North American Journal of Economics and Finance 64 (2023) 101845

Contents lists available at ScienceDirect

North American Journal of Economics and Finance

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

Forecasting stock return volatility in data-rich environment: A new powerful predictor

Zhifeng Dai [a,b,*], Xiaotong Zhang [a], Tingyu Li [a]

[a] School of Mathematics and Statistics, Changsha University of Science and Technology, China
[b] Hunan Provincial Key Laboratory of Mathematical Modeling and Analysis in Engineering, Changsha 410114, Hunan, China

ARTICLE INFO

Keywords:
Partial least squares approach
Stock return volatility
Out-of-sample forecast
Asset allocation

ABSTRACT

We forecast stock return volatility by using the partial least squares approach that extract a powerful predictor from data-rich environment. Empirical results indicate that the new index has superior out-of-sample forecasting performance than the existing indexes, and the discovery is consistent with the in-sample predictive power. Specifically, the application of the new-index is extended to the allocation of investment portfolios to support mean–variance investors obtain considerable economic gains. In addition, our results are robust to various checks. Overall, our findings confirm that the partial least squares approach can effectively improve stock return volatility forecasts in a data-rich environment, successfully outperforming the competitive models and far surpassing the benchmark model.

1. Introduction

Predicting stock returns volatility in financial markets (see, for example, Corsi, 2009, Asgharian, Hou, & Javed, 2013; Choudhry, Papadimitriou, & Shabi, 2016; Nonejad, 2017; Paye, 2012) seems to have received much more attention than predicting stock returns (e.g., Zhang, Zeng, Ma, and Shi (2019); Welch and Goyal (2008); Lustig, Roussanov, and Verdelhan (2014)) recently. This phenomenon can be explained by the fact that the prediction of stock return volatility plays an essential role in many financial fields, such as portfolio allocation, risk management and asset pricing.

The seminal work by Schwert (1989) demonstrates that sources of financial volatility are closely related to real and nominal macroeconomic volatility, economic activity and financial leverage. However, recent papers by Paye (2012) and Christiansen, Schmeling, and Schrimpf (2012) show that although some variables such as treasury spread and default returns can theoretically affect stock volatility, it is difficult to find an individual variable that can predict stock volatility exceed that of benchmark models. In addition, Nonejad (2017) shows that the predictive accuracy of most individual variables fails to exceed that of benchmark models because of parameter instability and model uncertainty. As practitioners involved in asset allocation and risk management are very interested in out-of-sample forecasting results, it is important to provide a methodology that can significantly improve these forecasts. Therefore, we need a forecasting method that can extract powerful predictors of stock return volatility in a data-rich environment and obtain more accurate forecasting results.

From a behavioral finance perspective, Baker and Wurgler (2006, 2007) and Huang, Jiang, Tu, and Zhou (2015) focused on verifying that investor sentiment variables are predictors for predicting stock return. The principal component extracted by Baker and

* Corresponding author.
E-mail address: zhifengdai823@163.com (Z. Dai).

https://doi.org/10.1016/j.najef.2022.101845
Received 1 April 2021; Received in revised form 8 September 2022; Accepted 17 November 2022
Available online 21 November 2022

Wurgler (2006, 2007) through principal component analysis is the first principal component which is the best combination of all the proxies. Since all agents may have approximation errors that are part of their variation, the first principal component may contain a large number of common approximation errors (Boivin & Ng, 2006; Bai & Ng, 2008), which can affect the predictive power of the first principal component.

Given this, we use the partial least squares (PLS) as an alternative approach to predict stock returns volatility. The PLS method pioneered by Wold (1966, 1975) can construct a real but unobservable factor to link a series of variables and reasonably explain stock return. So (2013) indicates that it is necessary to deal with errors correctly, as investors will be particularly cautious about the problem of predictable errors from researcher forecasts. Kelly and Pruitt's (2013, 2015) extended the PLS method to construct the optimum PLS-index by efficiently weighting and combining each of the individual variables in accordance with their ability to predict future returns. The greatest advantage of PLS-index is that it can effectively alleviate the impact of error on the forecasting process without considering the fundamental source of error. Another important reason why the PLS approach is superior to other methods is that information classified as an approximation error is not discarded. Using this approach, Huang et al. (2015) show that new indicators developed by the PLS approach are more efficient than those developed using the Baker and Wurgler (2006, 2007) approach when forecasting stock returns.

