Research in International Business and Finance 75 (2025) 102723

Contents lists available at ScienceDirect

Research in International Business and Finance

journal homepage: www.elsevier.com/locate/ribaf

Volatility forecasting and volatility-timing strategies: A machine learning approach

Dohyun Chun [a,1], Hoon Cho [b,2], Doojin Ryu [c,*,3]

[a] College of Business Administration, Kangwon National University, Gangwon-do, South Korea
[b] College of Business, Korea Advanced Institute of Science and Technology, Seoul, South Korea
[c] Department of Economics, Sungkyunkwan University, Seoul, South Korea

ARTICLE INFO

JEL Classification:
C52
C53
C55
G11
G17

Keywords:
Asset allocation
Machine learning
Volatility forecasting
Volatility-timing portfolio
Risk management

ABSTRACT

Recent increases in stock price volatility have generated renewed interest in volatility-timing strategies. Based on high-dimensional models including machine learning, we predict stock market volatility and apply them to improve the performance of volatility-timing portfolios. Using various evaluation methods, we verify that those machine learning models have better prediction performances relative to the standard volatility models. Asset allocation results suggest that volatility-timing portfolios constructed using machine learning models tend to outperform the market, with higher average returns during the volatile market period. Our empirical evidence supports the application of machine learning in the construction of volatility-timing portfolios.

1. Introduction

Recently, increases in stock price volatility have generated renewed interest in investment strategies that mitigate the risk associated with stock price movements. Volatility-timing strategies, which decrease risk exposure in response to the high volatility and vice versa, could meet this demand. Seminal works of Fleming et al. (2001,2003) discover that the portfolio times with volatility provide significant economic value to risk-averse investors. Subsequent studies show significant empirical performance for various types of volatility-timing strategies based on volatility managed (Cederburg et al., 2020; Moreira and Muir, 2017), volatility targeting (Barroso and Santa-Clara, 2015), and risk targeting (Bollerslev et al., 2018) portfolios. These portfolios are described by a small position in the risky asset during volatile periods and an aggressive position during stable periods.

Our research is primarily motivated by the recent surge in stock price volatility and the consequent renewed interest in investment strategies that effectively mitigate the associated risks (Božović, 2024; Ma et al., 2022; Pezzo et al., 2023). Volatility-timing strategies, which adjust risk exposure based on market volatility levels, have shown promise in this regard (Qiao et al., 2023; Taylor, 2023; Yin et al., 2023). However, the effectiveness of these strategies heavily relies on accurate volatility forecasts. This dependency creates a

* Corresponding author.
E-mail addresses: dohyunc@kangwon.ac.kr (D. Chun), hooncho@kaist.ac.kr (H. Cho), sharpjin@skku.edu (D. Ryu).
[1] https://orcid.org/0000-0003-3031-4011
[2] https://orcid.org/0000-0003-2322-320X
[3] https://orcid.org/0000-0002-0059-4887

https://doi.org/10.1016/j.ribaf.2024.102723
Received 6 November 2022; Received in revised form 13 August 2024; Accepted 19 December 2024
Available online 25 December 2024

compelling need for advanced forecasting techniques that can capture the complex dynamics of market volatility.

Accurate volatility prediction has long been important in the field of finance. Numerous studies have improved volatility prediction by developing prediction models or identifying significant drivers in worldwide markets (Chen et al., 2022; Chun et al., 2019, 2020; Kim et al., 2017, 2018; Park et al., 2017; Shim et al., 2015; Song et al., 2016, 2018). The generalized autoregressive conditional heteroscedasticity (GARCH) model (Bollerslev, 1986; Engle, 1982) and the heterogeneous autoregressive (HAR) model (Corsi, 2009; Duan et al., 2024; Han et al., 2015, Tan et al., 2024) are representative conditional volatility models. Comprehensive works of Christiansen et al. (2012), Dangl and Hailing (2012), Lee and Ryu (2018), Nonejad (2017), Paye (2012), and Vrontos et al. (2021) show that macroeconomic and financial variables significantly improve the performance of the volatility prediction model and explain the volatility dynamics. In light of this, machine learning models have been applied to integrate the information content of predictors and generate information-intensive forecasts. For example, volatility prediction models based on penalized linear models (Audrino and Knaus, 2016; Audrino et al., 2020; Carr et al., 2019, Niu et al., 2023), tree-based models (Luong and Dokuchaev, 2018; Mittnik et al., 2015), and neural networks (Bucci, 2020; Fernandes et al., 2014; Song et al., 2024) perform well in a data-rich environment.

In this study, we forecast aggregate stock market volatility using machine learning methods and utilize them to improve the performance of volatility-timing portfolios. In particular, we first collect a wide range of economic and financial indicators that are known to or are likely to affect financial volatility. Our predictor set includes 43 variables that can be categorized into six groups: the U.S. stock market, index option, investor sentiment, financial conditions, global market, and commodity-related variables. Then, by integrating their predictive information through machine learning models, we predict daily, weekly, and monthly U.S. stock market volatilities. For the aggregate U.S. stock market volatility, Standard & Poor's (S&P) 500 realized volatilities (RVs) are estimated using intraday data. For the prediction, we use high-dimensional models that are frequently used in volatility prediction. This includes the least absolute shrinkage and selection operator (LASSO), ridge, the gradient boosting regression tree (GBRT), and model averaging methods.

Next, we compare daily, weekly, and monthly out-of-sample performances of prediction models, investigating from January 2007 to December 2020. To assess the statistical significance of the improvement in prediction accuracy, we apply the model confidence set (MCS) of Hansen et al. (2011), the hypothesis-testing frameworks of Clark and West (CW; 2007) and Giacomini and White (GW; 2006), and the out-of-sample R^2 presented by Campbell and Thompson (2008) and Welch and Goyal (2008). The evaluation results indicate that the prediction performances of machine learning models are significantly better than the benchmark HAR model, in line with previous studies (Christensen et al., 2023). Finally, we construct volatility-timing portfolios using volatility forecasts from each prediction model. We find that portfolios based on machine learning models outperform the benchmark market portfolio. Specifically, portfolios based on the LASSO forecasts have the highest returns, Sharpe ratios, and certainty equivalent returns in most cases, providing at least 24 % and at most 42 % higher Sharpe ratios than the market portfolio. A subsample analysis reveals that these portfolios have higher average returns during volatile market periods as well.

This study contributes to the literature in three ways, addressing specific areas in current research. First, our study extends the literature on financial volatility prediction by employing a comprehensive approach. In this study, we generate information-intensive volatility forecasts by exploiting the predictive information of a diverse set of predictors. We also revisit studies related to volatility forecasting and organize the evaluation frameworks. To test the predictive ability of prediction models, we adopt MCS, CW, and GW frameworks. Moreover, as performances of volatility-timing portfolios greatly rely on the accuracy of volatility estimations, studies often assess the economic value of volatility forecasts through those portfolios (Jayawardena et al., 2016; Qu and Zhang, 2022; Wang et al., 2016). Evaluating both statistical accuracy and economic performance, we demonstrate that machine learning-based volatility models produce statistically and economically superior forecasts.

Second, we contribute to the growing literature on machine learning applications in finance, with a focus on volatility-timing strategies. Machine learning has demonstrated its potential to improve stock return prediction and portfolio construction (Chinco et al., 2019; Chun et al., 2024; Freyberger et al., 2020; Gu et al., 2020, 2021; Kozak et al., 2020). Machine learning is also successful in financial volatility prediction. Nevertheless, to the best of our knowledge, few studies directly applied them to improve the performance of volatility-timing portfolios. We fill this gap and meet the investors' demand for risk management via machine learning by providing evidence that the use of machine learning models can lead to an improvement in the performance of volatility-timing portfolios, particularly during volatile market periods.

Third, our study enhances the identification of fundamental predictors of stock volatility, addressing the limitations of restricted predictor selection in existing research. Our predictor set includes well-known volatility predictors such as stock market indicators (Carr and Wu, 2017; Christiansen et al., 2012; Nagel, 2012), option and futures market variables (Eraker and Wu, 2017; Griffin and Shams, 2017; Johnson, 2017; Xing et al., 2010), investor sentiment (Audrino et al., 2020; Siganos et al., 2017), and macroeconomics indicators (Amendola et al., 2019; Chiu et al., 2018; Ryu et al., 2022; Yu and Ryu, 2021). We note that high-dimensional prediction models organize the set of predictors that contain significant predictive information. Thus, we can assess the predictability of individual predictors using appropriate feature importance measures. Our prediction results suggest that aggregate stock market bid-ask spread, index option-implied volatilities, high-yield bond spread, Google search volume index, and the Chinese stock market variables among others are significant predictors for the U.S. stock market volatilities.

This paper proceeds as follows. Section 2 introduces realized volatility estimators and predictor variables. Section 3 describes the methodologies used in this study. Section 4 reports empirical results, including out-of-sample prediction accuracy and the performance of volatility-timing portfolios. Section 5 concludes the paper.

2. Volatility prediction

2.1. Realized volatility estimation

For the target variable, we employ the daily, weekly, and monthly RVs of the S&P 500 index over the period spanning from January 2, 2004, to December 31, 2020, a total of 4311 trading days. The daily (open-to-open) RV is decomposed into open-to-close and close-to-open RVs. For the open-to-close RV, we use intraday data on the S&P 500 index, collected from 9:30 a.m. to 4:00 p.m. EST, provided by the CBOE. High-frequency financial data contains valuable information but also includes market microstructural noise. To reduce this noise and produce more accurate estimations, various noise-adjusted RV estimators have been developed, including 5-minute (Andersen et al., 1999), pre-averaging (PRV; Jacod et al., 2009), kernel (Barndorff-Nielsen et al., 2008), and multiscale (Zhang, 2006; Zhang et al., 2005) RV estimators.[4] This study uses PRV, which is known to be more efficient for one-second and five-second data.[5] PRV applies a kernel-like weighting function to locally average prices and calculates returns using nearby pre-averaged prices:

$$\widehat{PRV}_n = \frac{1}{\phi_{D,h}(f)} \frac{h}{h-D} \sum_{d=1}^{h-D+1} \left(\overline{Y}(t_d)^2 \quad -\frac{1}{2}\widehat{Y}_d\right),$$

$$\overline{Y}(t_d) = \sum_{l=1}^{D-1} f\left(\frac{l}{D}\right)\left[Y_{t,d+l} - Y_{t,d+l-1}\right],$$

$$\widehat{Y}_d = \sum_{l=1}^{D} \left(f\left(\frac{l}{D}\right) - f\left(\frac{l-1}{D}\right)\right)^2 \left(Y_{t,d+l} - Y_{t,d+l-1}\right)^2,$$

$$\phi_{D,h}(f) = \sum_{l=1}^{D} f\left(\frac{l}{D}\right)^2, \quad f(x) = \min(x, 1-x) \tag{1}$$

where $Y_{t,d}$ denotes the d^{th} price on day t, h signifies the number of time intervals within the day, and D represents the length of the observation window (Aït-Sahalia and Jacod, 2014).[6]

It is to be noted that the high-frequency stock market index is not accessible during the market closure. In order to estimate the close-to-open RV, the square of the overnight return is utilized, as this is the widely accepted method for determining overnight volatility (Koopman et al., 2005; Martens, 2002). In consequence, the daily RV is defined as

$$\widehat{RV}_{oo,n} = \widehat{PRV}_n + r_{co,n}^2, \tag{2}$$

where $\widehat{PRV}_n$ is the open-to-close PRV estimate, and $r_{co,n}$ is the close-to-open return on day n. Finally, daily (RV_d), weekly (RV_w), and monthly (RV_m) target RVs can be calculated as

$$RV_{d,n+1} = \widehat{RV}_{oo,\ n+1}, \quad RV_{w,\ n+1} = \frac{\sum_{i=n+1}^{n+5} \widehat{RV}_{oo,i}}{5}, \text{and} RV_{m,n+1} = \frac{\sum_{i=n+1}^{n+20} \widehat{RV}_{oo,i}}{20} \tag{3}$$

For better statistical properties, we consider the logarithm of volatilities. Fig. 1 shows the daily S&P 500 returns and RV estimates during our sample period.

2.2. Predictor variables

The predictor variables we used in this study can be categorized into six groups, described as follows.

(a) The U.S. stock market variables: Existing studies uncovered various stock market indicators to predict market volatility. The predictor variables related to stock characteristics are the open-to-close S&P 500 return (*SPX*), the total daily trading volume ($) of the New York Stock Exchange (NYSE, *Totdvol*), S&P 500 ETF (SPX, VOO, IVV) trading volume (*SPXETFvol*), market aggregate bid-ask spread-based illiquidity (*BAS*), Amihud's (2002) illiquidity measure (*Amihud*), Nagel's (2012) short-term reversal factor (*STR*), an aggregate measure of earnings news (*Earnnews*), and the realized skewness (*Rskew*) and kurtosis (*Rkurt*) of equity returns.

[4] Andersen et al. (2006) decompose the inter-day variance into intraday variance and autocovariance and find the optimal number of time intervals that minimizes the bias from the noise. See also Chatziantoniou et al. (2019) and Degiannakis and Filis (2017) for further details.

[5] 5-minute RVs are also employed for the empirical analysis but we only report the results for the PRV owing to their similarity. The results for the 5-minute RV can be provided upon request.

[6] The data frequency varies over time, with a 15-second frequency from 2004 to 2011 (h=1540), a 5-second frequency from 2012 to 2014 (h=4680), and varying from one to five seconds from 2014 to July 23, 2015. After July 24, 2015, data is collected every second, totaling 23,400 intraday data points per day except on early closing days. Furthermore, the choice of D plays a key role in constructing the PRV estimator. For the optimal convergence, we set $D = \sqrt{h}$ (Kim et al., 2016).

