month Treasury bills and the S&P 500 in order to stabilize volatility over time around a target volatility of 10% annualized. Given a forecast of the S&P 500's daily volatility $\hat{\sigma}_t$, the asset weights and returns of the TV strategy are given by the equations below. We apply a leverage constraint of 150% to equity weights (w_t).

$$w_t = \min\left(\frac{10\%}{\sqrt{252}\hat{\sigma}_t}, 1.5\right)$$

$$R_{t+1}^{TV} = w_t R_{t+1} + (1 - w_t) R_{t+1}^{f}$$

The main metric we use to evaluate volatility control is the root mean squared error (RMSE) of the difference between the target volatility (10%) and the rolling volatility of the strategy over a 21-day window. A lower value corresponds to better volatility control. For ease of exposition, we will sometimes refer to this measure as the volatility of volatility (vol-of-vol), even though the measure captures both vol-of-vol (deviations around the long-term volatility of the strategy) and bias (the gap between long-term volatility and target volatility).

The strategy returns run from August 11, 2006 to November 30, 2023. This period features the global financial crisis, COVID-19, heightened volatility in 2022, as well as relatively calm periods in the 2010s. Therefore, our setup may be informative about the forecasting accuracy and volatility control of different forecasts across varying volatility levels.

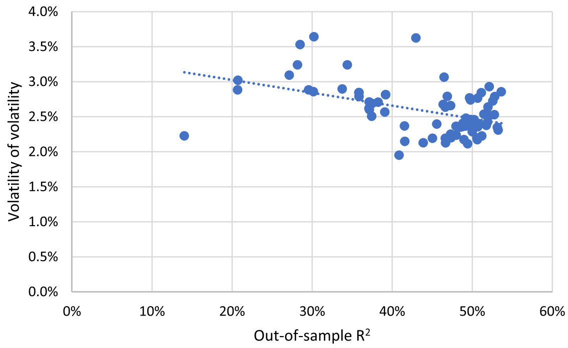
4 Forecast Selection and Volatility Control

4.1 Forecasting accuracy and volatility control

Exhibit 3 plots the volatility control measure against out-of-sample R^2 across all standalone and regression-based volatility forecasts. The overall negative correlation suggests that higher out-of-sample R^2 is associated with lower vol-of-vol, though the relation is weak. Several forecasts with similar out-of-sample R^2 lead to sharply different volatility control.

The disconnect arises because of the differing impacts of forecasting errors on out-of-sample R^2 and volatility control. Since the out-of-sample R^2 is based on the mean squared error, overestimation and underestimation of the same magnitude will have the same impact. In contrast, underestimation can have a bigger impact on the vol-of-vol. Because equity weight in the TV strategy is proportional to the inverse of the forecast ($1/x$), when the leverage constraint is not binding, for

Exhibit 3. Out-of-sample R^2 vs. vol-of-vol.



The data runs from August 11, 2006 to November 30, 2023. The charts show the out-of-sample R^2 and the vol-of-vol measure (RMSE of rolling 21-day volatility relative to target) across all standalone and regression-based volatility forecasts based on two types of scaling (scaling to full-day RV or squared daily returns). See Appendix A for more details on the volatility forecasts and evaluation measures.

the same magnitude of forecasting error, underestimation increases equity weight more than overestimation decreases it.

Consider a TV strategy with a 10% target, true equity volatility of 50%, and a forecast of 40%: The equity weight of the strategy would be 25% when it "should" have been 20%, and the volatility of the TV strategy will be $25\% * 50\% = 12.5\%$, 2.5% higher than the target of 10%. If the forecast is 60% instead, the equity weight would be 16.67% and the strategy volatility is $16.67\% * 50\% = 8.33\%$, 1.67% lower than the target. Although the magnitude of the over- and underestimation is the same (both 10%), the deviation of the TV strategy's volatility from target is larger with underestimation.

Moreover, forecasting errors of the same magnitude have a bigger impact on volatility control during calmer periods than during volatile periods. Continuing with the case of an underestimation of 10%, consider a true equity volatility of 20% instead of 50%. In this case, the volatility forecast is 10%, the equity weight is 100%, and the strategy's volatility is $100\% * 20\% = 20\%$, 10% higher than the target of 10% (vs. 2.5% higher than the target in the example above).

In Appendix B, Exhibit B.1 shows that while switching to a forecasting evaluation metric that scales forecast errors by the volatility level and penalizes underestimation more severely (QLIKE) helps to address some of these issues, there is still considerable noise in the relation between forecasting accuracy and vol-of-vol. These results provide a rationale for focusing on volatility control when selecting forecasts to build a TV strategy.

4.2 Standalone vs. regression-based forecasts

Exhibit 4 shows the vol-of-vol measure across different TV strategies. Parentheses in the labels denote the lookback window for standalone estimators and forecasting horizon for regressions.

All TV strategies achieve meaningful reduction in vol-of-vol relative to the static strategy, with reductions ranging from 40% to 70%. Focusing on the "Not averaging with VIX" columns, one main takeaway is that forecasts based on variants of five-minute RVs seem to achieve better volatility control than those that do not use intraday data, especially when scaled to squared daily returns. Interestingly, the different types of scaling seem to have a bigger effect on standalone forecasts than on regression-based forecasts.

When scaled to squared daily returns, full-day or intraday RV with short lookback windows (one and five days) are competitive with the best regression-based forecasts. Although the VIX and estimators based on squared daily returns (SQ-D and ES) perform worse than RV estimators with short lookback windows, they still reduce the vol-of-vol measure by about half.

To help interpret the numbers provided in Exhibit 4, Exhibit 5 shows the rolling volatility of

Exhibit 4. Vol-of-vol of TV strategies based on standalone and regression-based forecasts.