However, there is no literature examining the predictive role of the PLS method on stock return volatility. To fill this research gap, this paper uses the PLS method to construct new predictors to forecast stock return volatility. This paper uses monthly data ranging from January 1991 to December 2018, for a series of variables, including 5 investor sentiment indexes, 14 macroeconomic fundamental indicators, and 18 technical indicators. The investor sentiment indexes are in Baker and Wurgler (2006, 2007). The macroeconomic variables are proposed by Welch and Goyal (2008). Technical indicators are constructed using stock return prices and stock counts. In this paper, a PLS approach is used to extract powerful predictors of stock return volatility from 5 investor sentiment indexes, 14 macroeconomic variables and 18 technical indicators, respectively, and information on all variables is combined by PLS_ALL. Through in-sample and out-of-sample empirical analysis and robustness tests, we find that the predictors constructed with PLS significantly predict stock return volatility compared to other predictors.

The empirical results of this paper are presented below. Firstly, in the univariate regression model, most of the predictors are insignificant for both in-sample and out-of-sample forecasts, which is due to the limited information contained in the univariate regression model, resulting in its weaker forecasting power. In in-sample analysis, the PLS index is a strong predictor of stock return volatility and the in-sample predictive effect of the PLS index is much greater than that of other predictors, including the equal weights index (EW) and individual indexes. In the out-of-sample analysis, the PLS index (PLS_ALL) extracted from all variables by the PLS method has the best out-of-sample forecasts with a R_{os}^2 of 14.743 %. The PLS_ALL index outperformed the new macroeconomic variables index (PLS_ECON) and the new investor sentiment index (PLS_UMS) in terms of forecasting power in a data-rich environment. In summary, the PLS index shows strong in-sample and out-of-sample predictive performance.

Secondly, this paper analyses the economics of the PLS index from an asset allocation perspective. A mean–variance investor using the PLS approach to asset allocation between stocks and risk-free treasury bills achieves positive growth in the portfolio's certainty equivalent returns gain (CER gain), compared to negative growth in the univariate regression model and the portfolio approach. The PLS index generated by the PLS approach yields more accurate forecasts than other predictors. Specifically, the PLS_ALL index based on all predictors obtained a CER of 16.177 %, which is the highest increase in CER of almost all models. The growth in CER reveals that the PLS index based on a data-rich environment is not only statistically significant, but also economically significant when the risk aversion coefficient is three.

Finally, the paper tests the robustness of the empirical results to business cycles, forecast window sizes, risk aversion coefficients and transaction costs. The results show that macroeconomic variables, investor sentiment indicators, technical indicators and the PLS-index have better predictive power in recession than expansion. In particular, the PLS index has strong predictive power in both expansion and recession periods, and the PLS index's predictive power being stronger during the more recent period. Although CER gains vary with different risk aversion coefficients, all indexes based on the PLS methodology obtain greater CER gains than other predictors.

The main contributions of this paper are as follows. Firstly, we use the PLS approach to forecast stock return volatility. Unlike other methods, the PLS approach isolates the approximation error that is unrelated to predicting stock return volatility. The empirical evidence shows that the new PLS index is a powerful predictor of stock return volatility, which makes an important contribution to the literature on predicting stock return volatility. Secondly, in addition to analyzing the predictive validity of the predictors constructed using the PLS method from a statistical perspective, this paper also analyses the economic benefits of the predictors constructed using the PLS method from an economic perspective by calculating CER gains.

The remainder of the paper is structured as follows. Section 2 describes the data and their sources. Section 3 describes the methodology used in this paper. Section 4 shows the in-sample empirical results. Section 5 shows the out-of-sample empirical results. Section 6 conducts the asset allocation analysis. Section 7 presents the robustness checks. Section 8 concludes the full paper.

2. Data

2.1. Stock return volatility

Following Nonejad (2017) and Ma, Liu, Wahab, and Zhang (2018), we use the sum of the squares of the daily return volatilities of the S&P 500 to measure the monthly return volatilities of stocks.

$$RV_t = \sum_{i=1}^{N_t} R_{i,t}^2 \tag{1}$$

$$R_{i,t} = \log(P_{i,t}) - \log(P_{i-1,t}) \tag{2}$$

where N_t denotes the total number of trading days in the stock market in month t, $R_{i,t}$ is the daily stock return of the S&P 500 index on the i th trading day in month t, and $P_{i,t}$ denotes the daily price of the S&P 500 index on the i th trading day in month t. Following Paye (2012), we use the natural logarithm of monthly stock return volatility, defined as $V_t = log(RV_t)$. It is worth noting that OLS-based statistical analysis is inaccurate when the errors are non-normally distributed. According to the arguments of Andersen, Bollerslev, Diebold, and Ebens (2001), the distribution of V_t is approximately normal distribution.