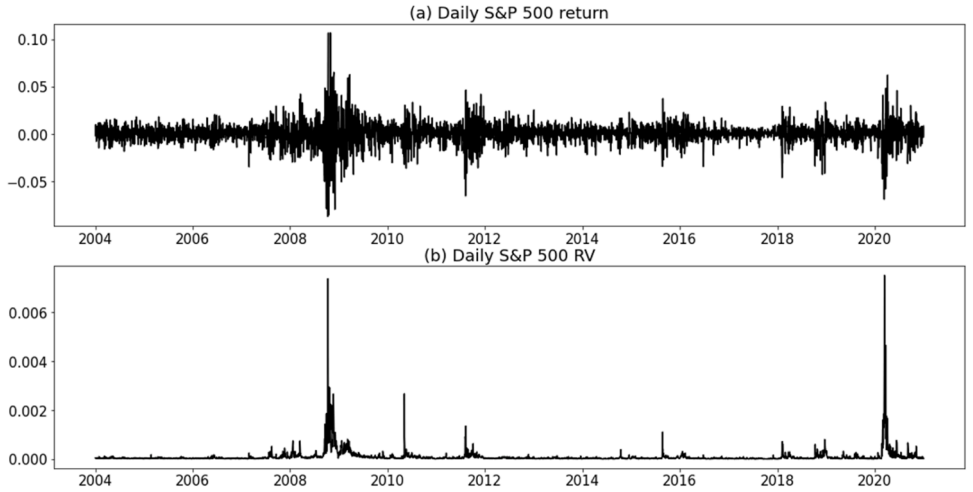

Fig. 1. Daily S&P 500 index returns and realized volatility estimates. *Notes*. This figure illustrates daily open-to-open S&P 500 returns (upper panel) and the PRV (lower panel) for the period January 2, 2004, to December 31, 2020. The year is shown on the *x*-axis.

(b) Index option-related variables: Informed traders often realize their information through options and futures markets. Therefore, the options and futures market indicators are widely used to predict stock market behaviors. We incorporate variables including a volatility index that measures the implied volatility of S&P 500 index options (*VIX*), the implied volatility of at-the-money options (*ATMIV*), the put-call ratio for U.S. equities and the S&P 500 index (*PCR_EQT* and *PCR_SPX*, respectively), and the disparity between out-of-the-money (OTM) put and call option implied volatilities, calculated for two different strike prices (volatility slope; *IVslope1*, *IVslope2*).

(c) Investor attention, sentiment, and economic uncertainty variables: Investors' irrational behavior can affect the stock market (Kim and Ryu, 2020, 2021, 2022; Kim et al., 2021, 2021, 2022; Lee and Ryu, 2024; Ryu and Yu, 2022; Ryu et al., 2023a, 2023b; Seok et al., 2022, 2024). We utilize variables to quantify investor attention and sentiment, including daily aggregate abnormal trading volumes (*Abtv*), a comprehensive Google search volume index encompassing keywords related to the stock market, S&P 500, and Nasdaq (*GSV*), the daily sentiment index derived from news articles (*News*), the variability in returns across different firms (*Stdret*), the economic policy uncertainty index (*EPU*), and a U.S. economic surprise index (*SURP*).

(d) Interest rate and financial condition variables: Numerous academic studies have documented the inverse relationship between financial instability and economic cycles. In the context of macroeconomic conditions, we examine the interest rate spreads such as the difference between the yield on corporate bonds and government bonds (*DEF*), the difference between the yield on long-term government bonds and short-term government bonds (*TERM*),

TED spread (TED): The difference between the yield on three-month Treasury bills and three-month certificates of deposit issued by the AAA-rated banks (*TED*), and the difference between the yield on high-yield corporate bonds and government bonds (*HYBSpread*). To capture key macroeconomic indicators, we use the Goldman Sachs Financial Condition Index (*FINCON*) as a proxy for the overall state of macroeconomic conditions.

(e) Global market variables: Global market indicators are essential predictors of U.S. stock volatilities, owing to financial globalization and overnight information. To capture the overseas market movements, we include the Hang Seng (*HSX*), Shanghai Composite (*SCX*), Nikkei 225 (*NIKKEI*), EURO STOXX 50 (*STOXX*), and FTSE100 (*UKX*) returns, the All Country World Index of Morgan Stanley Capital International (MSCI ACWI index, *ACWI*), as well as PRV estimators for the Hang Seng (*HSXvol*), NIKKEI 225 (*NIKKEIvol*), and EURO STOXX 50 (*STOXXvol*). Considering the different time zones, we use the Asian data on day *t* and the European data on day *t*–1 to predict the U.S. stock market volatility on day *t*. Furthermore, we consider a range of global economic variables, such as the U.S. dollar index (*Dollarind*), the Emerging Market Bond Index Spread by J.P. Morgan (*EMBISpread*), and the Citi Economic Surprise Indices for China and EU (*SURP_CN* and *SUPR_EU*, respectively).

(f) Commodity variables: A large strand of the literature shows that commodities have predictive power over stock markets. The last set of predictor variables is related to the price of commodities. This includes the Commodity Research Bureau (CRB) commodity index (*CRBCMDT*), the price of gold futures (*Gold*), the price of WTI futures (*WTI*), and the implied volatility index of oil ETFs (*OilVol*).

Table 1 summarizes the descriptions and sources of the variables.

3. Methodologies

3.1. Machine learning models

Machine learning simply identifies the best functional relationships between target and predictor variables based on statistical properties. Collaborated with economists' intuition, machine learning is widely adopted for equilibrium asset pricing and other

Table 1
Summary of predictor variables.

Predictor	Description	Source
U.S. stock market variables		
SPX	S&P 500 index return	Bloomberg
Totdvol	NYSE trading volume	Bloomberg
SPXETFtdv	S&P 500 ETF trading volume	Bloomberg
BAS	Market aggregate bid-ask spread measure	CRSP
Amihud	Market aggregate Amihud measure	CRSP
STR	Short-term reversal factor	French's website
Earnnews	Total number of quarterly earnings news	CRSP
Rskew	Realized skewness	Intraday data from CBOE
Rkurt	Realized kurtosis	Intraday data from CBOE
Index option-related variables		
VIX	CBOE volatility index: VIX	Bloomberg
ATMIV	ATM call and put option IVs	Bloomberg
IVslope1	Difference between OTM put and call option IVs	Bloomberg
IVslope2	Difference between deep OTM put and call option IVs	Bloomberg
PCR_EQT	CBOE equity put/call volume ratios	CBOE
PCR_SPX	S&P500 index put/call volume ratios	CBOE
Investor attention, sentiment, and economic uncertainty variables		
Abtv	Abnormal NYSE dollar volume	CRSP
GSV	Google search volume keywords: "stock market"&"S&P 500"&"Nasdaq"	Google
News	News sentiment index of Shapiro et al. (2022)	FRBSF
Stdret	Cross-sectional return standard deviation	CRSP
EPU	Economic policy uncertainty	World News Bank Service
SURP	U.S. Citi economic surprise index	Bloomberg
Interest rate and financial condition variables		
DEF	Default spread (BAA - AAA)	Moody's
TERM	Term spread (10Y-3M)	Moody's
TED	TED spread (3 M LIBOR – 3 M T-bill)	FRB
HYBSpread	Spread between low-grade bond yield and spot treasury curve	ICE
FINCON	Goldman Sachs U.S. financial conditions index return	Bloomberg
Global market variables		
HSX	Hang Seng Index return	Hong Kong Exchanges
SCX	Shanghai Composite Index return	Shanghai Stock Exchange
NIKKEI	Nikkei225 index return	Tokyo Stock Exchange
STOXX	EURO STOXX index return	EURO STOXX
UKX	FTSE 100 index return	Bloomberg
ACWI	MSCI ACWI index return	Bloomberg
HSXvol	PRV of HSX	FirstRate Data
NIKKEIvol	PRV of NIKKEI	FirstRate Data
STOXXvol	PRV of STOXX	FirstRate Data
Dollarind	Dollar index return	Bloomberg
EMBISpread	J.P. Morgan EMBI global spread	Bloomberg
SURP_CN	China Citi economic surprise index	Bloomberg
SURP_EU	EU Citi economic surprise index	Bloomberg
Commodity variables		
CRBCMDT	CRB commodity index return	CRB
Gold	Gold futures return	Bloomberg
WTI	WTI futures return	Bloomberg
OilVol	Oil ETF implied volatility	CBOE

Notes. This table summarizes the predictor variables by reporting their symbols, descriptions, and sources.

prediction studies in the field of finance (Bang and Ryu, 2023; Kim et al., 2021a, 2021b, 2022, Park and Ryu, 2021a, 2021b). Machine learning models play an important role in this study because they are used to integrate the drivers' predictive information and predict aggregate stock market volatility.

In this subsection, we describe three machine learning models, the first of which is the LASSO (Tibshirani, 1996). The LASSO is a representative penalized regression that uses L_1-penalization for coefficient shrinkage and variable selection (Bang et al., 2024; Bang and Ryu, 2024). The LASSO minimizes estimation error by considering the penalty parameter ν:

$$a^s, \quad \boldsymbol{b}^s = \underset{\alpha,\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^{N} \{(y_{n+1} - \alpha - \boldsymbol{X_n}\boldsymbol{\beta})^2 + \nu|\boldsymbol{\beta}|\}, \tag{4}$$

where y is a target variable, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_R)^{\mathrm{T}}$ and $\boldsymbol{X}_n = \left(X_{1,n},\ X_{2,n},\ \ldots,\ X_{R,n}\right)$ are vectors of parameters and explanatory variables, respectively, the superscript T denotes the transpose, and R is the number of predictors. The penalty parameter ν determines the intensity of coefficient shrinkage and variable selection. Moreover, we consider a ridge regression (Hoerl and Kennard, 1970), which is a penalized regression that uses an L_2-penalization. In contrast to the LASSO, ridge regression mitigates multicollinearity with

parameter shrinkage. The ridge regression minimizes the following equation:

$$\alpha^s, \quad \boldsymbol{b}^s = \underset{\alpha\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{N}\sum_{n=1}^{N}\{(y_{n+1} - \alpha - \boldsymbol{X_n\beta})^2 + \nu|\boldsymbol{\beta}|^2\}. \tag{5}$$

Again, ν determines the regularization strength in the ridge regression.

However, penalized linear regression models cannot account for nonlinear and interactive relationships between predictor variables. Here, regression trees are representative machine learning approaches that can deal with these characteristics. A tree is designed along a recursive partitioning to pick and split the predictor variables to be as homogeneous as possible in relation to the target variable. This sequential branching approximates the functional relationship within each partition. To avoid overfitting, trees are often regularized using boosting. One of the most popular supervised machine learning algorithms, the gradient boosting regression tree (GBRT), applies the gradient boosting algorithm to a regression tree (Friedman, 2001). The GBRT generates optimal forecasts by combining forecasts from a variety of trees. Based on variance reduction, it provides feature importance scores for each predictor variable. The detailed process of hyperparameter tuning for each machine-learning model is further explained in Section 3.4.

3.2. Model averaging methods

The usual model selection approaches often suffer from parameter instability and model uncertainty. To use the information contained in all candidate models, researchers take an average over all possible models, applying the weight vector that maximizes the information efficiency (Steel, 2020). In the Bayesian model averaging (BMA) framework, the optimal weight vector is derived based on Bayesian inference (Hoeting et al., 1999; Koop, 2003). Specifically, models for all possible combinations of the explanatory variables are estimated. Then, their weighted average is calculated, where the optimal weights correspond to the posterior probability. For instance, the optimal expectation of the target variable y is calculated as

$$E(y|\boldsymbol{X}) = \sum_{j=1}^{J} p(M^j|y,\boldsymbol{X}) \times y|M^j,\boldsymbol{X} \text{where} p(M^j|y, \quad \boldsymbol{X}) = \frac{p(y|M^j,\boldsymbol{X})p(M^j)}{\sum_{i=1}^{J} p(y|M^i,\boldsymbol{X})p(M^i)}. \tag{6}$$

Here, M^j denotes a forecast model j, J is the total number of candidate models, and $p(M^j|y, \boldsymbol{X})$ denotes the model's posterior probability. Christiansen et al. (2012), Dangl and Hailing (2012), Nonejad (2017), and Wang et al. (2016) apply the BMA to forecast stock market volatility, and find that integrating information using the BMA improves the out-of-sample performance. Instead of obtaining a weighted average, one can simply select the forecast combination with the highest posterior probability, called Bayesian model selection (BMS). Considering the large dimensional set of variables, we employ the birth-death Markov chain Monte Carlo (MCMC) sampler. As we do not have enough prior knowledge, we assume a uniform model prior, and assign unit information prior to Zellner's g-prior. We confirm that the empirical results are robust to the MCMC sampler and prior settings. The posterior inclusion probability of each predictor represents the sum of the probability that the predictor is included in each candidate model, which can be used as a proxy for the variable importance in a BMA inference.