	Scaled to full-day RV		Scaled to squared daily returns	
	Not averaging with VIX	Averaging with VIX	Not averaging with VIX	Averaging with VIX
Static	6.1%			
Standalone				
SQ-D (5d)	3.6%	2.4%	2.9%	2.1%
SQ-D (21d)	3.5%	2.7%	3.0%	2.5%
ES	3.2%	2.7%	2.9%	2.6%
FRV (1d)	2.9%	2.3%	2.2%	2.1%
FRV (5d)	2.4%	2.3%	2.0%	2.2%
FRV (21d)	3.2%	2.7%	2.9%	2.6%
Intraday RV (1d)	3.1%	2.4%	2.4%	2.2%
Intraday RV (5d)	2.8%	2.4%	2.2%	2.2%
Intraday RV (21d)	3.6%	2.8%	3.1%	2.6%
VIX	2.8%	2.8%	2.8%	2.8%
Regression-based				
HAR (1d)	2.2%	2.4%	2.1%	2.4%
HAR (5d)	2.4%	2.5%	2.4%	2.5%
HAR (21d)	2.8%	2.7%	2.9%	2.8%
HExp (1d)	2.2%	2.4%	2.1%	2.4%
HExp (5d)	2.3%	2.5%	2.4%	2.5%
HExp (21d)	2.6%	2.6%	2.8%	2.7%
MIDAS (1d)	2.2%	2.4%	2.2%	2.4%
MIDAS (5d)	2.3%	2.5%	2.4%	2.5%
MIDAS (21d)	2.7%	2.7%	2.9%	2.8%
Mean	2.8%	2.5%	2.6%	2.5%

The data runs from August 11, 2006 to November 30, 2023. The table shows the vol-of-vol measure (root mean squared error of rolling 21-day volatility from target) for TV strategies based on all standalone and regression-based volatility forecasts using two types of scaling (scaling to full-day RV or squared daily returns) as well as the simple averages of VIX and each of the forecasts. See Appendix A for more details on the volatility forecasts and strategy construction.

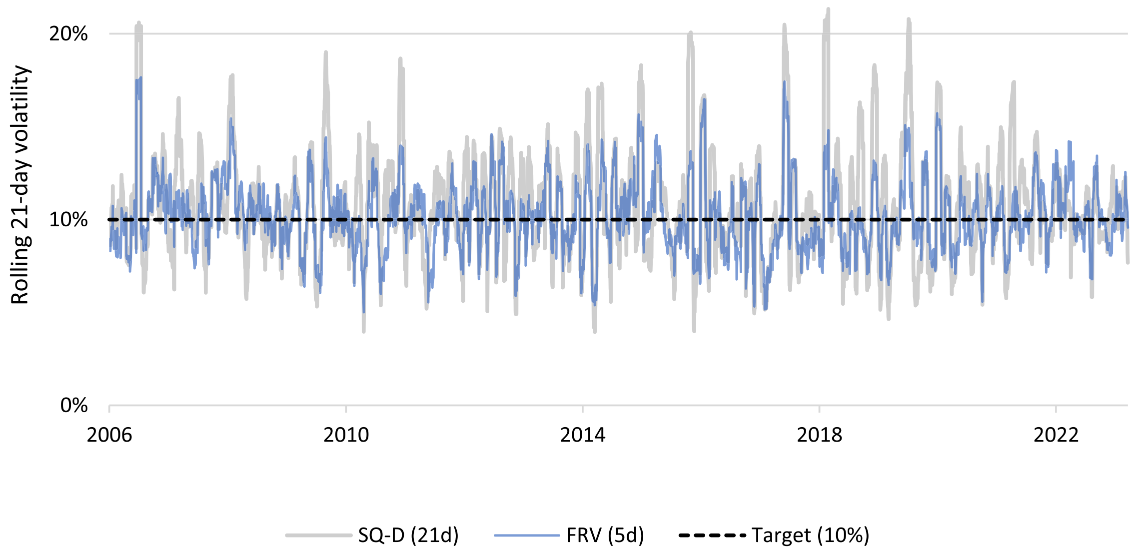
returns for TV strategies based on two forecasts: SQ-D (21d) and FRV (5d), both scaled to squared returns. The vol-of-vol values are 3.0% and 2.0%, respectively. The strategy based on FRV (5d) very rarely exhibits rolling volatility over 15%, while SQ-D (21d) reaches the 20% threshold, twice the target volatility, on several occasions. This example confirms that while the difference in vol-of-vol between TV and the static allocation is the most dramatic, differences between TV strategies based on different forecasts are significant too, hence our emphasis on the link between forecast selection and volatility control.

So far, we have not considered turnover and trading costs. Their impact is especially relevant for forecasts with very short lookback windows (such as one day) that produce attractive volatility control but may result in excessive turnover. We address the issue in Section 5.

4.3 Averaging with VIX

Theoretically, the VIX captures the expected volatility of the S&P 500 index over the next month. Since it embeds the forward-looking expectations of market participants, it should be a better predictor of S&P 500 volatility than measures that rely solely on backward-looking information. However, we find that the VIX is not the best predictor for the next-day S&P 500 volatility and does not lead to the most effective volatility control. This may be caused by the horizon discrepancy between the VIX (next month) and our forecast target (next day), the approximation errors inherent in the construction

Exhibit 5. Rolling volatility of TV strategies based on two different forecasts with the same target volatility of 10%.