2.2. Investor sentiment indexes

Investor sentiment indexes are widely used in the study of stock markets, and they have a significant impact on stock market volatility. In this paper, five investor sentiment indexes[1] proposed by Baker and Wurgler (2006, 2007) are as following:

1. Dividend premium (PDND): we use the first six-month moving average of value weighted dividend premium constructed by Baker and Wurgler (2006) to smooth the noise;
2. First-day returns of IPOs (RIPO): the data is defined as the monthly weighted average of the average first day returns of IPO shares in the past 12 months;
3. Number of IPOs (NIPO): the term is described as a monthly moving average of the number of IPOs in the past 12 months;
4. Close-end fund discount rate (CEFD): value-weighted average difference between the net asset values of closed-end stock mutual fund shares and their market prices;
5. Stock shares in new issues (EQTI): it is interpreted as the total monthly stock issuance divided by the sum of the current month's stock and debt issuance, and uses the moving average of the previous six months.

2.3. Macroeconomic variables

We use the 14 macroeconomic variables proposed byWelch and Goyal (2008)[2] to examine monthly stock return volatility, which include Dividend-price ratio (DP), Dividend yield (DY), Earnings-price ratio (EP), Dividend-payout ratio (DE), Stock risk premium volatility (RVOL), Book-to-market ratio (BM), Net stock expansion (NTIS), Treasury bill rate (TBL), Long-term yield (LTY), Long-term return (LTR), Term spread (TMS), Default yield spread (DFY), Default return spread (DFR) and Inflation (INFL).

2.4. Technical indicators

Technical indicators have received more attention from academics than macroeconomic variables when it comes to predicting stock volatility. We use past prices or volumes as a basis to identify future trends. Much of the literature examines momentum strategies (e.g. Fuertes, Miffre, & Rallis, 2010; Moskowitz, Ooi, & Pedersen, 2012), moving averages and channel strategies (e.g. Szakmary, Shen, & Sharma, 2010) in stock futures trading.

We use 18 technical indicators based on three popular trend following trading strategies. The first strategy is based on the Moving Average (MA) rule. It produces a buy signal ($S_{i,t} = 1$) when the short-term moving average stock price exceeds the long-term and vice versa to generate a sell signal ($S_{i,t} = 0$).

$$S_{i,t} = \begin{cases} 1 & if MA_{s,t} \geq MA_{i,t} \\ 0 & if MA_{s,t} < MA_{i,t} \end{cases} \tag{3}$$

$$MA_{j,t} = \left(\frac{1}{j}\right) \sum_{i=0}^{j-1} P_{t-i} \; for \; j = s, l \tag{4}$$

where P_t is the level of a stock price index, $s(l)$ is the length of the short (long) $MA(s < l)$. In fact, short MA will detect recent price movement, because the MA rules imply stock price trends. We research monthly MA rules with $s = 1, 2, 3$ and $l = 9, 12$.

The second strategy is based on momentum (MOM) rules and generates the following signals:

$$S_{i,t} = \begin{cases} 1 & if \quad P_t \geq P_{t-m} \\ 0 & if \quad P_t < P_{t-m} \end{cases} \tag{5}$$

In the rule, a sell signal indicates that stock price is lower than its level of m periods ago, which implies "negative" momentum and low stock returns. We compute monthly signals for $m = 1, 2, 3, 6, 9, 12$.

The last strategy is based on the trading volume (VOL) rule. Volume data is rarely used to identify market trends, but it is a useful

[1] https://apps.olin.wustl.edu/faculty/zhou/zworkingpapers.html.
[2] The data can be downloaded from the following websites: https://www.hec.unil.ch/agoyal/ (Welch & Goyal, 2008).

measure. We begin by defining:

$$OBV_t = \sum\nolimits_{k=1}^{t} VOL_k D_k \tag{6}$$

where VOL_k is a measure of the trading volume during period k. When $P_k - P_{k-1} \geq 0$, the value of D_k is 1, otherwise -1. The following define OBV as:

$$S_{i,t} = \begin{cases} 1 & if MA_{s,t}^{OBV} \geq MA_{i,t}^{OBV} \\ 0 & if MA_{s,t}^{OBV} < MA_{i,t}^{OBV} \end{cases} \tag{7}$$

$$MA_{j,t}^{OBV} = \left(\frac{1}{j}\right) \sum\nolimits_{i=0}^{j-1} OBV_{t-i} \ for\ j = s, l \tag{8}$$

A relatively high trading volume with rising prices in a recent period usually forms a buy signal, which is a positive market trend. The corresponding VOL indicator with VOL lengths of s and l is defined as VOL(s, l). I calculate the monthly signals for $s = 1, 2, 3$ and $l = 9, 12$.