In contrast to Bayesian averaging frameworks, frequentist methods do not rely on the priors of the parameter values, and focus mainly on the asymptotic behavior of the estimators. From a linear regression perspective, frequentist model averaging (FMA) estimators can be expressed as

$$\boldsymbol{b}^{FMA} = \sum_{j=1}^{J} \omega^j \quad \boldsymbol{b}^j, \tag{7}$$

where $\boldsymbol{b}^j$ and ω^j are an estimator and a weight, respectively, corresponding to the candidate model j. Based on information criteria, Buckland et al. (1997) suggest optimal weights as

$$\omega^j = \frac{\exp\left(-\frac{\mathscr{I}^j}{2}\right)}{\sum_{i=1}^{J}\exp\left(-\frac{\mathscr{I}^i}{2}\right)}, \tag{8}$$

where $\mathscr{I}^j$ is an information criterion for model j. Hansen (2007) introduces a least-squares model averaging approach that minimizes the Mallows criterion, namely, Mallows model averaging (MMA). Consider the homoscedastic regression model given by

$$y_n = \sum_{j=1}^{\infty} \beta^j X_{j,n} + \epsilon_{0,n}, \tag{9}$$

where $E(\epsilon_{0,n}|\boldsymbol{X}_n) = 0$, $E\left(\epsilon_{0,n}^2\middle|\boldsymbol{X}_n\right) = \sigma^2$, and $\boldsymbol{X}_n = (X_{1,n}, X_{2,n}, \ldots)$. Assume that the first κ^j variables are used in the model j and $0 < \kappa^1 < \kappa^2 < \ldots$. Then, the model j can be written as

$$y_n = \sum_{j=1}^{\kappa^j} \beta^j X_{j,n} + \epsilon_n^j \text{or} \boldsymbol{Y} = \boldsymbol{X}^j\boldsymbol{\beta}^j + \boldsymbol{\epsilon}^j, \tag{10}$$

where $\boldsymbol{Y}$ is a $N \times 1$ target vector, $\boldsymbol{X}^j$ is a $N \times \kappa^j$ matrix, and $\boldsymbol{\beta}^j$ is a $\kappa^j \times 1$ parameter vector. Then, the optimal weight of the MMA can be obtained as

$$\widehat{\boldsymbol{\omega}}^{\mathrm{MMA}} = \underset{\boldsymbol{\omega}}{argmin}\left\{(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{b}) + 2\sigma^2\Phi(\boldsymbol{\omega})\right\}, \tag{11}$$

where $\boldsymbol{b} = \sum_{j=1}^{J}\omega^j\begin{pmatrix}\boldsymbol{b}^j\\ 0\end{pmatrix}$, $\Phi(\boldsymbol{\omega}) = \mathrm{tr}(\Psi(\boldsymbol{\omega}))$, $\mathrm{tr}(\cdot)$ is the trace of a matrix, $\Psi(\boldsymbol{\omega}) = \sum_{j=1}^{R}\omega^j \quad \Psi^j$, and $\Psi^j = \boldsymbol{X}^j\left(\boldsymbol{X}^{j,\mathrm{T}}\boldsymbol{X}^j\right)^{-1}\boldsymbol{X}^{j,\mathrm{T}}$. Hansen and Racine (2012) extend the MMA to introduce jackknife model averaging (JMA), a method for heteroscedastic and non-nested models. Under the JMA framework, optimal weights can be obtained as

$$\widehat{\boldsymbol{\omega}}^{\mathrm{JMA}} = \underset{\boldsymbol{\omega}}{argmin}\left\{(\boldsymbol{Y}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega}))^{\mathrm{T}}(\boldsymbol{Y}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega}))\right\}, \tag{12}$$

where $\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \sum_{j=1}^{J}\omega^j\widetilde{\boldsymbol{\mu}}^j$, $\widetilde{\boldsymbol{\mu}}^j = \left(\widetilde{\mu}_1^j, \quad \widetilde{\mu}_2^j, \ldots, \quad \widetilde{\mu}_N^j\right)^{\mathrm{T}}$, $\widetilde{\mu}_n^j = \boldsymbol{X}_n^j\left(\boldsymbol{X}_{-n}^{j,\mathrm{T}}\boldsymbol{X}_{-n}^j\right)^{-1}\boldsymbol{X}_{-n}^{j,\mathrm{T}}\boldsymbol{Y}_n$, and $\boldsymbol{X}_{-n}$ represents the matrix without the n^{th} row. Using $\widehat{\boldsymbol{\omega}}^{\mathrm{MMA}}$ and $\widehat{\boldsymbol{\omega}}^{\mathrm{JMA}}$, we obtain optimal estimators and forecasts across all candidate models.

3.3. Benchmark models

As a benchmark, we consider a HARSJ model, which incorporates exogenous regressors, the signed jump variation, and the negative semi-variance in the HAR model of Corsi (2009). Patton and Sheppard (2015) define a signed jump variation as

$$Jump_n = SV_{pos,n} - SV_{neg,n}, \tag{13}$$

where $SV_{pos,n} = \sum_{k=1}^{K} ir_{n,k}^2 1_{[ir_{n,k}\geq 0]}$ and $SV_{neg,n} = \sum_{k=1}^{K} ir_{n,k}^2 1_{[ir_{n,k}<0]}$ denote positive and negative semi-variances, respectively, while $\mathrm{ir_{n,k}}$ is the $\mathrm{k^{th}}$ five-minute return on day n, K stands for the total number of five-minute returns on day n, and $1_{[\bullet]}$ is an indicator function. Then, the HARSJ model can be expressed as follows:

$$RV_{n+1} = \alpha + \vartheta_d RV_{d,n} + \vartheta_w RV_{w,n} + \vartheta_m RV_{m,n} + \vartheta_{nj} SV_{neg,n} + \vartheta_j Jump_n + \epsilon_{n+1}, \tag{14}$$

where $RV_{d,n}$, $RV_{w,n}$, and $RV_{m,n}$ are daily, weekly, and monthly historical RVs, respectively.[7]

Additionally, we calculate a simple combination of individual forecasts based on the HARSJ model. The HARSJ-X model, which incorporates the exogenous regressor $X_{r,n}$ in the HARSJ model, can be described as follows:

$$RV_{r,\quad n+1} = \alpha + \vartheta_d RV_{d,n} + \vartheta_w RV_{w,n} + \vartheta_m RV_{m,n} + \vartheta_{nj} SV_{neg,n} + \vartheta_j Jump_n + \beta_r X_{r,\quad n} + \epsilon_{r,\quad n+1}. \tag{15}$$

We obtain individual HARSJ-X forecasts $\widehat{RV}_{r,n+1}$ by incorporating each predictor variable into the exogenous regressor $X_{r,n}$. Then, given the individual HARSJ-X forecasts $\widehat{RV}_{r,n+1}$, the simple forecast combinations are calculated as

$$\widehat{RV}_{c,n+1} = \sum_{r=1}^{R}\delta_{r,n+1}\widehat{RV}_{r,n+1}, \tag{16}$$

where $\widehat{RV}_{r,n+1}$ denotes the forecast from the HARSJ-X model r. A mean forecast combination (MFC) is obtained by assigning equal weights to each forecast component (i.e., $\delta_r = \frac{1}{R}$). The discounted mean squared prediction error (*DMSPE*) uses the inverse of the historical MSPE for the weights, given by

$$\delta_{r,n+1} = \frac{\left(\widehat{RV}_{r,n} - RV_n\right)^{-2}}{\sum_{i=1}^{R}\left(\widehat{RV}_{i,n} - RV_n\right)^{-2}}. \tag{17}$$

3.4. Out-of-sample prediction and evaluation methods

Based on each high-dimensional model, we obtain the daily, weekly, and monthly RV forecasts as follows:
Model s prediction:

$$\widehat{RV}_{s,\quad n+1} = g^s(\boldsymbol{X}_n), \tag{18}$$

where $g^s(\cdot)$ is an estimated function based on the model s and $\boldsymbol{X}_n$ represents the predictor variables available on day n. The prediction

[7] We examine extensions of the HAR model as reference models and observe that our HARSJ model exhibits superior predictive performance for the U.S. stock market RVs, compared to alternative models. Nevertheless, the main findings of this study are robust to the benchmarks.

model g^s are recursively estimated from 500-day moving windows, which are further divided into 400 days for training and 100 days for validation. Through this validation procedure, we determine the optimal hyper-parameters for each machine learning model using a systematic grid search approach. For LASSO and ridge regression, we explore regularization parameters (ν) ranging from 10^{-3} to 1, using logarithmically spaced values. For GBRT, we tune the maximum depth (1−6), number of trees (1−100), and learning rate (0.01–0.1). The best-performing parameters on the validation set are then selected for out-of-sample predictions. For the weekly and monthly predictions, we used the direct method rather than the iterated method. While the iterated method is generally known to be more accurate, it is also more sensitive to model misspecification (Marcellino et al., 2006). Since our forecasts rely heavily on the predictor set, we believe that the costs of the iterated forecasting method (e.g., increased model bias and computational burden) would outweigh its benefits in our case. We assume an investor produces daily, weekly, and monthly volatility forecasts every morning, conditioning on the most recent information. Out-of-sample predictions are conducted from January 3, 2007, to December 31, 2020 (N = 3498).

For the statistical evaluation of the out-of-sample predictive performance, we consider four statistics based on mean squared prediction errors (MSPEs). The first is the model confidence set (MCS) procedure proposed by Hansen et al. (2011). Given a specific loss function, the MCS demonstrates a statistically superior subset of competing forecasts. The MCS procedure allows us to test whether a single forecasting model dominates its competitors based on their out-of-sample performance. For example, Wang et al. (2016) compare various volatility forecasting models using the MCS procedure, and show that dynamic model averaging generates better forecasts. Samuels and Sekkel (2017), Amendola et al. (2020), and Chiang et al. (2021) identify the best subset of the forecasts and produce superior forecasts using the MCS procedure. Following these studies, we perform a block bootstrap procedure of 5000 resamples.

To test the large-sample and finite-sample prediction performance, we apply the frameworks of Clark and West (2007); CW) and Giacomini and White (2006); GW), respectively. The CW and GW statistics test different aspects of the equal predictive accuracy of the model: the CW statistic focuses on the large-sample performance, whereas the GW statistic deals with the finite-sample performance. The detailed evaluation procedures are as follows. Let $MSPE_s$ be the MSPE of the model s forecasts and $\widehat{RV}_{BM,n}$ and $MSPE_{BM}$ be the fitted RV and MSPE, respectively, from the benchmark model. Then, the difference between $MSPE_s$ and $MSPE_{BM}$ is the finite-sample prediction accuracy of the HARSJ-X model s relative to the benchmark. Thus, the GW statistic tests the null hypothesis H_0: $E[MSPE_{BM} - MSPE_s] = 0$, as follows:[8]

$$GW_s = \frac{\overline{\xi}_s}{\sqrt{Var(\xi_{s,n})/N}}, \quad \overline{\xi}_s = \frac{1}{N}\sum_{n=1}^{N}\xi_{s,n}, \quad \xi_{s,n} = \left(RV_n - \widehat{RV}_{BM,n}\right)^2 - \left(RV_n - \widehat{RV}_{s,n}\right)^2 \tag{19}$$

Intuitively, the GW statistic evaluates the finite-sample improvement in terms of the prediction error. It has a negative value when the cost (i.e., prediction error) of incorporating additional predictors outweighs the benefit (i.e., prediction accuracy). In contrast, the CW statistic tests the large-sample prediction performance by adjusting the parameter estimation error. The CW statistic is calculated as follows:

$$CW_s = \frac{\overline{\xi}_{adj,s}}{\sqrt{\mathrm{Var}\left(\xi_{adj,s,n}\right)/N}},$$

$$\overline{\xi}_{adj,s} = \frac{1}{N}\sum_{n=1}^{N}\xi_{adj,s,n} = MSPE_{BM} - MSPE_s + \frac{1}{N}\sum_{n=1}^{N}\left(\widehat{RV}_{BM,n} - \widehat{RV}_{s,n}\right)^2,$$

$$\xi_{adj,s,n} = \left(RV_n - \widehat{RV}_{BM,n}\right)^2 - \left[\left(RV_n - \widehat{RV}_{s,n}\right)^2 - \left(\widehat{RV}_{BM,n} - \widehat{RV}_{s,n}\right)^2\right] \tag{20}$$

Note that the term $\frac{1}{N}\sum_{n=1}^{N}\left(\widehat{RV}_{BM,n} - \widehat{RV}_{s,n}\right)^2$ in $\overline{\xi}_{adj,s}$ deals with the variance increment from the parameter estimation. Note that if the explanatory variable causes the target variable, but does not improve the prediction accuracy, the CW test will be rejected, whereas the GW test will not.

Finally, we consider the out-of-sample R^2, following Welch and Goyal (2008) and Campbell and Thompson (2008). The out-of-sample R^2 of the model (R^2_{oos}) and the increase in R^2_{oos} compared with the benchmark model (ΔR^2_{oos}) are calculated as

$$R^2_{oos} = 1 - \frac{MSPE_s}{MSPE_{HM}}, \Delta R^2_{oos} = \left(1 - \frac{MSPE_s}{MSPE_{HM}}\right) - \left(1 - \frac{MSPE_{BM}}{MSPE_{HM}}\right) = \frac{MSPE_{BM} - MSPE_s}{MSPE_{HM}} \tag{21}$$

where $MSPE_{HM}$ denotes the MSPE of the historical mean model. The ΔR^2_{oos} value captures the relative prediction error, scaled by the MSPE of the historical mean model.

[8] We apply a Newey–West heteroscedasticity and autocorrelation-adjusted estimator for the variance of the statistics.

Table 2
Evaluation metrics overview.

Metric	Purpose	Section
MSPE & MAPE	Gauges prediction accuracy	4.2
MCS	Identifies the top model set	4.2
CW test	Assesses nested model accuracy differences	4.2
GW test	Examines finite-sample predictive ability	4.2
R^2_{oos}	Compares against benchmark	4.2
Sharpe ratio	Evaluates risk-adjusted performance	4.3
CER	Measures risk-adjusted economic value	4.3
Realized utility	Assesses mean-variance utility-based value	4.3

Notes. This table summarizes the evaluation metrics used in our study, presenting each metric's name, purpose, and the corresponding section(s) of the paper where it is discussed and applied.

4. Empirical results

4.1. Evaluation framework

Our study employs a comprehensive set of evaluation metrics to assess the performance of volatility forecasting models. The statistical metrics include the MSPE and MAPE, which quantify prediction accuracy. We utilize the MCS procedure to identify statistically equivalent top-performing models. The CW test examines predictive accuracy differences in nested models, while the GW test evaluates competing models' finite-sample predictive capabilities. Additionally, we use the out-of-sample R^2 to contrast model performance against a benchmark.

A crucial aspect of our evaluation framework is the assessment of the economic value of our forecasts through portfolio construction. We construct volatility-timing portfolios based on our forecasts and evaluate their performance using widely-used financial indicators. These include the Sharpe ratio, which measures risk-adjusted performance, CER, which quantifies the risk-adjusted economic value to investors, and realized utility, based on a mean-variance utility function. Additionally, we implement VIX futures trading strategies to directly assess the economic value of our forecasts using a tradable asset. These economic measures offer crucial insights into the real-world applicability and potential profitability of our forecasting models, complementing the statistical evaluation and highlighting the economic benefits of our approach. Table 2 summarizes the evaluation indicators used in our study.