The data runs from August 11, 2006, to November 30, 2023. The rolling volatility for each strategy is the standard deviation of daily returns over the rolling 21 days. Both strategies rebalance daily between the S&P 500 and one-month Treasury bills. The TV strategies target 10% volatility. The gray line, SQ-D (21d), uses the average of squared daily, close-to-close log returns over the last 21 trading days, scaled to squared returns, as the volatility forecast. The blue line, FRV (5d), uses the average of full-day RV over the last five days scaled to squared returns. See Appendix A for details about strategy construction.

of VIX, or the presence of a time-varying risk premium in implied volatility.[3]

However, combining backward- and forward-looking information about volatility may still produce better volatility forecasts than those based on past data alone. Indeed, past studies have shown that combining forecasts based on historical data with option-implied measures of market expected volatility may lead to better predictions (Poon and Granger, 2003 and references therein; Todorov and Zhang, 2022). Moreover, studies show that simple averages of forecasts often outperform more complex combinations (Diebold and Shin, 2019 and references therein). Motivated by this evidence, we explore the effect of averaging with implied volatility on volatility control for both standalone and regression-based estimators.

As shown in the "Averaging with VIX" columns of Exhibit 4, the addition of the VIX reduces vol-of-vol for most standalone forecasts, with the exception of five-day intraday and five-day full-day RV scaled to squared returns. The reduction is especially pronounced for forecasts with the lowest performance. For example, the average of SQ-D (5d), i.e., daily squared returns over the last five days, and the VIX (scaled to squared returns) yields a vol-of-vol measure of 2.1%. This value is not only meaningfully lower than either estimator taken alone (2.8% for the VIX, 2.9% for the SQ-D (5d)), but also the second lowest in the entire table. Notably, this estimator does not require the use of intraday data, which may be attractive when the use of intraday data is impractical.

For regression-based estimators, the effect of averaging with the VIX on volatility control is more muted and mixed. We see slight improvements for regression models with a forecasting horizon of 21 days but slight impairments for those with a forecasting horizon of one or five days.

5 Implementation Considerations

In this section, we explore the practical aspects of TV strategies, including their turnover, trading costs and rebalancing methodology. For brevity, we show results for a subset of nine strategies that are representative of the full set: the 1-, 5-, and 21-day averages of full-day RV, on their own and averaged with the VIX, as well as the HAR model with forecasting horizons of 1, 5, and 21 days. The VIX is scaled to full-day RV, while other estimators are naturally on that scale. This subset covers both standalone estimators and regression-based estimators. The different lookback windows and forecasting horizons allow us to explore the tradeoffs between forecast responsiveness and stability. The inclusion of VIX averaging allows us to explore how averaging forecasts may affect implementation.

5.1 Rebalancing timing and frequency

Due to implementation constraints, there may be a gap between when a volatility forecast is available and when rebalancing can be carried out. The first two columns of Exhibit 6 compare the volatility control achieved under rebalancing with the latest information vs. a forecast lagged by one day. The vol-of-vol under a one-day lag is slightly higher, but still achieves a considerable reduction relative to the static strategy. This result suggests that TV strategies are practical even when there are reasonable lags between forecasting and rebalancing.

Next, we explore the impact of rebalancing frequency on volatility control. The last two columns of Exhibit 6 show results for calendar-based approaches in which rebalancing is carried out every 5 or 21 trading days. While it is not surprising that vol-of-vol increases with decreasing rebalancing frequency, weekly rebalancing preserves a substantial fraction of the volatility control ability. The reduction in vol-of-vol relative to static is nonnegligible even at the monthly frequency, at approximately 1.5 percentage points (25% in relative terms) for HAR and combinations of FRV and VIX.

Exhibit 6. Vol-of-vol of TV strategies under different rebalancing constraints.

Rebalancing frequency	Daily	Daily	Weekly	Monthly
Rebalancing timing	No lag	One-day lag	No lag	No lag
Static	6.1%	6.1%	6.1%	5.9%
FRV (1d)	2.9%	3.3%	3.7%	6.6%
FRV (5d)	2.4%	2.7%	3.1%	5.4%
FRV (21d)	3.2%	3.5%	3.8%	5.8%
FRV (1d) + VIX	2.3%	2.5%	2.8%	4.8%
FRV (5d) + VIX	2.3%	2.5%	2.7%	4.6%
FRV (21d) + VIX	2.7%	2.9%	3.2%	4.9%
HAR (1d)	2.2%	2.4%	2.7%	4.6%
HAR (5d)	2.4%	2.6%	2.7%	4.5%
HAR (21d)	2.8%	3.0%	3.1%	4.5%

The data runs from August 11, 2006 to November 30, 2023. The table shows the vol-of-vol measure (root mean squared error of rolling 21-day volatility from target) for selected TV strategies. The static strategy has fixed weights (50% in equities and 50% in one-month Treasury bills) that yield a volatility of 10% over the sample period. The results for one-day lag use the value of the forecast at the end of the previous trading day when forecasting volatility at the end of the current trading day. Weekly and monthly rebalancing use the latest forecasting information but rebalance every 5 and 21 trading days. See Appendix A for more details on the volatility forecasts and strategy construction.

These results suggest that the concept of volatility targeting can be implemented through strategies that rebalance less frequently. In practice, we could better balance the tradeoff between rebalancing frequency and volatility control through more thoughtful rebalancing methods than the simple calendar-based approach, such as weighing the expected reduction in vol-of-vol vs. turnover or adding intra-period rebalancing when volatility changes are above a certain threshold.

5.2 *Turnover and trading costs*

Consistent with the example of Section 2, Exhibit 7 shows that TV strategies have had higher exposure to equities than the static strategy on average. All strategies experience a large range of equity exposures over the sample period, as shown by the standard deviation. The turnover figures, calculated as the daily average of the absolute change in equity exposure, are reported in the third column of Exhibit 7. Turnover is sharply higher for forecasts with short lookback windows or short forecasting horizons. Daily one-way turnover reaches 22% for the TV strategy based on yesterday's FRV and reaches as low as 1.8% for the 21-day average of FRV.