The technical indicators based on the MA, momentum and volume rules proposed by Sullivan, Timmermann, and White (1999) represent the volatility trends in the stock market. The technical indicators in equations (3), (5) and (7) are obtained by using daily data and monthly volume data of the S&P 500. In addition, given the lag rule and availability of technical indicators, all data are from January 1990.

3. Forecasting and evaluation methods

3.1. Predictive models

Conrad and Loch (2015) show that realized volatility is highly persistent, i.e. the aggregation effect of volatility. This implies that past volatilities contain predictive information about future volatilities. According to Wang, Wei, Wu, and Yin (2018), it is reasonable to apply an autoregressive model (AR) as the standard predictive regression framework for forecasting stock volatility. The benchmark AR model is defined as follows:

$$V_{t+1} = \omega + \sum\nolimits_{i=0}^{p-1} \alpha_i V_{t-i} + \varepsilon_{t+1} \tag{9}$$

where V_{t+1} is t + 1-th month stock volatility, V_{t-i} is i + 1-th lag of V_{t+1}. ω and α_i are constant coefficient and lag coefficient respectively, ε_{t+1} is independent error term and conforms to identically normal distribution. According to Paye (2012), we set the lag order p to 6 when our data are monthly fluctuations, as this length allows us to obtain autocorrelated information on stock volatility.

To investigate the impact of other predictors on stock volatility, we added exogenous regressors, i.e. other individual predictors, to equation (9) to obtain the following model:

$$V_{t+1} = \omega + \sum\nolimits_{i=0}^{p-1} \alpha_i V_{t-i} + \beta X_t + \varepsilon_{t+1} \tag{10}$$

where X_t represents predictors in t-th month that affect stock volatility. The influence of other factors on the stock market volatility is reflected in the parameter β. We use least squares estimation (OLS) to estimate the coefficients in (10). In a standard regression framework, the null hypothesis of unpredictability, $\beta = 0$, can be tested using the standard t-statistic. To address the possibility of serial correlation in the sample data, Newey-West covariance correction is used in the calculation.

We use recursive estimation windows to generate out-of-sample stock volatility forecasts. Firstly, the overall length of the data sample is denoted as T. The first M samples are then used as the initial training samples and the remaining $T - M$ samples are used as the test samples. Using the recursive window estimation method, the first out-of-sample stock volatility prediction can be given by the following equation:

$$\widehat{V}_{M+1} = \widehat{\omega}_M + \sum\nolimits_{i=0}^{p-1} \widehat{\alpha}_{i,M} V_{M-i} + \widehat{\beta}_M X_M \tag{11}$$

where $\widehat{\omega}_M$, $\widehat{\alpha}_{i,M}(i = 1, 2, ..., p)$and $\widehat{\beta}_M$ are the least squares regression estimates of model (10) for ω, $\alpha_i(i = 1, 2, ..., p)$ and β. The estimation results are obtained from regressing a constant, $\{V_t\}_{t=i}^{M-p+i-1}(i = 1,2,...,p)$ and $\{X_t\}_{t=p}^{M-1}$ on $\{V_{t+1}\}_{t=p}^{M}$. By adding one sample to the initial sample in order, the second out-of-sample stock volatility forecast can be given by:

$$\widehat{V}_{M+2} = \widehat{\omega}_{M+1} + \sum\nolimits_{i=0}^{p-1} \widehat{\alpha}_{i,M+1} V_{M-i+1} + \widehat{\beta}_{M+1} X_{M+1} \tag{12}$$

Similarly, using least squares regression estimation to calculate the parameters $\widehat{\omega}_{M+1}$, $\widehat{\alpha}_{i,M+1}(i = 1,2,...,p)$ and the value of $\widehat{\beta}_{M+1}$, and then use the sample from period $M+1$ to obtain the forecast value for period $M + 2$. Continuing with the above steps, the length of the test sample is increased sequentially to gradually generate all remaining stock volatility forecasts.

3.2. Multivariate predictive analysis

3.2.1. Portfolio forecast

In this section, we use a portfolio forecasting approach to integrate useful information from univariate forecasts. Bates and Granger (2001) suggest that portfolio forecasts can exhibit better forecasting performance than original forecasts. Portfolio forecasting has attracted a lot of attention. According to Rapach, Strauss, and Zhou (2010), portfolio forecasting can produce more reliable forecasts by reducing mean-squared forecast errors. However, there is little literature examining the impact of a range of predictors on stock volatility. Our paper focuses on portfolio forecasting of macroeconomic variables, technical indicators, etc. to improve the forecasting of stock volatility.