4.2. Prediction results

In this subsection, we assess the out-of-sample predictive performance of the high-dimensional models relative to the benchmark forecasting. Table 3 reports the MSPE and mean absolute percentage error (MAPE), as well as GW and CW test results and the ΔR^2_{oos} value, relative to the benchmark HARSJ model. The results demonstrate the higher predictability and accuracy of forecasts that are based on machine learning and model averaging methods, compared to the benchmark HARSJ model. The improvement in prediction accuracy, captured by CW, GW, and ΔR^2_{oos}, are statistically significant in most cases. Specifically, the LASSO generates the most accurate forecasts among others, regardless of the forecasting horizon. The LASSO forecasts have the largest ΔR^2_{oos} and the smallest MSPE and MAPE, and belong to the 1 % MCS for all forecasting horizons.

4.3. Volatility-timing strategies

Volatility-timing strategies refer to asset allocation strategies conditioning on volatility. In constructing the volatility-timing portfolio, we consider the asset allocation problem of an investor who determines the optimal weighting of a market portfolio and a risk-free asset based on market volatility forecasts. Let w_t^p be the portion of an investment in the market portfolio on day t. Then, the portion of the investment in the risk-free asset is given by $1 - w_t^p$, and the portfolio return on day t ($r_{p,t}$) is expressed as

$$r_{p,t} = w_t^p r_{m,t} + (1 - w_t^p) r_{f,\quad t}, \tag{22}$$

where $r_{m,t}$ and $r_{f,t}$ are the market portfolio return and the risk-free rate, respectively. In this study, we regard the open-to-open S&P 500 return and the three-month T-bill rate as the market portfolio return and the risk-free rate, respectively.[9]

Volatility-timing portfolios are constructed by adjusting the weights in the risky asset based on the inverse of their conditional return volatility:

$$w_t^{vt} = \frac{1}{\gamma} \frac{V_t}{\widehat{\sigma}_{m,t}}, \tag{23}$$

[9] To assess the real investment performance, we skip 15 seconds after the market opens. Moreover, we restrict the range of w_t^p from −3.0–3.0, considering the investment in the index portfolio with the exchange-traded fund.

Table 3
Out-of-sample prediction accuracy of high-dimensional models.

Model	MSPE	MAPE	CW	GW	ΔR^2_{oos}
Panel A. Daily					
LASSO	0.304 * **	0.464 * **	12.708 * **	7.281 * **	3.109
Ridge	0.337	0.516	11.507 * **	−0.361	−0.311
GBRT	0.342	0.493	11.140 * **	−1.414	−0.912
BMA	0.310	0.473	14.362 * **	4.186 * **	2.539
BMS	0.316	0.479	14.113 * **	2.903 * **	1.901
JMA	0.308	0.477	14.518 * **	3.972 * **	2.663
MMA	0.308	0.477	14.349 * **	3.972 * **	2.681
DMSPE	0.330	0.478	5.832 * **	3.589 * **	0.345
MFC	0.330	0.479	6.376 * **	5.068 * **	0.358
Panel B. Weekly					
LASSO	0.271 * **	0.400 * **	8.503 * **	4.470 * **	2.729
Ridge	0.290	0.419	9.058 * **	0.319	0.275
GBRT	0.307	0.437	10.254 * **	−1.970	−1.996
BMA	0.279	0.421	10.990 * **	1.782 * *	1.661
BMS	0.284	0.426	10.844 * **	1.028	1.005
JMA	0.274	0.459	11.173 * **	1.762 * *	2.271
MMA	0.276	0.459	11.510 * **	1.601 *	2.108
DMSPE	0.287	0.409	5.257 * **	4.150 * **	0.633
MFC	0.287	0.410	4.898 * **	3.980 * **	0.593
Panel C. Monthly					
LASSO	0.360 * **	0.432 * **	10.211 * **	4.503 * **	5.765
Ridge	0.361	0.436	11.708 * **	4.068 * **	5.577
GBRT	0.376	0.481	11.203 * **	1.603 *	3.140
BMA	0.391	0.487	12.145 * **	0.387	0.786
BMS	0.396	0.488	12.108 * **	−0.041	−0.084
JMA	0.369	0.567	12.024 * **	2.058 * *	4.376
MMA	0.374	0.512	11.898 * **	1.498 *	3.536
DMSPE	0.391	0.445	4.269 * **	2.026 * *	0.752
MFC	0.389	0.446	5.494 * **	3.676 * **	1.177

Notes. This table reports the prediction accuracy of the high-dimensional models. The table reports the MSPE and MAPE; * ** means that the model belongs to the 1 % MCS. The table displays the CW and GW statistics; * ** , * *, and * denote statistical significance at the 1 %, 5 %, and 10 % levels, respectively. The table also demonstrates an increase in the out-of-sample R^2 in comparison to the benchmark HARSJ model (ΔR^2_{oos}). The variances of these statistics are computed to account for heteroscedasticity and autocorrelation.

where V_t is the target volatility, and $\widehat{\sigma}_{m,t}$ is the fitted volatility of the stock market, conditional on the past information set. For the target volatility, we apply both the fixed target and the moving variance of the stock market return. In accordance with the prediction horizons, we construct daily, weekly, and monthly portfolios, as follows. An investor rebalances a daily portfolio based on the daily volatility forecasts. For the weekly portfolios, the investor predicts the weekly volatility and forms the portfolio on day t, holds the position for a week, and rebalances the portfolio again on day $t+4$. The monthly portfolios are constructed in the same way, with a 20-day holding period.

For portfolios based on the high-dimensional model, we use daily, weekly, and monthly volatility forecasts from the out-of-sample prediction in Section 4.1. In addition to the high-dimensional models, we consider the MFC and DMSPE introduced in Section 3.3. Moreover, two standard volatility models are adopted, the first of which is the HARSJ model, presented in Eq. (14). The second model is the GARCH (1,1) model:

$$r_t = \mu + \varepsilon_t^g, \quad \varepsilon_t^g | I_{t-1} \sim N(0, \quad h_t), \quad h_t = \alpha^g + \beta^g h_{t-1} + \tau^g (\varepsilon_{t-1}^g)^2 \tag{24}$$

where h_t is the conditional volatility of the return, α^g, β^g, and τ^g are model parameters, and I_t denotes the information set. We estimate the daily, weekly, and monthly conditional volatilities using the HARSJ and GARCH models. For the comparison, we also consider a buy-and-hold market portfolio in which an investor buys a market portfolio at the beginning of the investment period, holds it without adjustment, and sells it at the end of the investment period. This can be obtained by setting $w_t^p = 1$ in Eq. (22).

For the performance measures, we employ the Sharpe ratio and the certainty equivalent returns (CERs). The Sharpe ratio captures the risk-return tradeoff by calculating the expected excess return in units of standard deviations. The CER is the risk-free return that is equivalent to holding a risky portfolio. A higher CER portfolio provides higher utility to an investor, considering their risk aversion. The Sharpe ratio and CER are computed as

$$\textit{Sharpe ratio} = \frac{E\left[r_{p,t} - r_{f,t}\right]}{\sigma_{p,t}} = \frac{\frac{\sum_{t=1}^{T}\left(r_{p,t} - r_{f,t}\right)}{T}}{\sqrt{\frac{\sum_{t=1}^{T}\left(r_{p,t} - r_{f,t}\right)^2}{T-1}}},$$

$$CER = \quad E[r_{p,t}] - \frac{1}{2}\gamma\sigma_{p,t}^2 = \frac{1}{T}\sum_{t=1}^{T} r_{p,t} - \frac{1}{2}\gamma\frac{\sum_{t=1}^{T} {r_{p,t}}^2}{T-1}, \tag{25}$$

respectively, where γ is the investor's risk aversion coefficient. We report the annualized Sharpe ratio and CER for $\gamma = 3$.

Tables 4, 5, and 6 show the performance of the daily, weekly, and monthly volatility targeting portfolios, respectively. The first three columns represent the volatility targeting portfolios with fixed target volatility corresponding to an annualized volatility of 15 %. The last three columns stand for those with the target volatilities corresponding to the one-year rolling standard deviations of the S&P 500 returns.[10] In general, portfolios constructed based on high-dimensional models exhibit superior performance compared to the static market portfolio, characterized by higher average returns, Sharpe ratios, and CERs over the evaluation period.[11] However, these portfolios do not always outperform the portfolio based on the HARSJ model. For instance, only the LASSO outperforms the HARSJ model for daily portfolios, whereas the LASSO, GBRT, JMA, and MMA models (for weekly portfolios) and the LASSO, ridge, and GBRT models (for monthly portfolios) outperform the HARSJ model in terms of the Sharpe ratio.

We need to focus on the LASSO. Portfolios that time with the LASSO forecasts perform best, regardless of the strategies and forecasting horizons. They enjoy the highest average returns, Sharpe ratios, and CERs in most cases: at least 24 %, and at most 42 % higher Sharpe ratios than the market portfolio. Fig. 2 summarizes the findings, showing the cumulative log returns of the monthly volatility targeting portfolios, ordered by the final cumulative returns in descending order. Here, we find that the volatility-timing portfolios outperform the market portfolio in general, and the LASSO portfolios consistently have the highest cumulative returns over the 14-year testing sample.

Fig. 3 illustrates the proportion of investment in the market portfolio (w_t^p) for each monthly rebalanced portfolio. The black dashed line represents the scale-adjusted monthly market RVs. As anticipated, we find that the investments in the risk asset are relatively small during high volatility periods such as the global financial crisis and the COVID pandemic, while they are relatively high during low volatility periods.

Next, we perform a subperiod analysis to thoroughly examine the performance of the volatility-timing portfolios in different market conditions. To determine the market conditions, we use three representative criteria. Firstly, we evaluate the aggregate stock market volatility and define a high volatility period when the monthly RV is among the top 30 % during our sample period. Secondly, we consider the recession indicator provided by the National Bureau of Economic Research (NBER), with a value of 1 for recessionary periods and 0 for expansionary periods. Finally, we evaluate the aggregate stock market returns and define a market downturn when the monthly S&P 500 returns are in the bottom 30 %. Fig. 4 displays the cumulative returns of the volatility-timing portfolios with market conditions. The green, grey, and red areas represent high market volatility, recession, and low market return periods, respectively.

Table 7 presents the average returns of monthly volatility targeting portfolios in different market conditions. Panels (a), (b), and (c) represent market conditions determined by market volatility, recession indicator, and market return, respectively. High-dimensional models have higher average returns compared to the benchmark HAR model in high volatility periods as seen in Panel (a). LASSO (-8.19 %), ridge (-11.19 %), and JMA (-10.54 %) based portfolios have higher average returns compared to the HAR-based (-13.24 %) and the market portfolio (-16.54 %). Furthermore, Panel (b) indicates that LASSO (-11.32 %) and ridge (-13.55 %) based portfolios perform better than others during recession periods related to market crashes like the global financial crisis and the COVID pandemic. It can be concluded that incorporating volatility forecasts from machine learning models can enhance the performance of volatility-timing portfolios, particularly during volatile market conditions.

Panel (c) of Table 7 presents the relationship between the performance of volatility-timing portfolios and stock market returns. The "*Low*" column in Panel (c) indicates that the portfolios do not effectively hedge market downturns. However, the portfolios perform better after low return periods, as shown in the "*After low*" column, which represents the average return over one month after a low return period. During this period, LASSO (24.15 %), ridge (21.81 %), and JMA (23.54 %) based portfolios outperform the HAR-based (19.43 %) and the market (11.91 %) portfolios, suggesting that the portfolios respond to market shocks and adjust the allocation in risky assets properly.

The last strategy examined is the risk-targeting strategy of Bollerslev et al. (2018). The utility maximization problem of a mean-variance investor is expressed as

$$max_{w_t^p} \quad U\{E(r_{p,t+1}|I_t), \sigma_{p,t+1}\} = \quad E(r_{p,t+1}|I_t) - \frac{1}{2}\gamma\sigma_{p,t+1}^2. \tag{26}$$

Solving the maximization problem, we obtain the optimal portfolio weight for the investor as follows:

$$w_t^{mv} = argmax_{w_t^p} U(r_{p,t+1}) = \frac{1}{\gamma}\frac{E(r_{m,t+1}|I_t) - r_{f,t+1}}{\sigma_{m,t+1}^2}. \tag{27}$$

[10] We verify that the main findings remain consistent when considering other target volatilities and gammas. The results for other target volatilities and gammas are available upon request.

[11] About the general issues on the portfolio performance measures and their applications in global markets, refer to the recent studies of Ham et al. (2019), Lee et al. (2024), Lee et al. (2023), and Seok et al. (2023).

Table 4
Performance of daily volatility targeting portfolios.

	Fixed target 15 %			One-year rolling window		
Model	Avg.ret (%)	Sharpe	CER	Avg.ret (%)	Sharpe	CER
LASSO	11.55	0.560	6.083	14.06	0.593	6.631
Ridge	11.18	0.532	5.542	13.38	0.552	5.668
GBRT	11.26	0.537	5.646	13.90	0.578	6.262
BMA	10.70	0.518	5.282	12.88	0.549	5.694
BMS	10.81	0.523	5.381	13.00	0.554	5.804
JMA	11.06	0.537	5.648	13.17	0.563	5.994
MMA	11.09	0.539	5.684	13.21	0.564	6.022
DMSPE	11.28	0.538	5.665	13.82	0.577	6.255
MFC	11.35	0.543	5.748	13.89	0.581	6.348
HARSJ	11.32	0.541	5.712	13.87	0.580	6.324
GARCH	8.87	0.481	4.705	10.93	0.531	5.533
Market	8.55	0.418	3.480	8.55	0.418	3.480

Notes. This table shows the performance of the daily volatility targeting portfolios based on the daily volatility forecasts from each model, reporting the annualized average return (*Avg.ret*), Sharpe ratio (*Sharpe*), and CER. The first three columns represent volatility targeting portfolios with fixed target volatility corresponding to an annualized volatility of 15 %. The last three columns stand for the target volatility corresponding to the one-year rolling standard deviation of the market returns. The rows labeled "*Market*" present the results for the buy-and-hold market portfolio.