Averaging the VIX with FRV-based forecasts sharply reduces turnover. For example, averaging the VIX with the five-day average of FRV reduces daily one-way turnover from 5.9% to 3.7%. This result suggests that forecast averaging may facilitate implementation by reducing turnover. The pattern is present when averaging VIX with a wider set of estimators than the three combinations we consider here (see Exhibit 8). The reduction in daily turnover is about 2% on average and is particularly meaningful for standalone forecasts with short lookback periods and regression-based forecasts with short forecasting horizons.

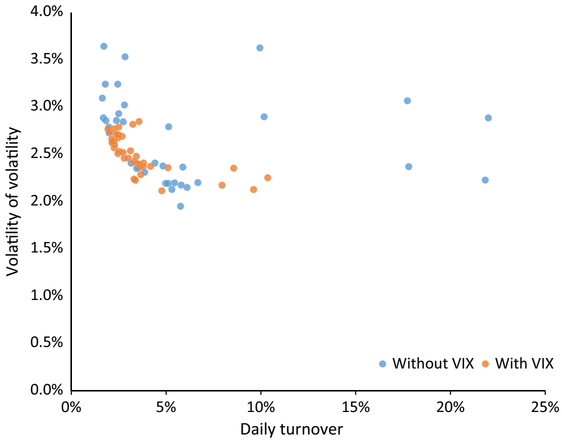
Finally, the last two columns of Exhibit 7 examine the impact of turnover on returns under conservative trading costs assumptions. We first present

Exhibit 7. Returns, turnover and trading costs for selected strategies.

	Average equity allocation	SD of equity allocation	Daily one-way turnover	Average daily return, annualized, and 95% CI	Return reduction after T-costs
Static	50%	0%	0.2%	6.37% ± 3.94%	−0.01%
FRV (1d)	91%	41%	22.0%	10.88% ± 5.10%	−0.56%
FRV (5d)	85%	37%	5.9%	9.02% ± 4.94%	−0.15%
FRV (21d)	80%	34%	1.8%	8.54% ± 4.81%	−0.05%
FRV (1d) + VIX	79%	31%	10.4%	8.64% ± 4.59%	−0.26%
FRV (5d) + VIX	77%	30%	3.7%	7.94% ± 4.62%	−0.09%
FRV (21d) + VIX	75%	28%	2.4%	7.72% ± 4.55%	−0.06%
HAR (1d)	76%	28%	6.7%	8.56% ± 4.51%	−0.17%
HAR (5d)	72%	24%	4.8%	8.22% ± 4.33%	−0.12%
HAR (21d)	66%	18%	2.7%	7.84% ± 4.11%	−0.07%

The data runs from August 11, 2006 to November 30, 2023. The average and standard deviation of equity allocation are obtained by taking the time-series mean and standard deviation of the target weight. Daily one-way turnover is calculated as the absolute difference between the equity weight at the end of the trading day and the target weight for the next trading day. Average returns are average daily, simple returns multiplied by 252. Standard errors were obtained by stationary block bootstrap (Politis and Romano, 1994) with mean block size 25 and 10,000 replicas and multiplied by 1.96 to form a 95% confidence interval. Trading costs are assumed to be 1 basis point for both S&P 500 and bills. See Appendix A for more details on volatility forecasts, strategy construction, and trading cost methodology.

Exhibit 8. Vol-of-vol vs. daily turnover.



The data runs from August 11, 2006 to November 30, 2023. The chart shows the vol-of-vol measure (root mean squared error of rolling 21-day volatility from target) and the average daily one-way turnover (average absolute difference between end-of-day equity weight immediately before rebalancing and equity weight after rebalancing) for TV strategies based on all standalone and regression-based volatility forecasts using two types of scaling (scaling to full-day RV or squared daily returns) as well as the simple averages of VIX and each of the forecast. VIX and forecasts averaged with VIX are highlighted in orange. See Appendix A for more details on the volatility forecasts and strategy construction.

average daily returns (annualized) in the absence of trading costs, along with 1.96 standard errors (95% confidence interval).[4] In our roughly 20-year period, TV strategies achieved higher average returns than static, though there is substantial uncertainty about their magnitude. In other words, we cannot say that the returns are reliably different between TV strategies and the static strategy or across TV strategies. We then show the reduction in average returns with transaction costs of 1 basis point (bp) applied to two-way turnover. As we argue in Appendix A, since the TV strategies we study could be implemented with S&P 500 futures and one-month Treasury bills, our assumed trading costs are arguably higher than those currently achievable in practice. In addition, our estimates can be considered as an upper bound because they are based on daily rebalanced strategies. The impact of different trading costs can be approximated from our results since trading costs are roughly a linear function of turnover: For instance, under trading costs of 2 bps, one would expect the reduction in average return for FRV (21d) to be around -0.10% rather than -0.05%.

The TV strategies based on one-day lookback window or forecasting horizon see the larger reduction in average returns after subtracting trading costs, with the largest reduction of -0.56% for the standalone FRV (1d). However, all other strategies exhibit average return reductions of -0.15% or less. Combined with the results in Exhibit 4, a five-day lookback or forecasting horizon seems to strike an attractive tradeoff between lowering vol-of-vol and controlling trading costs.

Collectively, our results provide evidence for the feasibility of TV strategies. The use of judicious rebalancing rules (thresholds, hold ranges, etc.), patient and flexible trading, forecasts that strike a balance between turnover and volatility control, and the selection of liquid instruments can all increase the real-world feasibility of TV strategies.

6 Conclusion

Portfolios with constant target weights are inadequate for helping investors maintain stable risk exposures. By dynamically adjusting portfolio weights, target volatility strategies offer an effective tool for investors to stabilize portfolio volatility. When designing such strategies, it is helpful to recognize the distinction between forecasting accuracy and volatility control. Our results suggest that a variety of volatility forecasts can be used to substantially reduce fluctuations in volatility relative to strategies with static weights. In particular, simple forecasts that require few parameters and no tuning can often be competitive with more complex models. Combining simple forecasts based on past returns and implied volatility can lead to better volatility control and/or lower turnover. Finally, our analysis suggests that TV strategies based on liquid assets can be implemented in the presence of realistic trading costs, delays between forecasting and rebalancing, or constraints on rebalancing frequency.