Table 5
Performance of weekly volatility targeting portfolios.

	Fixed target 15 %			One-year rolling window		
Model	Avg.ret (%)	Sharpe	CER	Avg.ret (%)	Sharpe	CER
LASSO	12.06	0.600	6.176	14.49	0.629	6.709
Ridge	11.33	0.555	5.264	13.53	0.582	5.641
GBRT	12.17	0.589	5.954	14.94	0.638	6.909
BMA	11.17	0.561	5.397	13.58	0.604	6.194
BMS	11.14	0.558	5.335	13.52	0.600	6.096
JMA	11.62	0.591	5.989	13.76	0.619	6.540
MMA	11.66	0.593	6.033	13.81	0.622	6.591
DMSPE	11.97	0.587	5.917	14.30	0.614	6.369
MFC	11.79	0.579	5.748	14.17	0.609	6.245
HARSJ	11.80	0.579	5.751	14.17	0.609	6.251
GARCH	8.90	0.448	3.212	11.71	0.514	4.149
Market	8.42	0.484	4.072	8.42	0.484	4.072

Notes. This table shows the performance of the weekly volatility targeting portfolios based on the weekly volatility forecasts from each model, reporting the annualized average return (*Avg.ret*), Sharpe ratio (*Sharpe*), and CER. The first three columns represent volatility targeting portfolios with fixed target volatility corresponding to an annualized volatility of 15 %. The last three columns stand for the target volatility corresponding to the one-year rolling standard deviation of the market returns. The rows labeled "*Market*" present the results for the buy-and-hold market portfolio.

Table 6
Performance of monthly volatility targeting portfolios.

	Fixed target 15 %			One-year rolling window		
Model	Avg.ret (%)	Sharpe	CER	Avg.ret (%)	Sharpe	CER
LASSO	12.71	0.695	7.729	14.86	0.688	7.897
Ridge	12.04	0.665	7.162	13.70	0.677	7.596
GBRT	12.04	0.672	7.263	13.26	0.656	7.175
BMA	11.22	0.584	5.730	13.02	0.599	5.983
BMS	11.18	0.583	5.705	12.91	0.594	5.876
JMA	11.80	0.632	6.605	13.67	0.651	7.086
MMA	11.75	0.629	6.547	13.63	0.648	7.037
DMSPE	11.81	0.652	6.920	13.45	0.660	7.268
MFC	11.30	0.634	6.575	13.19	0.651	7.061
HARSJ	11.21	0.629	6.477	13.09	0.646	6.969
GARCH	6.81	0.326	0.358	8.40	0.349	−0.218
Market	7.80	0.500	4.186	7.80	0.500	4.186

Notes. This table shows the performance of the monthly volatility targeting portfolios based on the monthly volatility forecasts from each model, reporting the annualized average return (*Avg.ret*), Sharpe ratio (*Sharpe*), and CER. The first three columns represent volatility targeting portfolios with fixed target volatility corresponding to an annualized volatility of 15 %. The last three columns stand for the target volatility corresponding to the one-year rolling standard deviation of the market returns. The rows labeled "*Market*" present the results for the buy-and-hold market portfolio.

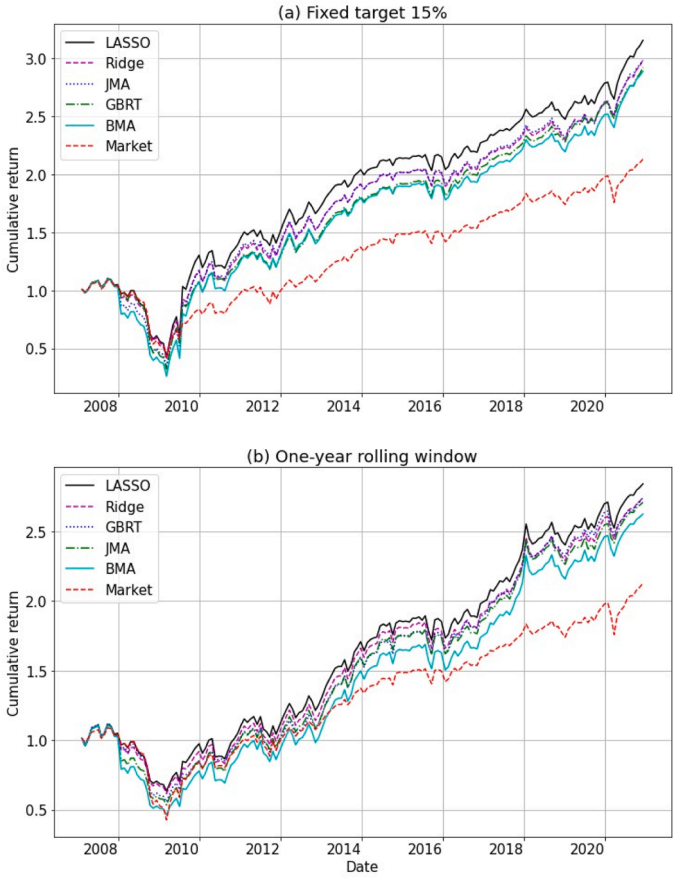

Fig. 2. Cumulative return of monthly volatility targeting portfolios. *Notes*. This figure presents the cumulative log returns of monthly volatility targeting portfolios (a) with an annualized target volatility of 15 % and (b) with the target volatility of the one-year rolling standard deviation of the market returns. The legend is ordered by the final cumulative returns in descending order. The evaluation period is from January 3, 2007, to December 31, 2020 (3498 days).

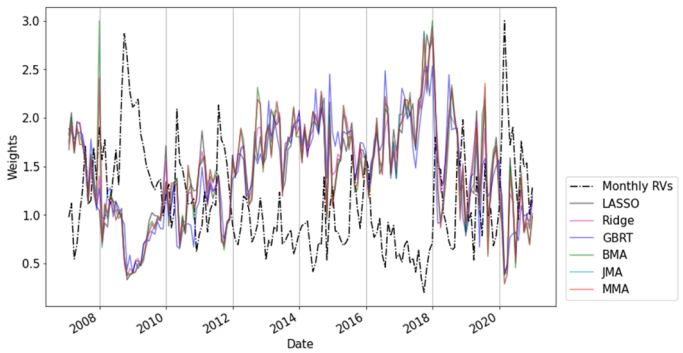

Fig. 3. Portion of the investment in the market portfolio for monthly volatility targeting portfolios. *Notes*. This figure presents the portion of the investment in the market portfolio for monthly volatility targeting portfolios based on the one-year rolling standard deviation of the market returns. The dashed black line represents log-transformed and scale-adjusted monthly RVs. The portfolio construction period is from January 3, 2007, to December 31, 2020 (3498 days).

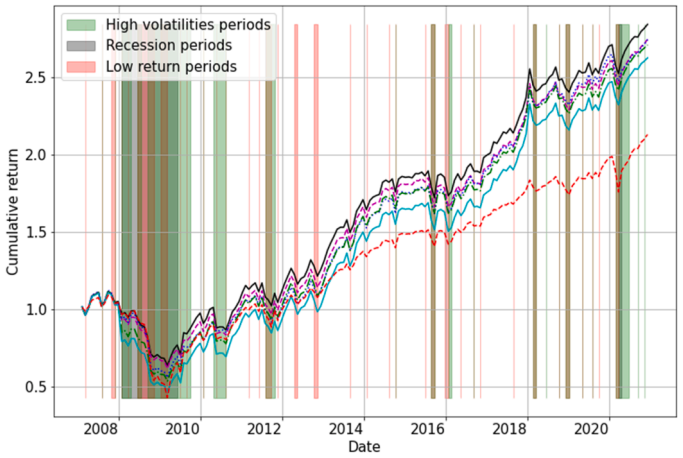

Fig. 4. Cumulative returns of monthly volatility targeting portfolios with the market conditions. *Notes*. This figure presents the cumulative log returns of monthly volatility targeting portfolios based on the one-year rolling standard deviation of the market returns. To determine the market conditions, we use three representative criteria. Firstly, we evaluate the aggregate stock market volatility and define a high volatility period when the monthly RV is among the top 30 % during our sample period. Secondly, we consider the recession indicator provided by the National Bureau of Economic Research (NBER), with a value of 1 for recessionary periods and 0 for expansionary periods. Finally, we evaluate the aggregate stock market returns and define a market downturn when the monthly S&P 500 returns are in the bottom 30 %. The green, grey, and red areas represent high market volatility, recession, and low market return periods, respectively. The evaluation period is from January 3, 2007, to December 31, 2020 (3498 days).

Table 7
Average returns of monthly volatility targeting portfolios in different market conditions.

	(a) Market volatility		(b) Recession indicator		(c) Market return		
	High	Low	Recession	Expansion	Low	High	After low
LASSO	−8.19	24.95	−11.32	18.26	−65.15	49.90	24.15
Ridge	−11.19	24.59	−13.55	17.23	−65.28	48.29	21.81
GBRT	−12.11	24.37	−18.41	17.37	−64.11	47.15	18.61
BMA	−12.55	24.22	−23.01	17.70	−68.74	48.84	22.69
BMS	−12.83	24.19	−24.27	17.74	−68.99	48.78	22.20
JMA	−10.54	24.28	−17.52	17.72	−66.29	48.70	23.54
MMA	−10.63	24.26	−17.67	17.70	−66.40	48.68	23.52
DMSPE	−12.13	24.66	−14.99	17.15	−65.21	47.91	19.80
MFC	−12.99	24.65	−16.09	16.99	−65.26	47.55	19.75
HAR	−13.24	24.63	−16.51	16.94	−65.38	47.46	19.43
GARCH	−22.06	21.74	−41.22	14.84	−75.42	45.11	9.87
Market	−16.54	18.46	−23.46	11.85	−53.70	34.73	11.91

Notes. This table shows the average return of the monthly volatility targeting portfolios based on the one-year rolling standard deviation of the market returns, in different market conditions. We use (a) market volatility, (b) recession indicator, and (c) market return to determine the market conditions, corresponding to Panels (a), (b), and (c), respectively. Firstly, we evaluate the aggregate stock market volatility and define a high volatility period when the monthly RV is among the top 30 % during our sample period (column '*High*'). Secondly, we consider the recession indicator provided by the National Bureau of Economic Research (NBER), with a value of 1 for recessionary periods and 0 for expansionary periods. Finally, we evaluate the aggregate stock market returns and define a low return period when the monthly S&P 500 returns are in the bottom 30 % (column '*Low*') Panels (a), (b), (c) represent market conditions determined by market volatility, recession indicator, and market return, respectively. The "*After low*" column represents the average return over one month after a low return period. The table represents the annualized percentage return.

Assuming a constant Sharpe ratio, we can express Eqs. (26) and (27) as follows (by omitting the constant):

$$U\left(r_{p,t+1}\right) = w_t^p \bullet \varrho \bullet \sigma_{m,t+1} - \frac{\gamma}{2}\left(w_t^p\right)^2 \sigma_{m,t+1}^2 \text{and} w_t^{rt} = \frac{\varrho}{\gamma}\frac{1}{\widehat{\sigma}_{m,t+1}^2} \tag{28}$$

where $\varrho = \frac{E\left[r_{m,t+1} - r_{f,t+1}\right]}{\sigma_{m,t+1}}$. Then, the realized utility of the investor can be written as

$$\textit{Realized utility} = \frac{\varrho^2}{\gamma}\frac{\sigma_{m,t+1}}{\widehat{\sigma}_{m,t+1}} - \frac{\varrho^2}{2\gamma}\frac{\sigma_{m,t+1}^2}{\widehat{\sigma}_{m,t+1}^2}. \tag{29}$$

In this study, we set $\varrho = 0.3$ and $\gamma = 3$. Note that the realized utility depends on the ratio of the true and the estimated volatilities,

Table 8
Realized utility of risk-targeting portfolios.

	Daily		Weekly		Monthly	
	Realized utility	*t*-stat	Realized utility	*t*-stat	Realized utility	*t*-stat
LASSO	1.377	2.868 * **	1.405	1.556	1.406	2.389 * *
Ridge	1.332	−1.177	1.403	0.776	1.280	−0.873
GBRT	1.365	−1.909 *	1.386	−1.594	1.375	−0.503
BMA	1.372	0.455	1.390	−0.901	1.387	0.229
BMS	1.369	−0.555	1.386	−1.162	1.385	0.112
JMA	1.369	−0.323	1.396	−0.342	1.393	0.732
MMA	1.369	−0.337	1.395	−0.479	1.394	0.736
DMSPE	1.372	0.564	1.401	1.068	1.398	2.525 * *
MFC	1.372	1.059	1.399	3.913 * **	1.390	3.091 * **
HARSJ	1.371	-	1.398	-	1.384	-
GARCH	0.914	−19.198 * **	0.898	−10.512 * **	0.862	−5.341 * **

Notes. This table reports the realized utility of risk-targeting portfolios, based on volatility forecasts from each model. *t*-stat is used to test the significance of the realized utility relative to the utility of the portfolio based on the HARSJ model; * ** , * *, and * denote statistical significance at the 1 %, 5 %, and 10 % levels, respectively. *Daily*, *Weekly*, and *Monthly* represent results for daily, weekly, and monthly volatility forecasts, respectively.

that is, the volatility prediction accuracy. Table 8 reports the realized utility of the risk-targeting strategies and the *t*-statistics for testing the significance of the realized utility difference, relative to the utility from the HARSJ-based portfolios. Similar to previous results, we find that not all high-dimensional models but the LASSO deliver higher utility than the benchmark HARSJ model. The positive and significant *t*-statistics imply that differences are significant for daily and monthly portfolios. The superiority of LASSO forecasts will be further discussed in the next section.

Table 9
Significant variables for out-of-sample prediction of the high-dimensional model.

	LASSO	Ridge	GBRT	JMA	MMA	BMA
Panel A. Daily						
1	HSX	BAS	SPX	ATMIV	ATMIV	BAS
2	BAS	ATMIV	VIX	BAS	BAS	HSX
3	ATMIV	HSXvol	NIKKEI	VIX	VIX	ATMIV
4	HSXvol	VIX	UKX	HSX	HSX	HSXvol
5	VIX	GSV	BAS	SPX	SPX	VIX
6	NIKKEI	HSX	Earnnews	HSXvol	FINCON	NIKKEI
7	HYBSpread	OilVol	HYBSpread	FINCON	HSXvol	HYBSpread
8	Earnnews	Stdret	TERM	ACWI	OilVol	Earnnews
9	STR	NIKKEI	SPXETFtdv	OilVol	ACWI	EMBISpread
10	UKX	HYBSpread	Totdvol	HYBSpread	HYBSpread	SPX
Panel B. Weekly						
1	HSX	HSXvol	SPX	ACWI	ACWI	HSX
2	HSXvol	Stdret	STR	FINCON	HSX	HSXvol
3	VIX	GSV	VIX	HSX	FINCON	STR
4	ATMIV	ATMIV	STOXX	VIX	ATMIV	GSV
5	SURP_EU	BAS	HYBSpread	ATMIV	VIX	NIKKEI
6	STR	VIX	Totdvol	SPX	SPX	VIX
7	NIKKEI	HSX	SPXETFtdv	HSXvol	HSXvol	SURP_EU
8	GSV	NIKKEI	DEF	SURP_EU	SURP_EU	IVslope1
9	HYBSpread	HYBSpread	Abtv	OilVol	GSV	BAS
10	IVslope1	OilVol	SURP	GSV	OilVol	ATMIV
Panel C. Monthly						
1	OilVol	ATMIV	SPX	VIX	VIX	VIX
2	News	VIX	DEF	ATMIV	ATMIV	News
3	VIX	OilVol	VIX	OilVol	OilVol	OilVol
4	PCR_EQT	HSXvol	Totdvol	News	News	PCR_EQT
5	HSXvol	News	ATMIV	PCR_EQT	PCR_EQT	ATMIV
6	GSV	BAS	SPXETFtdv	SURP_CN	SURP_CN	SURP_EU
7	ATMIV	GSV	STOXX	SPX	SPX	HSXvol
8	Earnnews	PCR_EQT	HYBSpread	SURP_EU	SURP_EU	SURP_CN
9	SURP_EU	Earnnews	UKX	HSXvol	GSV	Earnnews
10	SURP_CN	NIKKEI	IVslope2	FINCON	HSXvol	SURP

Notes. This table reports the significant variables for the out-of-sample prediction of each model and forecasting horizon. The measures of variable importance are presented in Section 4.3. Full results are provided in the Appendix.

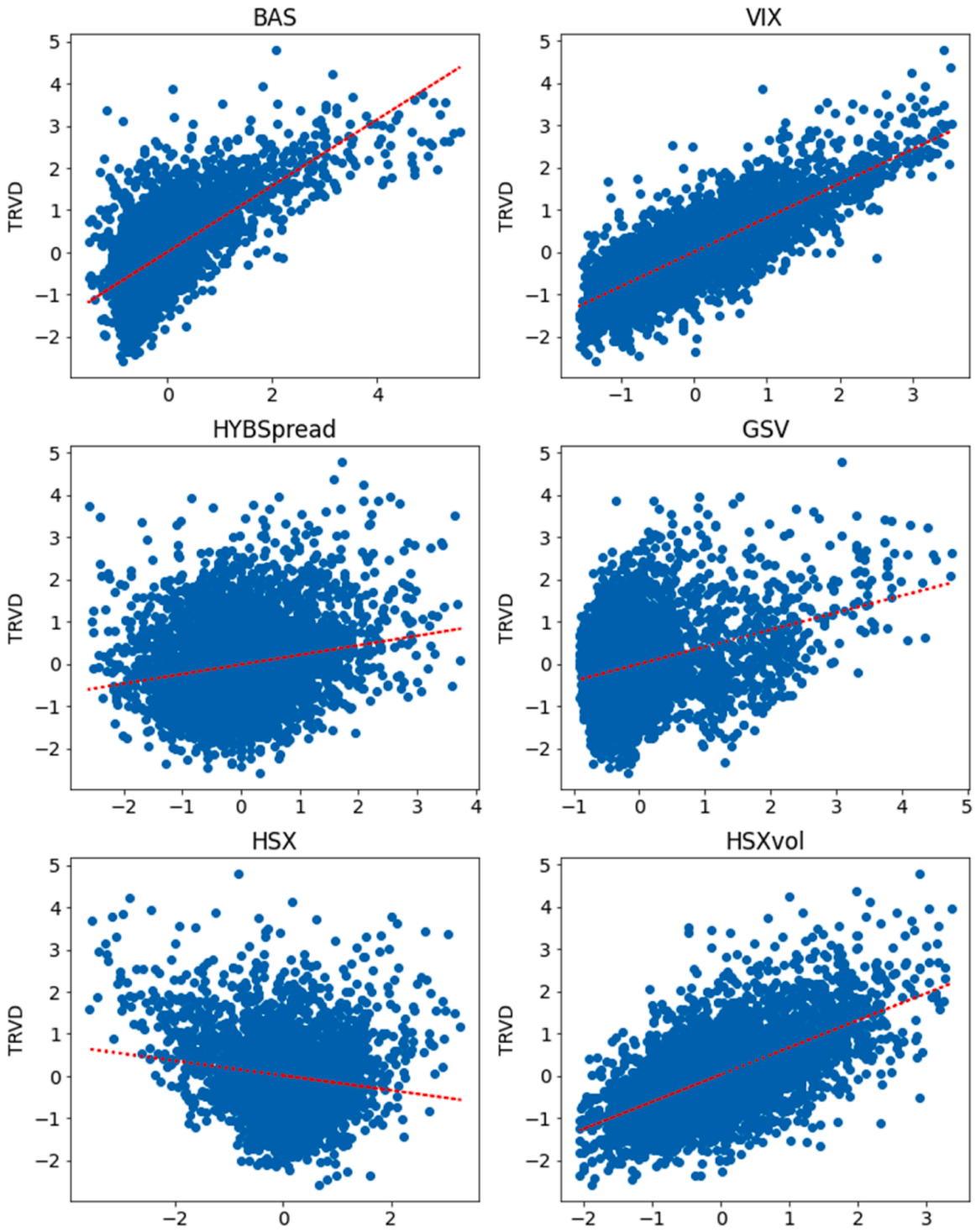

Fig. 5. Scatter plots between one-day-ahead daily S&P 500 RV and key predictor variables. *Notes*. This figure illustrates scatter plots between the one-day-ahead daily S&P 500 RV and bid-ask spread (*BAS*), VIX, high-yield bond spread (*HYBSpread*), Google search volume index (*GSV*), Hang Seng return (*HSX*), and Hang Seng return volatility (*HSXvol*). The red dotted line illustrates the regression line.

4.4. Identifying key predictors

In this subsection, we investigate the predictability of individual predictors and identify common key predictors. To investigate the role of predictors in the prediction model, we apply variable importance measures corresponding to each high-dimensional model, following Chun et al. (2023). For example, for the LASSO, we consider three evaluation indicators: the percentage included (PI), average absolute coefficient (AAC), and an increase in adjusted R^2 values of the encompassing regressions (ΔR_q^2). The PI gauges the number of out-of-sample predictions containing the predictor, calculated as

$$PI_r = \frac{\text{The number of out} - \text{of} - \text{sample predictions that select the predictor}r}{\text{Total number of out} - \text{of} - \text{sample predictions}} \times 100(\%) \tag{30}$$

Second, the average absolute coefficient (AAC) of a predictor is obtained by averaging the absolute coefficients over out-of-sample predictions. Third, we calculate the difference between the adjusted R^2 of the following regressions:

$$\begin{aligned} &(\text{R1}) \quad RV_n = \alpha_1 + \zeta_r \widehat{RV}_{r,n} + \omega_{r,n}, \\ &(\text{R2}) \quad RV_n = \alpha_2 + \varphi_r \widehat{RV}_{r,n} + \varphi_s \widehat{RV}_{s,n} + \upsilon_{s,r,n}, \end{aligned} \tag{31}$$

where $\widehat{RV}_{r,n}$ and $\widehat{RV}_{s,n}$ are the RV forecasts from HARSJ-X model r and the high-dimensional model s, respectively. We define ΔR_q^2 as an increase in the adjusted R^2 of regression (R2) relative to that of regression (R1). Intuitively, smaller ΔR_q^2 signifies that the predictor r plays an important role in the high-dimensional model s forecast. In accordance with the model properties, we determine the mean ranking of the PI, AAC (in descending order), and ΔR_q^2 (in ascending order) for LASSO, AAC and ΔR_q^2 for the ridge regression, that of the PI, feature importance, and ΔR_q^2 for the GBRT, that of the PI, AAC, and ΔR_q^2 for the JMA and MMA, and that of the posterior inclusion probabilities and ΔR_q^2 for the BMA.

Table 9 displays key predictor variables for out-of-sample prediction of high-dimensional models, sorted by corresponding feature importance measures. We discover that the stock market aggregated bid-ask spread (*BAS*), option implied volatility indices (*VIX*, *ATMIV*), high-yield bond spread (*HYBSpread*), Google search volume index (*GSV*), and the Chinese stock market return and their volatility (*HSX* and *HSXvol*) among others are common key predictors.

These findings are in line with previous studies. First, a large bid-ask spread signifies disagreement on asset prices, resulting in high volatility. Rahman et al. (2002) also find that including bid-ask spreads and trading volumes in a GARCH model improved its forecasting performance. Chen et al. (2018) provide further evidence that overall market illiquidity is associated with overall market volatility and can be used as a reliable indicator of economic activity.

Second, informed traders often realize their information through options and futures markets. Therefore, the options and futures market indicators are widely used to predict stock market behaviors. In particular, Eraker and Wu (2017), Griffin and Shams (2017), and Johnson (2017) among others demonstrate that index option implied volatilities have significant predictability on realized volatility.

Third, many studies describe the countercyclicality of financial volatility (Veronesi, 1999; Bansal and Yaron, 2004; Mele, 2007; Brunnermeier and Pedersen, 2008). Conrad and Loch (2015) find that the macroeconomic environment, as measured by the unemployment rate, the term spread, corporate profits, and housing starts, affects long-term stock market risk. Chiu et al. (2018) demonstrate that monthly market volatility is related to both macroeconomics and sentiment. Interest rate spreads such as high-yield bond spreads can serve as a proxy for the default probability, which predicts market downturn (Stock and Watson, 1989; Bernanke, 1990; Friedman and Kuttner, 1998). Consequently, the widening of high-yield bond spread is related to an increase in stock market risk.

Fourth, investor attention can signal the disclosure of material information, leading to significant shifts in asset prices (Andrei and Hasler, 2015; Park et al., 2022; Bang et al., 2023). Additionally, investor sentiment or mood has a significant impact on the stock market and can therefore be used to predict stock volatility (Antweiler and Frank, 2004; Ho et al., 2013; Shu and Chang, 2015; Smith, 2012). In this study, we calculate the sum of the daily Google search volume of keywords 'stock market', 'S&P 500', and 'Nasdaq' to construct the Google search volume index. It gauges investor sentiment and attention regarding the aggregate stock market, which predicts stock market volatility.

Finally, there are two main reasons why Chinese stock market indicators exhibit predictive power for U.S. stock market volatility. First, the increasing globalization of financial markets has led to synchronized movements across global equity markets (Bekaert and Harvey, 1997; Dungey et al., 2005; Karolyi, 2003; King and Wadhwani, 1990). Second, due to their time lag, the Asian stock markets contain information over the time period when the U.S. stock market is closed (i.e., overnight information). Studies including Tsiakas (2008), Ahoniemi and Lanne (2013), and Jayawardena et al. (2016) show that predictability for stock market volatility is significantly improved by incorporating overnight information.

We note that common key predictor variables have a theoretical linear relationship with stock volatility. Fig. 5 shows scatter plots between these variables and the one-day-ahead daily S&P 500 RV. Here, we find that not only the volatility measures but also the other variables have a linear relationship with the target variable.

Each prediction model has distinct properties in dealing with the high-dimensional set of variables. Therefore, their prediction performance depends on the characteristics of the target and explanatory variables. We suppose that the LASSO outperforms the others

because of the sparsity and linearity of the predictor variables. Financial variables are known to have a low signal-to-noise ratio, and studies demonstrate that only a small number of volatility drivers significantly improve out-of-sample prediction performances. If few predictors contain the predictive information and the predictors are linearly correlated with the target variable, the parsimonious linear model is likely to perform well in out-of-sample forecasting, owing to the bias-variance tradeoff. In particular, the prediction bias decreases with the number of explanatory variables as the model's explanatory power increases. However, in finite-sample predictions, additional predictors increase the parameter estimation error. Accordingly, adding extra predictors has both benefits (i.e., bias reduction) and costs (i.e., an increase in the prediction variance), leading to the bias-variance tradeoff for the finite-sample prediction. The LASSO identifies an efficient subset of the explanatory variables and simplifies the model. This reduced model has statistical advantages, owing to its stability and addressable estimates. Accordingly, the LASSO has an advantage in using a sparse set of variables that are linearly correlated with the target variable.

4.5. Robustness checks

The use of a statistical loss function alone can mislead forecasters, and therefore, economic criteria should be applied as well for evaluation of the predictive ability (Becker et al., 2015; Chkili et al., 2014; Jiang et al., 2014). In this section, we conducted supplementary out-of-sample evaluations to assess the predictive power of the volatility forecasts. First, we apply a forecast-encompassing test (Fair and Shiller, 1990). This approach compares the forecasting performance of competing predictions to determine which one is more informative (Kourtis et al., 2016):

Forecast encompassing regression:

$$RV_t = \omega + \sum_{i=s}^{S} \varphi_i \widehat{RV}_{i,t} + \upsilon_t, \tag{32}$$

where S is the total number of the prediction models. We note that if a volatility forecast $\widehat{RV}_{s,t}$ contains no information about the true volatility, the regression coefficient ϕ_s is zero, and if $\widehat{RV}_{s,t}$ contains superior predictive information about the true volatility, ϕ_s would be significant. Table 10 shows the results. In general, LASSO-based forecasts have superior information for the true RVs. In particular, RV forecasts based on LASSO, GBRT, BMS, and MMA (for daily prediction), forecasts based on LASSO and GBRT (for weekly prediction), and forecasts based on the LASSO have significant and superior information about true RVs.