We plan to explore at least two extensions of this research in future work. First, given the large standard errors observed in our sample, we plan to study the returns of TV strategies over a longer sample of US data and international data. While extending the sample will limit the forecasts we can use (especially those with intraday data), it may allow for more reliable conclusions around performance expectations. Second, we plan to explore machine learning methods such as decision trees and neural networks for volatility forecasting. It will be interesting to assess whether machine learning methods can extract more information about future volatility through a more flexible functional form *and* lead to more effective volatility control.

Appendix A Data and methodology

We consider two broad classes of volatility estimators: standalone forecasts and forecasts based on regression models. To construct these forecasts, we use close-to-close daily returns, overnight returns, and intraday returns, all based on Refinitiv data on high-frequency S&P 500 (.INX) prices from 1997 to 2023.[5] Throughout this section, we use r to denote log returns, whereas R denotes simple returns. When daily open or close values are unavailable in the Refinitiv dataset, we impute them using Bloomberg data. Our sample period choice is driven by the availability of reliable five-minute price data for the S&P 500. The length of our sample compares favorably with other studies that use intraday data; see Endnote 2.

We use price returns to forecast volatility and total returns to construct the returns of TV strategies. Total daily returns for the S&P 500, inclusive of dividends and corrected for corporate actions, are sourced from Bloomberg.

A.1 Standalone forecasts

Our first standalone estimator is the equally-weighted average of **sq**uared **d**aily close-to-close log returns (labeled **SQ-D**) over the last week (five trading days) or month (21 trading days). Most work that discusses the potential benefits of target volatility strategies (Lo, 2016; Moreira and Muir, 2017) use this volatility measure over the past month to showcase their results.

We also present the exponentially smoothed (**ES**) average of squared daily close-to-close log returns with $\lambda = 0.94$. Exponential smoothing is widely used among practitioners and the smoothing parameter λ is typically set at a value of 0.94 with daily returns, following the *RiskMetrics*^{TM} framework (Andersen *et al.*, 2013). The formula, using a lookback window with one year (252 trading days), is given by:

$$ES_t = (1-\lambda)\sum_{k=0}^{251}\lambda^k r_{t-k}^2$$

Our next measure is the intraday realized variance (**RV**) based on 78 five-minute log returns between open and close[6]:

$$RV_t = \sum_{i=1}^{78} r^2_{\left\{t-1+\frac{i-1}{78}\to t-1+\frac{i}{78}\right\}}$$

We use equally-weighted averages of past intraday RV as predictors, with three lookback windows: 1, 5, and 21 days. RV is widely used in the literature because it leverages the information contained in intraday price movements without requiring complex modeling choices. Both empirical and theoretical studies find that the five-minute sampling interval works well for volatility estimation and forecasting (Liu *et al.*, 2015; Bollerslev *et al.*, 2018 and references therein).

We also construct a measure that covers the entire day, the full-day RV (**FRV**), by adding to RV the squared overnight return.

$$FRV_t = RV_t + \log\left(\frac{Open_t}{Close_{t-1}}\right)^2$$

Finally, we use the daily close of the **VIX** index as a predictor, with data obtained from CBOE's website.[7] In contrast to other predictors that use backward-looking information, implied volatility reflects the most up-to-date forward-looking expectations of market participants.

A.2 Regression-based forecasts

Our second class of estimators consists of widely used regression-based estimators. We use the heterogeneous autoregressive (HAR) model of Corsi (2009), the mixed data sampling (MIDAS) model of Ghysels *et al.* (2006), and the heterogeneous exponential (HExp) model of Bollerslev *et al.* (2018). We use full-day RV to build both the response variable and regressors in these models.

The HAR model is given by the equation below.

$$\begin{aligned}\overline{FRV}_{t+1,t+k} = \alpha &+ \beta_d FRV_t + \beta_w \overline{FRV}_t^w \\ &+ \beta_m \overline{FRV}_t^m + \beta_q \overline{FRV}_t^q \\ &+ \varepsilon_{t+k}\end{aligned}$$

$\overline{FRV}_{t+1,t+k}$ is the average full-day RV over the next k trading days, with k denoting the forecasting horizon of either 1, 5, or 21 trading days. The superscripts w, m, and q denote week (five trading days), month (21 trading days), and quarter (63 trading days), the lookback window over which the regressors are computed. Like Li and Tang (2023), we augment the original HAR model with a regressor that captures the volatility over the last quarter.

The MIDAS model is given by the equations below.

$$\overline{FRV}_{t+1,t+k} = \alpha + \beta MIDAS_t^{(k)} + \varepsilon_{t+k}$$

$$\begin{aligned}MIDAS_t^{(k)} &= \frac{1}{\sum_{i=1}^{L} a_i}(a_1 FRV_t + a_2 FRV_{t-1} \\ &\quad + \cdots + a_L FRV_{t-L+1}) \\ a_i &= \left(\frac{i}{L}\right)^{\theta_1 - 1}\left(1 - \frac{i}{L}\right)^{\theta_2 - 1} \\ &\quad \times \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}, \quad i = 1, \ldots, L\end{aligned}$$

The regressor in the MIDAS model is a weighted average of past full-day RV up to a maximum lag. The coefficients a_i are flexible functions of two tuning parameters θ_1 and θ_2, which allow coefficients to vary smoothly as a function of the lag. By contrast, in the HAR model, the coefficient on FRV_t is $\beta_d + \beta_w + \beta_m + \beta_q$, while the coefficient on lags up to FRV_{t-4} is $\beta_w + \beta_m + \beta_q$, inducing a step function-like change in the coefficients. The MIDAS term is indexed by the forecasting horizon because the optimal tuning parameters will vary based on the forecasting horizon. We choose a maximum lag $L = 50$, set $\theta_1 = 1$, and tune θ_2, in line with Ghysels *et al.* (2006) and Bollerslev *et al.* (2018).