Second, to directly assess the economic value of volatility forecasts, we implement trading strategies using VIX futures. In particular, objective-based evaluation criteria based on the trading strategy can be expressed as follows (Degiannakis and Filis, 2022).

Trading rule 1 (long-only): When the RV forecasts for the next month ($\widehat{RV}_{m,\ n+1}$) are greater than the present monthly RV ($RV_{m,n}$), the trader takes long positions in VIX futures on day n. Subsequently, the return on day n can be calculated as

$$r_{p1,n} = Q_n \times \frac{VIXF_{n+1} - VIXF_n}{VIXF_n}, \tag{33}$$

where $VIXF_n$ is the VIX futures price on day n and $Q_n = \begin{cases} 1 & if \quad \widehat{RV}_{m,\ n+1} \geq RV_{m,n} \\ 0 & if \quad \widehat{RV}_{m,\ n+1} < RV_{m,n} \end{cases}$.

Trading rule 2 (long-short): When the RV forecasts for the next month ($\widehat{RV}_{m,\ n+1}$) are greater (smaller) than the present monthly RV ($RV_{m,n}$), the trader takes long (short) positions in VIX futures on day n. Subsequently, the return on day n can be calculated as

$$r_{p2,n} = Q_n \times \frac{VIXF_{n+1} - VIXF_n}{VIXF_n}, \tag{34}$$

where $Q_n = \begin{cases} 1 & if \quad \widehat{RV}_{m,\ n+1} \geq RV_{m,n} \\ -1 & if \quad \widehat{RV}_{m,\ n+1} < RV_{m,n} \end{cases}$.

We assume the transaction costs per trade to be 0.5 % and the monthly return is limited to 50 % or less to eliminate abnormal situations. The investment period is from October 14, 2009, to December 31, 2020. Table 11 shows the average annual return of the VIX futures portfolio based on each machine-learning model. The columns named "*Long-only*" and "*Long-Short*" represent the performance of long-only (i.e., trading rule 1) and long-short (i.e., trading rule 2) portfolios, respectively. "*Turnover*" represents the number of tradings during the investment period. We find that the machine learning-based portfolios, especially those based on GBRT, Ridge, and LASSO, have significantly higher average returns than the other portfolios.

5. Conclusion

The purpose of this study is to improve the performance of volatility-timing portfolios by forecasting aggregate stock market volatility using machine learning models. The machine learning approaches integrate predictive information contained in a wide range of predictors and provide information-intensive forecasts corresponding to the time-varying market state. Using various performance measures, we find that machine learning models provide accurate forecasts in general. Furthermore, we construct daily, weekly, and monthly volatility-timing portfolios based on volatility forecasts from each prediction model. The asset allocation results suggest that machine learning-based volatility-timing portfolios, especially the LASSO-based portfolios outperform others in general, characterized

Table 10
Forecast encompassing test results.

	Daily		Weekly		Monthly	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
LASSO	0.570 * **	(0.191)	0.714 * **	(0.247)	0.207 * *	(0.086)
Ridge	0.010	(0.031)	−0.387 *	(0.215)	0.121	(0.086)
GBRT	0.480 * **	(0.170)	0.792 * **	(0.234)	0.810	(0.262)
BMA	0.788 * **	(0.212)	−0.435	(0.250)	−0.263	(0.186)
BMS	−0.528	(0.242)	0.298 *	(0.177)	0.030	(0.237)
JMA	−5.222	(1.895)	0.629 *	(0.382)	0.437	(0.242)
MMA	5.130 * **	(1.939)	−0.590	(0.363)	−0.210	(0.116)
DMSPE	−1.042	(0.345)	0.138	(0.382)	0.147	(0.128)
MFC	−0.405	(1.140)	−0.648	(0.376)	−0.414	(0.294)
HAR	1.259 *	(1.109)	0.564	(0.490)	0.433	(0.392)

Notes. This table reports the coefficient of forecast encompassing regression φ_i (*Coef.*) and their standard errors (*S.E.*). * ** , * *, and * denote statistical significance at the 1 %, 5 %, and 10 % levels, respectively. *Daily*, *Weekly*, and *Monthly* represent results for daily, weekly, and monthly volatility forecasts, respectively.

Table 11
Average returns of the VIX futures portfolio.

	Long-only	Long-Short	Turnover
LASSO	15.99	31.82	65
Ridge	17.49	34.81	63
GBRT	19.82	39.48	60
BMA	12.05	23.94	76
BMS	10.12	20.08	76
JMA	11.61	23.05	74
MMA	7.45	14.74	74
DMSPE	5.61	−19.52	63
MFC	4.58	−21.58	65
HAR	3.60	−23.53	65

Notes. This table reports the annualized average returns (%) of the VIX futures portfolios based on the monthly RV forecasts. The columns named "*Long-only*" and "*Long-Short*" represent the long-only and long-short portfolios based on trading rules 1 and 2, respectively. The portfolio construction methods are described in Section 4.4. "*Turnover*" represents the number of tradings during the investment period. The investment period is from October 14, 2009, to December 31, 2020.

by higher average returns, Sharpe ratios, and CERs. The machine learning techniques identify the key factors for the prediction of U.S. stock market volatility, which include the aggregate stock market bid-ask spread, implied volatilities of index options, spreads on high-yield bonds, the Google search volume index, and variables related to the Chinese stock market.

This study has several limitations. First, the analysis does not include deep learning models, which may capture more complex patterns in volatility dynamics. Second, real-world factors such as trading costs and execution timing may constrain the practical applicability of the approach. Third, the economic intuition behind some key variables identified by the models, particularly their causal relationships with volatility, requires further investigation. These limitations suggest promising avenues for future research, including exploring more advanced machine learning techniques, incorporating real-world trading constraints, and analyzing the economic mechanisms driving volatility predictions.

Conflict of interest

Author Dohyun Chun declares that he has no conflict of interest. Author Hoon Cho declares that he has no conflict of interest. Author Doojin Ryu declares that he has no conflict of interest. There is no conflict of interest regarding the publication of this paper.

Funding

No funding.

CRediT authorship contribution statement

Hoon Cho: Conceptualization, Investigation, Methodology, Validation. **Doojin Ryu:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Dohyun Chun:** Data curation,

Formal analysis, Software, Visualization, Writing – original draft.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ribaf.2024.102723.

Data availability

Data will be made available on request.

References

Ahoniemi, K., Lanne, M., 2013. Overnight stock returns and realized volatility. Int. J. Forecast. 29 (4), 592–604.
Aït-Sahalia, Y., Jacod, J., 2014. High-Frequency Financial Econometrics. Princeton University Press.
Amendola, A., Candila, V., Gallo, G.M., 2019. On the asymmetric impact of macro–variables on volatility. Econ. Model. 76, 135–152.
Amendola, A., Braione, M., Candila, V., Storti, G., 2020. A model confidence set approach to the combination of multivariate volatility forecasts. Int. J. Forecast. 36 (3), 873–891.
Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. J. Financ. Mark. 5 (1), 31–56.
Andersen, T.G., Bollerslev, T., Lange, S., 1999. Forecasting financial market volatility: sample frequency vis-a-vis forecast horizon. J. Empir. Financ. 6 (5), 457–477.
Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2006. Volatility and correlation forecasting. Handb. Econ. Forecast. 1, 777–878.
Andrei, D., Hasler, M., 2015. Investor attention and stock market volatility. Rev. Financ. Stud. 28 (1), 33–72.
Antweiler, W., Frank, M.Z., 2004. Is all that talk just noise? The information content of internet stock message boards. J. Financ. 59 (3), 1259–1294.
Audrino, F., Knaus, S.D., 2016. Lassoing the HAR model: a model selection perspective on realized volatility dynamics. Econom. Rev. 35 (8-10), 1485–1521.
Audrino, F., Sigrist, F., Ballinari, D., 2020. The impact of sentiment and attention measures on stock market volatility. Int. J. Forecast. 36 (2), 334–357.
Bang, J., Ryu, D., 2023. CNN-based stock price forecasting by stock chart images. Rom. J. Econ. Forecast. 26 (3), 120–128.
Bang, J., Ryu, D., 2024. ESG factors and the cross-section of expected stock returns: a LASSO-based approach. Financ. Res. Lett. 65, 105482.
Bang, J., Ryu, D., Webb, R.I., 2023b. ESG controversy as a potential asset-pricing factor. Financ. Res. Lett. 58 (Part A), 104315.
Bang, J., Ryu, D., Yu, J., 2023. ESG controversies and investor trading behavior in the Korean market. Financ. Res. Lett. 54, 103750.
Bang, J., Kang, Y., Ryu, D., 2024. Potential pricing factors in the Korean market. Financ. Res. Lett. 67 (Part B), 105946.
Bansal, R., Yaron, A., 2004. Risks for the long run: a potential resolution of asset pricing puzzles. J. Financ. 59 (4), 1481–1509.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. Econometrica 76 (6), 1481–1536.
Barroso, P., Santa-Clara, P., 2015. Momentum has its moments. J. Financ. Econ. 116 (1), 111–120.
Becker, R., Clements, A.E., Doolan, M.B., Hurn, A.S., 2015. Selecting volatility forecasting models for portfolio allocation purposes. Int. J. Forecast. 31 (3), 849–861.
Bekaert, G., Harvey, C.R., 1997. Emerging equity market volatility. J. Financ. Econ. 43 (1), 29–77.
Bernanke, B., 1990. On the predictive power of interest rates and interest rate spreads. New England Economic Review, Federal Reserve Bank of Boston, November-December, 51-68.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econ. 31 (3), 307–327.
Bollerslev, T., Hood, B., Huss, J., Pedersen, L.H., 2018. Risk everywhere: modeling and managing volatility. Rev. Financ. Stud. 31 (7), 2729–2773.
Božović, M., 2024. VIX-managed portfolios. Int. Rev. Financ. Anal. 95, 103353.
Brunnermeier, M.K., Pedersen, L.H., 2008. Market liquidity and funding liquidity. Rev. Financ. Stud. 22 (6), 2201–2238.
Bucci, A., 2020. Realized volatility forecasting with neural networks. J. Financ. Econ. 18 (3), 502–531.
Buckland, S.T., Burnham, K.P., Augustin, N.H., 1997. Model selection: an integral part of inference. Biometrics 53 (2), 603–618.
Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Rev. Financ. Stud. 21 (4), 1509–1531.
Carr, P., Wu, L., 2017. Leverage effect, volatility feedback, and self-exciting market disruptions. J. Financ. Quant. Anal. 52 (5), 2119–2156.
Carr, P., Wu, L., Zhang, Z., 2019. Using machine learning to predict realized variance. arXiv preprint, arXiv:1909.10035. ⟨https://arxiv.org/abs/1909.10035⟩.
Cederburg, S., O'Doherty, M.S., Wang, F., Yan, X.S., 2020. On the performance of volatility-managed portfolios. J. Financ. Econ. 138 (1), 95–117.
Chatziantoniou, I., Degiannakis, S., Filis, G., 2019. Futures-based forecasts: how useful are they for oil price volatility forecasting? Energy Econ. 81, 639–649.
Chen, J., Han, Q., Ryu, D., Tang, J., 2022. Does the world smile together? A network analysis of global index option implied volatilities. J. Int. Financ. Mark. Inst. Money 77, 101497.
Chen, Y., Eaton, G.W., Paye, B.S., 2018. Micro (structure) before macro? The predictive power of aggregate illiquidity for stock returns and economic activity. J. Financ. Econ. 130 (1), 48–73.
Chiang, I.-H.E., Liao, Y., Zhou, Q., 2021. Modeling the cross-section of stock returns using sensible models in a model pool. J. Empir. Financ. 60, 56–73.
Chinco, A., Clark-Joseph, A.D., Ye, M., 2019. Sparse signals in the cross-section of returns. J. Financ. 74 (1), 449–492.
Chiu, C.-W.J., Harris, R.D., Stoja, E., Chin, M., 2018. Financial market volatility, macroeconomic fundamentals and investor sentiment. J. Bank. Financ. 92, 130–145.
Chkili, W., Hammoudeh, S., Nguyen, D.K., 2014. Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. Energy Econ. 41, 1–18.
Christensen, K., Siggaard, M., Veliyev, B., 2023. A machine learning approach to volatility forecasting. J. Financ. Econ. 21 (5), 1680–1727.
Christiansen, C., Schmeling, M., Schrimpf, A., 2012. A comprehensive look at financial volatility prediction by economic variables. J. Appl. Econ. 27 (6), 956–977.
Chun, D., Cho, H., Ryu, D., 2019. Forecasting the KOSPI200 spot volatility using various volatility measures. Phys. A Stat. Mech. Appl. 514, 156–166.
Chun, D., Cho, H., Ryu, D., 2020. Economic indicators and stock market volatility in an emerging economy. Econ. Syst. 44 (2), 100788.
Chun, D., Cho, H., Ryu, D., 2023. Discovering the drivers of stock market volatility in a data-rich world. J. Int. Financ. Mark. Inst. Money 82, 101684.
Chun, D., Kang, J., Kim, J., 2024. Forecasting returns with machine learning and optimizing global portfolios: Evidence from the Korean and US stock markets. Financ. Innov. 10, 124.
Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. J. Econ. 138 (1), 291–311.
Conrad, C., Loch, K., 2015. Anticipating long-term stock market volatility. J. Appl. Econ. 30 (7), 1090–1114.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econ. 7 (2), 174–196.
Dangl, T., Halling, M., 2012. Predictive regressions with time-varying coefficients. J. Financ. Econ. 106 (1), 157–181.
Degiannakis, S., Filis, G., 2017. Forecasting oil price realized volatility using information channels from other asset classes. J. Int. Money Financ. 76, 28–49.
Degiannakis, S., Filis, G., 2022. Oil price volatility forecasts: what do investors need to know? J. Int. Money Financ. 123, 102594.
Duan, H., Zhao, C., Wang, L., Liu, G., 2024. The relationship between renewable energy attention and volatility: a HAR model with Markov time-varying transition probability. Res. Int. Bus. Financ. 71, 102437.
Dungey, M., Fry, R., González-Hermosillo, B., Martin, V.L., 2005. Empirical modelling of contagion: a review of methodologies. Quant. Financ. 5 (1), 9–24.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 987–1007.
Eraker, B., Wu, Y., 2017. Explaining the negative returns to volatility claims: an equilibrium approach. J. Financ. Econ. 125 (1), 72–98.
Fair, R.C., Shiller, R.J., 1990. Comparing information in forecasts from econometric models. Am. Econ. Rev. 80 (3), 375–389.
Fernandes, M., Medeiros, M.C., Scharth, M., 2014. Modeling and predicting the CBOE market volatility index. J. Bank. Financ. 40, 1–10.
Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility-timing. J. Financ. 56 (1), 329–352.
Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility-timing using "realized" volatility. J. Financ. Econ. 67 (3), 473–509.
Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Rev. Financ. Stud. 33 (5), 2326–2377.
Friedman, B.M., Kuttner, K.N., 1998. Indicator properties of the paper—bill spread: lessons from recent experience. Rev. Econ. Stat. 80 (1), 34–44.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.
Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74 (6), 1545–1578.
Griffin, J.M., Shams, A., 2017. Manipulation in the VIX? Rev. Financ. Stud. 31 (4), 1377–1417.
Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33 (5), 2223–2273.
Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. J. Econ. 222 (1), 429–450.
Ham, H., Cho, H., Kim, H., Ryu, D., 2019. Time-series momentum in China's commodity futures market. J. Futures Mark. 39 (12), 1515–1528.
Han, H., Kutan, A.M., Ryu, D., 2015. Effects of the US stock market return and volatility on the VKOSPI. Economics 9 (1), 20150035.
Hansen, B.E., 2007. Least squares model averaging. Econometrica 75 (4), 1175–1189.
Hansen, B.E., Racine, J.S., 2012. Jackknife model averaging. J. Econ. 167 (1), 38–46.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.
Ho, K.-Y., Shi, Y., Zhang, Z., 2013. How does news sentiment impact asset volatility? Evidence from long memory and regime-switching approaches. North Am. J. Econ. Financ. 26, 436–456.
Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12 (1), 55–67.
Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14 (4), 382–417.
Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. Stoch. Process. Appl. 119 (7), 2249–2276.
Jayawardena, N.I., Todorova, N., Li, B., Su, J.-J., 2016. Forecasting stock volatility using after-hour information: evidence from the Australian Stock Exchange. Econ. Model. 52, 592–608.
Jiang, I.M., Hung, J.C., Wang, C.S., 2014. Volatility forecasts: do volatility estimators and evaluation methods matter? J. Futures Mark. 34 (11), 1077–1094.
Johnson, T.L., 2017. Risk premia and the VIX term structure. J. Financ. Quant. Anal. 52 (6), 2461–2490.
Karolyi, G.A., 2003. Does international financial contagion really exist? Int. Financ. 6 (2), 179–199.
Kim, D., Wang, Y., Zou, J., 2016. Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. Stoch. Process. Appl. 126 (11), 3527–3577.
Kim, H., Cho, H., Ryu, D., 2021a. Forecasting consumer credit recovery failure: classification approaches. J. Credit Risk 17 (3), 117–140.
Kim, H., Cho, H., Ryu, D., 2021b. Predicting corporate defaults using machine learning with geometric-lag variables. Invest. Anal. J. 50 (3), 161–175.
Kim, H., Cho, H., Ryu, D., 2022. Corporate bankruptcy prediction using machine learning methodologies with a focus on sequential data. Comput. Econ. 59, 1231–1249.
Kim, J., Park, Y.J., Ryu, D., 2017. Stochastic volatility of the futures prices of emission allowances: a Bayesian approach. Phys. A Stat. Mech. Appl. 465, 714–724.
Kim, J., Park, Y.J., Ryu, D., 2018. Testing CEV stochastic volatility models using implied volatility index data. Phys. A Stat. Mech. Appl. 499, 224–232.
Kim, K., Ryu, D., 2020. Predictive ability of investor sentiment for the stock market. Rom. J. Econ. Forecast. 23 (4), 33–46.
Kim, K., Ryu, D., 2021. Does sentiment determine investor trading behaviour? Appl. Econ. Lett. 28 (10), 811–816.
Kim, K., Ryu, D., 2022. Sentiment changes and the Monday effect. Financ. Res. Lett. 47 (Part B), 102709.
Kim, K., Ryu, D., Yu, J., 2021. Do sentiment trades explain investor overconfidence around analyst recommendation revisions? Res. Int. Bus. Financ. 56, 101376.
Kim, K., Ryu, D., Yang, H., 2021. Information uncertainty, investor sentiment, and analyst reports. Int. Rev. Financ. Anal. 77, 101835.
Kim, K., Ryu, D., Yu, J., 2022. Is a sentiment-based trading strategy profitable? Invest. Anal. J. 51 (2), 94–107.
King, M.A., Wadhwani, S., 1990. Transmission of volatility between stock markets. Rev. Financ. Stud. 3 (1), 5–33.
Koop, G.M., 2003. Bayesian econometrics. John Wiley & Sons Inc.
Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. J. Empir. Financ. 12 (3), 445–475.
Kourtis, A., Markellos, R.N., Symeonidis, L., 2016. An international comparison of implied, realized, and GARCH volatility forecasts. J. Futures Mark. 36 (12), 1164–1193.
Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. J. Financ. Econ. 135 (2), 271–292.
Lee, G., Ryu, D., 2018. Asymmetry in the stock price response to macroeconomic shocks: evidence from the Korean market. J. Bus. Econ. Manag. 19 (2), 343–359.
Lee, G., Ryu, D., 2024. Investor sentiment or information content? A simple test for investor sentiment proxies. North Am. J. Econ. Financ. 74, 102222.
Lee, J., Batten, J.A., Ham, H., Ryu, D., 2024. Does portfolio momentum beat analyst advice? Abacus 60 (2), 338–364.
Lee, J.E., Cho, H., Lee, G., Ryu, D., 2023. Does performance-chasing behavior matter? International evidence. J. Multinatl. Financ. Manag. 68, 100799.
Luong, C., Dokuchaev, N., 2018. Forecasting of realised volatility with the random forests algorithm. J. Risk Financ. Manag. 11 (4), 61.
Ma, T., Tee, K.H., Li, B., 2022. Timing the volatility risk of beta anomaly: evidence from hedge fund strategies. Int. Rev. Financ. Anal. 81, 102095.
Marcellino, M., Stock, J.H., Watson, M.W., 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. J. Econ. 135 (1-2), 499–526.
Martens, M., 2002. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. J. Futures Mark. 22 (6), 497–518.
Mele, A., 2007. Asymmetric stock market volatility and the cyclical behavior of expected returns. J. Financ. Econ. 86 (2), 446–478.
Mittnik, S., Robinzonov, N., Spindler, M., 2015. Stock market volatility: identifying major drivers and the nature of their impact. J. Bank. Financ. 58 (1), 1–14.
Moreira, A., Muir, T., 2017. Volatility-managed portfolios. J. Financ. 72 (4), 1611–1644.
Nagel, S., 2012. Evaporating liquidity. Rev. Financ. Stud. 25 (7), 2005–2039.
Niu, Z., Wang, C., Zhang, H., 2023. Forecasting stock market volatility with various geopolitical risks categories: new evidence from machine learning models. Int. Rev. Financ. Anal. 89, 102738.
Nonejad, N., 2017. Forecasting aggregate stock market volatility using financial and macroeconomic predictors: which models forecast best, when and why? J. Empir. Financ. 42, 131–154.
Park, D., Ryu, D., 2021a. A machine learning-based early warning system for the housing and stock markets. IEEE Access 9, 85566–85572.
Park, D., Ryu, D., 2021b. Forecasting stock market dynamics using bidirectional long short-term memory. Rom. J. Econ. Forecast. 24 (2), 22–34.
Park, H., Kim, M., Ryu, D., 2022. Heterogeneous investor attention to climate risk: evidence from a unique dataset. Invest. Anal. J. 51 (4), 253–267.
Park, S.Y., Ryu, D., Song, J., 2017. The dynamic conditional relationship between stock market returns and implied volatility. Phys. A Stat. Mech. Appl. 482, 638–648.
Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: signed jumps and the persistence of volatility. Rev. Econ. Stat. 97 (3), 683–697.
Paye, B.S., 2012. Déjà vol': predictive regressions for aggregate stock market volatility using macroeconomic variables. J. Financ. Econ. 106 (3), 527–546.
Pezzo, L., Wang, L., Zirek, D., 2023. Large scale mean-variance strategies in the US stock market. Res. Int. Bus. Financ. 66, 102062.
Qiao, W., Bu, D., Gibberd, A., Liao, Y., Wen, T., Li, E., 2023. When "time varying" volatility meets "transaction cost" in portfolio selection. J. Empir. Financ. 73, 220–237.
Qu, H., Zhang, Y., 2022. Asymmetric multivariate HAR models for realized covariance matrix: a study based on volatility-timing strategies. Econ. Model. 106, 105699.
Rahman, S., Lee, C.-F., Ang, K.P., 2002. Intraday return volatility process: evidence from NASDAQ stocks. Rev. Quant. Financ. Account. 19, 155–180.
Ryu, D., Yu, J., 2022. Sentiment-dependent impact of funding liquidity shocks on futures market liquidity. J. Futures Mark. 42 (1), 61–76.

Ryu, D., Webb, R.I., Yu, J., 2022. Foreign institutions and the behavior of liquidity following macroeconomic announcements. Financ. Res. Lett. 50, 103239.
Ryu, D., Ryu, D., Yang, H., 2023a. Whose sentiment explains implied volatility change and smile? Financ. Res. Lett. 55 (Part A), 103838.
Ryu, D., Ryu, D., Yang, H., 2023b. Investor sentiment and futures market mispricing. Financ. Res. Lett. 58 (Part C), 104559.
Samuels, J.D., Sekkel, R.M., 2017. Model confidence sets and forecast combination. Int. J. Forecast. 33 (1), 48–60.
Seok, S., Cho, H., Ryu, D., 2022. Scheduled macroeconomic news announcements and intraday market sentiment. North Am. J. Econ. Financ. 62, 101739.
Seok, S., Cho, H., Lee, J.E., Ryu, D., 2023. Indirect effects of flow-performance sensitivity on fund performance. Borsa Istanb. Rev. 23 (S1), S1–S14.
Seok, S., Cho, H., Ryu, D., 2024. Dual effects of investor sentiment and uncertainty in financial markets. Q. Rev. Econ. Financ. 95, 300–315.
Shapiro, A.H., Sudhof, M., Wilson, D.J., 2022. Measuring news sentiment. J. Econom. 228 (2), 221–243.
Shim, H., Kim, H., Kim, J.Y., Ryu, D., 2015. Weather and stock market volatility: the case of a leading emerging market. Appl. Econ. Lett. 22 (12), 987–992.
Shu, H.-C., Chang, J.-H., 2015. Investor sentiment and financial market volatility. J. Behav. Financ. 16 (3), 206–219.
Siganos, A., Vagenas-Nanos, E., Verwijmeren, P., 2017. Divergence of sentiment and stock market trading. J. Bank. Financ. 78, 130–141.
Smith, G.P., 2012. Google Internet search activity and volatility prediction in the market for foreign currency. Financ. Res. Lett. 9 (2), 103–110.
Song, W., Ryu, D., Webb, R.I., 2016. Overseas market shocks and VKOSPI dynamics: a Markov-switching approach. Financ. Res. Lett. 16, 275–282.
Song, W., Ryu, D., Webb, R.I., 2018. Volatility dynamics under an endogenous Markov-switching framework: a cross-market approach. Quant. Financ. 18 (9), 1559–1571.
Song, Y., Lei, B., Tang, X., Li, C., 2024. Volatility forecasting for stock market index based on complex network and hybrid deep learning model. J. Forecast. 43 (3), 544–566.
Steel, M.F., 2020. Model averaging and its use in economics. J. Econ. Lit. 58 (3), 644–719.
Stock, J.H., Watson, M.W., 1989. New indexes of coincident and leading economic indicators. NBER Macroecon. Annu. 4, 351–394.
Tan, Y., Tan, Z., Tang, Y., Zhang, Z., 2024. Functional volatility forecasting. J. Forecast. 43 (5), 1129–1730.
Taylor, N., 2023. The determinants of volatility-timing performance. J. Financ. Econ. 21 (4), 1228–1257.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodol.) 58 (1), 267–288.
Tsiakas, I., 2008. Overnight information and stochastic volatility: a study of European and US stock exchanges. J. Bank. Financ. 32 (2), 251–268.
Veronesi, P., 1999. Stock market overreactions to bad news in good times: a rational expectations equilibrium model. Rev. Financ. Stud. 12 (5), 975–1007.
Vrontos, S.D., Galakis, J., Vrontos, I.D., 2021. Implied volatility directional forecasting: a machine learning approach. Quant. Financ. 21 (10), 1687–1706.
Wang, Y., Ma, F., Wei, Y., Wu, C., 2016. Forecasting realized volatility in a changing world: a dynamic model averaging approach. J. Bank. Financ. 64, 136–149.
Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21 (4), 1455–1508.
Xing, Y., Zhang, X., Zhao, R., 2010. What does the individual option volatility smirk tell us about future equity returns? J. Financ. Quant. Anal. 45 (3), 641–662.
Yin, Z., Jiang, J., Qian, Z., 2023. How does the volatility-timing strategy perform in mutual funds portfolios. Int. Rev. Financ. 23 (1), 87–102.
Yu, J., Ryu, D., 2021. Effectiveness of the Basel III framework: procyclicality in the banking sector and macroeconomic fluctuations. Singap. Econ. Rev. 66 (3), 855–879.
Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observative: a multi-scale approach. Bernoulli 12 (6), 1019–1043.
Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Am. Stat. Assoc. 100 (472), 1394–1411.