Finally, the HExp model is given by the equations below.

$$\begin{aligned}\overline{FRV}_{t+1,t+k} = \alpha &+ \beta_d ExpRV_t^d \\ &+ \beta_w ExpRV_t^w \\ &+ \beta_m ExpRV_t^m \\ &+ \beta_q ExpRV_t^q + \varepsilon_{t+k}\end{aligned}$$

$$ExpRV_t^{CoM(\lambda)} = \frac{\sum_{j=1}^{252} e^{-j\lambda} FRV_{t-j+1}}{\sum_{j=1}^{252} e^{-j\lambda}}$$

$$CoM(\lambda) = \frac{e^{-\lambda}}{1 - e^{-\lambda}}$$

The center of mass (CoM) is the weighted average horizon corresponding to a given λ. By setting the center of mass equal to the lookback window of interest, one obtains a value of λ that is then used to compute the four regressors that enter the regression.

All regressions are estimated daily on an expanding window with an initial length of eight years. All volatility forecasts are winsorized at 1% annualized to prevent implausibly low or negative values. The tuning parameter for MIDAS (θ_2) is tuned annually based on the expanding window, in contrast with some previous studies that tune based on the full sample. All regressions are estimated via OLS; in unreported results, regressions estimated with a Huber loss failed to improve either volatility forecasting or control for the vast majority of specifications. In unreported results, we also estimated regressions using rolling four-year and eight-year windows, which failed to improve volatility control or forecasting performance.

A.3 Scaling

Some estimators, such as VIX or intraday RV, are not suitable for direct use as volatility forecasts. VIX would systematically overestimate realized volatility because it includes a volatility risk premium (Poon and Granger, 2003; Carr and Wu, 2009), while intraday RV would systematically underestimate because it does not account for overnight price movements. Therefore, we apply scaling to all the estimators that we use. The scaling factor is the ratio between the long-term average values of a volatility proxy and the estimator under consideration (Hansen and Lunde, 2005), computed on an expanding window with an initial length of eight years and updated daily. For both standalone and regression-based estimators, we present results for two types of scaling: scaling to full-day RV and scaling to squared daily close-to-close returns.[8]

For regression-based forecasts, the response variable is full-day RV, so the forecasts are effectively already "scaled" to full-day RV. For scaling to squared daily returns, we use full-day RV as the response variable in the regressions and scale the resulting forecasts instead of directly using squared daily returns as the response variable. This is because squared daily returns are a noisy proxy for volatility over short periods, such as one day. For example, an asset may have a squared daily return close to zero, even though substantial volatility may have occurred overnight or during the trading day. This makes squared daily returns less informative about the underlying volatility over short periods and more prone to outliers, which can affect regression estimates. Therefore, forecasting full-day RV, which is better-behaved and closer to theoretical definitions of volatility, then re-scaling it based on long-term ratios between full-day RV and squared daily returns, seems attractive for short-horizon forecasting.

A.4 Forecast averaging

We consider simple averages of the (scaled) VIX and each of the standalone and regression-based forecasts. In unreported results, for regression-based forecasts, we also generated results with the squared VIX as an additional regressor. This typically results in a negative intercept, making the model more prone to implausibly low, or even negative predictions (before winsorizing), which deteriorates volatility control. Therefore, we focus on simple averaging of forecasts in our main results.

A.5 Strategy construction and evaluation

We consider a target volatility portfolio that rebalances daily, at the close, between one-month Treasury bills and the S&P 500 in order to stabilize volatility over time around the target volatility $\sigma^* = 10\%$ annualized. The daily returns of

Treasury bills are obtained from Ken French's website.[9] Given a daily volatility forecast $\hat{\sigma}_t$ of the S&P 500, the asset weights and returns of the TV strategy are given by the equations below. We apply a leverage constraint of 150% to equity weights (w_t).

$$w_t = \min\left(\frac{\sigma^*}{\sqrt{252}\hat{\sigma}_t}, 1.5\right)$$

$$R_{t+1}^{TV} = w_t R_{t+1} + (1 - w_t) R_{t+1}^f$$

Our main focus is the relation between the volatility forecast $\hat{\sigma}_t$ and the volatility behavior of the TV strategy based on that forecast. Daily forecasts are estimated on an expanding window, with the initial window from January 12, 1998, to August 10, 2006, which covers 8 × 252 days of complete data.[10] Forecasts and strategy returns are generated from August 11, 2006, to November 30, 2023, giving us 4,302 values for evaluation.

We assess forecasts using three different metrics: out-of-sample R^2, quasi-likelihood (QLIKE), and volatility control (VC) of the TV strategy. The out-of-sample R^2 and QLIKE are two measures of forecasting accuracy, whereas VC measures how closely the TV strategy adheres to the volatility target over time.

The formula for the out-of-sample R^2 is

$$OOS\ R^2 = 1 - \frac{\sum_{t=t_0}^{T-1}(FRV_{t+1} - \hat{\sigma}_t^2)^2}{\sum_{t=t_0}^{T-1}(FRV_{t+1} - \bar{\sigma}_t^2)^2}$$

where $\bar{\sigma}_t^2$ is the average of full-day RV in the expanding window and t_0 and T are the first and last trading day of the testing sample. The out-of-sample R^2 measures the mean squared error (MSE) of the forecasts relative to that of the running average full-day RV and takes value from negative infinity to one. Negative values mean that the forecasts are even worse than simple running averages, while more positive values indicate smaller forecasting errors.

Our volatility control measure is defined as the root mean squared error (RMSE) of rolling 21-day volatility from target, given by the formula below.

$$Vol - of - vol = \sqrt{\frac{\sum_{t=t_0+21}^{T}(s_{t-20,t} - \sigma^*)^2}{T - (t_0 + 21) + 1}}$$

where $s_{t-20,t}$ denotes the annualized standard deviation of the TV strategy returns over rolling periods of 21 trading days. A lower value of this volatility control measure indicates that the TV strategy adheres to the volatility target more closely over time. Strictly speaking, this measure does not merely reflect volatility of volatility: a strategy can trivially achieve a volatility-of-volatility of zero by investing all assets in the risk-free asset. Instead, our measure reflects both volatility of volatility and bias, that is, the difference between the TV strategy's long-run volatility and its target σ^*. Nevertheless, because changes in volatility are often the main driver of the measure, and for expositional simplicity, we refer to the metric as vol-of-vol.

A.6 Trading cost methodology

We follow Lo (2016) and model trading costs as a fraction of the amount transacted, which apply to both S&P 500 and one-month T-bills. Two-way trading costs are approximately (Target equity weight – equity weight before rebalancing) × (Trading cost for equities + trading cost for bills).

We assume that trading costs are 1 bp for both S&P 500 and T-bills. Our assumption of 1 bp for one-month T-bills is conservative given the high liquidity of Treasury bills or money market instruments. In an older study based on a sample from December 30, 1996, to March 31, 2000, Fleming (2001) finds average bid–ask spreads in quoted discount rates of 0.71 bps and 0.74 bps for on-the-run three- and six-month bills, respectively.

Bessembinder *et al.* (2020) find average bid–ask spreads in quoted yields of 1 bp for two-year Treasury notes in recent years. Finally, our spot checks on Bloomberg confirm that the bid–ask spreads on one-month T-bills are well below those estimates.

The equity exposure can be implemented through highly liquid equity index futures. The bid–ask spread on E-mini S&P 500 futures is typically one quarter of a point (Lo, 2016), which, combined with a contract multiplier of \$50, leads to a bid–ask spread of $\$50 * 0.25 = \$12.5$ in dollar terms. Relative to the mid-point price, this amounts to a \$6.25 trading cost. Conservatively assuming that exchange fees and commissions bring the total cost to \$10 and an S&P 500 level of 2,000, this would translate to a trading cost of $10/(50 * 2000) = 1$ bp. For comparison, the S&P 500 was above 5,000 at the end of our sample, which would translate to a trading cost of $10/(50 * 5000) = 0.4$ bp. Therefore, our 1 bp is arguably an upper bound on prospective, real-world trading costs.

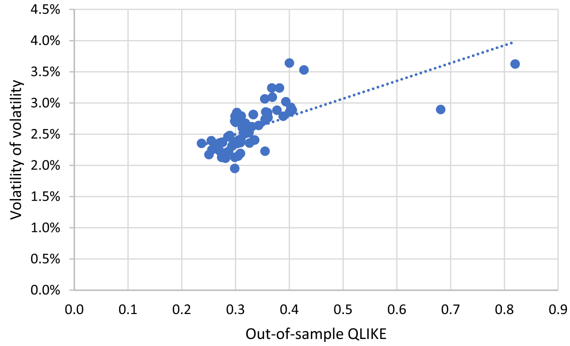
Appendix B Alternative evaluation metrics

Unlike the out-of-sample R^2, which puts equal weight on errors due to overestimation and underestimation, QLIKE penalizes underestimation more severely as can be seen from the formula below. This measure is always positive, with larger values indicating greater forecasting errors.

$$QLIKE = \frac{1}{T - t_0} \times \sum_{t=t_0}^{T-1} \left(\frac{FRV_{t+1}}{\hat{\sigma}_t^2} - \log \frac{FRV_{t+1}}{\hat{\sigma}_t^2} - 1 \right)$$

MSE (the key component of out-of-sample R^2) and QLIKE are both special cases of a loss function class characterized in Patton (2011). Members of that class share the attractive property

Exhibit B.1. QLIKE vs. vol-of-vol.



The data runs from August 11, 2006, to November 30, 2023. The charts show the out-of-sample QLIKE and vol-of-vol measures across all standalone and regression-based volatility forecasts based on two types of scaling (scaling to full-day RV or squared daily returns). See Appendix A for more details on the volatility forecasts and evaluation measures.

that the ranking of competing forecasts is robust to noise in the proxy used to represent "true" volatility. Both metrics are used by Liu *et al.* (2015) in their comparison of 400 volatility forecasts, with baseline results using QLIKE due to its potentially better ability to rank forecasts based on limited data (i.e., higher statistical power).

Exhibit B.1 shows a tighter link between QLIKE and volatility control than Exhibit 3 (based on out-of-sample R^2), with lower QLIKE values being associated with lower vol-of-vol. That said, the relation is still noisy, which again highlights the benefits of measuring forecasting accuracy and volatility control separately.

Notes

1 We abstract away from the effect of taxes. Their impact would likely be significant for TV strategies held in taxable accounts given the frequent and substantial rebalancing.

2 Our sample, which ranges from 1997 to 2023 (a full year of data is burned in to compute certain estimators), is more than twice as long than the 10-year periods considered in Ghysels *et al.* (2006) and Liu *et al.* (2015). Ghysels *et al.* (2006) use intraday data that starts in 1993, though their dataset is limited to Dow Jones Industrial Average constituents. Bollerslev *et al.* (2018) mention

in their Table 1 that the earliest equity series they use starts in late 1996, similar to our sample starting date.

3 In unreported results, we find that the VIX tends to exhibit better forecasting performance than other predictors for volatility over the next 21 days. While there are implied volatility indices at shorter tenors, we focus on the standard VIX given the short sample of the newer variants. For example, the historical data of one-day VIX starts in 2022.

4 Standard errors were obtained through stationary block bootstrap (Politis and Romano, 1994) with 10,000 replicas and mean block size of 25. We also considered block sizes of 10 and 50, as well as subsampling (Politis *et al.*, 1999) with the same three block sizes. All procedures yield similar standard errors.

5 Some five-minute prices are unavailable. Missing intraday five-minute prices are allocated based on the latest value. Days with more than six missing prices (i.e., 30 minutes worth) are dropped from the sample.

6 For the purpose of this study, we take prices at 9:35 am and 4:05 pm as the open and close prices, respectively. Our approach is consistent with Ahoniemi and Lanne (2013), who find that computing five-minute realized volatility between 09:35 am and 4:00 pm avoids contamination by stale opening prices and yields similar results as using Special Opening Quote prices. In validation exercises, we find that using 4:05 pm as the end of the trading day better matches published closing values of the S&P 500. In practice, one can use the exact close price to construct the volatility forecast or build in a lag between forecast construction and rebalancing (see Section 5.1).

7 https://www.cboe.com/tradable_products/vix/.

8 Using two types of scaling is interesting because while both full-day RV and squared daily returns are common volatility measures and highly correlated, they have slightly different magnitudes over our sample period. The average full-day RV is 18.1% annualized and the average squared daily return is 19.6% annualized.

9 https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Over our sample period, the T-bill return is the simple daily rate that, over the number of trading days in the month, compounds to one-month T-Bill rate from Ibbotson and Associates, Inc.

10 We drop days for which more than six intraday prices at five-minute intervals are missing. For estimators based on averages over multiple days, the weights are redistributed across non-missing days within the lookback window (e.g., if the 4th lag in the weekly term of HAR is missing, the 1st, 2nd, 3rd, and 5th lags receive 25% weight).

References

Ahoniemi, K. and Lanne, M. (2013). "Overnight Stock Returns and Realized Volatility," *International Journal of Forecasting* **29**, 592–604.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2013). "Financial Risk Measurement for Financial Risk Management," In *Handbook of the Economics of Finance* (Vol. 2, pp. 1127–1220). Elsevier.

Becker, R., Clements, A. E., and White, S. I. (2007). "Does Implied Volatility Provide Any Information beyond That Captured in Model-Based Volatility Forecasts?," *Journal of Banking & Finance* **31**, 2535–2549.

Bekaert, G. and Hoerova, M. (2014). "The VIX, the Variance Premium and Stock Market Volatility," *Journal of Econometrics* **183**, 181–192.

Bessembinder, H., Spatt, C., and Venkataraman, K. (2020). "A Survey of the Microstructure of Fixed-Income Markets," *Journal of Financial and Quantitative Analysis*, **55**, 1–45.

Bollerslev, T., Hood, B., Huss, J., and Pedersen, L. H. (2018). "Risk Everywhere: Modeling and Managing Volatility," *The Review of Financial Studies* **31**, 2729–2773.

Busch, T., Christensen, B. J., and Nielsen, M. Ø. (2011). "The Role of Implied Volatility in Forecasting Future Realized Volatility and Jumps in Foreign Exchange, Stock, and Bond Markets," *Journal of Econometrics* **160**, 48–57.

Carr, P. and Wu, L. (2009). "Variance Risk Premiums," *The Review of Financial Studies* **22**, 1311–1341.

Cederburg, S., O'Doherty, M. S., Wang, F., and Yan, X. S. (2020). "On the Performance of Volatility-Managed Portfolios," *Journal of Financial Economics* **138**, 95–117.

Christensen, K., Siggaard, M., and Veliyev, B. (2023). "A Machine Learning Approach to Volatility Forecasting," *Journal of Financial Econometrics* **21**, 1680–1727.

Corsi, F. (2009). "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics* **7**, 174–196.

Diebold, F. X. and Shin, M. (2019). "Machine Learning for Regularized Survey Forecast Combination: Partially-Egalitarian Lasso and Its Derivatives," *International Journal of Forecasting* **35**, 1679–1691.

Fleming, M. J. (2001). "Measuring Treasury Market Liquidity," FRB of New York Staff Report 133.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). "Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies," *Journal of Econometrics* **131**, 59–95.

Hansen, P. R. and Lunde, A. (2005). "A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data," *Journal of Financial Econometrics* **3**, 525–554.

Li, S. Z. and Tang, Y. (2023). "Automated Volatility Forecasting," SSRN Working Paper 3776915.

Liu, L. Y., Patton, A. J., and Sheppard, K. (2015). "Does Anything Beat 5-Minute RV? A Comparison of Realized Measures across Multiple Asset Classes," *Journal of Econometrics* **187**, 293–311.

Lo, A. W. (2016). "What Is an Index?," *Journal of Portfolio Management* **42**, 21.

Moreira, A. and Muir, T. (2017). "Volatility-Managed Portfolios," *The Journal of Finance* **72**, 1611–1644.

Patton, A. J. (2011). "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Journal of Econometrics* **160**, 246–256.

Patton, A. J. and Sheppard, K. (2015). "Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility," *Review of Economics and Statistics* **97**, 683–697.

Politis, D. N. and Romano, J. P. (1994). "The Stationary Bootstrap," *Journal of the American Statistical Association* **89**, 1303–1313.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). "Subsampling for Nonstationary Time Series," In: Subsampling. *Springer Series in Statistics* (New York, NY: Springer).

Poon, S. H. and Granger, C. W. J. (2003). "Forecasting Volatility in Financial Markets: A Review," *Journal of Economic Literature* **41**, 478–539.

Todorov, V. and Zhang, Y. (2022). "Information Gains from Using Short-Dated Options for Measuring and Forecasting Volatility," *Journal of Applied Econometrics* **37**, 368–391.

Keywords: Volatility-managed portfolios; volatility forecasting; realized volatility; asset allocation.