Statistically, the combination forecasts of V_t are calculated as weighted averages of the N individual forecasts at month t, which is expressed as:

$$\widehat{V}_{c,t+1} = \sum_{i=1}^{N} \omega_{i,t} \widehat{V}_{i,t+1} \tag{13}$$

where $\{\omega_{i,t}\}_{i=1}^{N}$ are defined as the combining weight of the i-th univariate forecast computed at t-th month.

Following Pettenuzzo, Timmermann, and Valkanov (2014) and Rapach et al. (2010), we consider five popular combining methods: mean, median, trimmed mean, DMSPE(1), and DMSPE(0.9). The mean combination forecast is the mean of $\{\widehat{V}_{i,t+1}\}_{i=1}^{N}$ and takes $\omega_{i,t} = \frac{1}{N}(i = 1, \cdots, N)$ for the individual predictions. The median combining methods takes the median of$\{\widehat{V}_{i,t+1}\}_{i=1}^{N}$. The trimmed mean combination forecast sets the weight of the maximum and minimum to zero and sets $\omega_{i,t} = \frac{1}{(N-2)}$) for the remainder of the individual predictions. We set the combining weights of the discount mean squared prediction error (DMSPE) in i-th individual forecast at t-th period as:

$$\omega_{i,t} = \frac{\phi_{i,t}^{-1}}{\sum_{j=1}^{N} \phi_{j,t}^{-1}} \tag{14}$$

$$\phi_{i,t} = \sum_{s=M}^{t-1} \theta^{t-1-s} \left(V_{s+1} - \widehat{V}_{i,s+1}\right)^2 \tag{15}$$

where θ is a discount factor and M is the length of initial estimation period. We should assign greater weights to individual predictions with lower MSPE values, so the DMSPE method can have better predictive performance in out-of-sample periods. When $\theta = 1$, there is no discounting, and generates the optimal combination forecast derived by Bates and Granger (2001). When $\theta < 1$, we find that larger weight is attached to the more accurate individual predictive regression. Following Rapach et al. (2010) and Zhang et al. (2019), we employ 1 and 0.9 for θ in the DMSPE methods.

3.2.2. Partial least squares regression

If the various predictors represent the different factors affecting the volatility of stock returns, then a regression model that includes information on all the predictors can significantly improve the forecasting of volatility. The commonly used multiple regression models may create overfitting problems. Following Kelly and Pruitt (2013, 2015), we use partial least squares (PLS) to construct new predictors to forecast stock return volatility. We assume a standard linear relationship between stock return volatility and an index of investor sentiment, macroeconomic variables and technical indicators:

$$E_t(V_{t+1}) = \alpha + \beta X_t \tag{16}$$

where $E_t(V_{t+1})$ is the expected stock return volatility at $t + 1$-th, X_t is real but unobservable part that is important to predict the stock return volatility. Stock return volatility equals its conditional expectations plus unpredictable shocks:

$$V_{t+1} = E_t(V_{t+1}) + \varepsilon_{t+1} = \alpha + \beta X_t + \varepsilon_{t+1} \tag{17}$$

where ε_{t+1} is an unpredictable component and independent of XT Let $s_t = (s_{1,t}, s_{2,t}, \ldots\ldots, s_{N,t})^T$ denote the N-vector of individual proxies for investor sentiment indicators, macroeconomic variables and technical indicators. We assume the factor structure of s_t as follows:

$$s_{i,t} = \varphi_{i,0} + \varphi_{i,1} X_t + \varphi_{i,2} E_t + e_{i,t}, i = 1, \cdots\cdots, N \tag{18}$$

where $\varphi_{i,1}$ represents the factor loadings for the sensitivity of various factors $s_{i,t}$ to changes in X_t. E_t is the common approximation error component for all agents, independent of volatility, while $e_{i,t}$ is the idiosyncratic noise associated only with measure i. Using this idea for reference, we effectively estimate the real but unobservable variable X_t by eliminating approximation error E_t and idiosyncratic error $e_{i,t}$.

To realize this idea, following Wold (1966, 1975) and Kelly and Pruitt (2013, 2015), we apply the partial least squares (PLS) method to efficiently extract X_t and filter out the irrelevant component E_t. The key idea of the PLS method is that PLS extracts predictors, X_t, from the cross-section based on the covariance of the predictors with stock volatility and selects an optimal linear combination of predictors to forecast. PLS can be implemented by the following three-step OLS regression.

In the first step, we run a regression of N time series. That is, for each individual predictor s_i, we run a time series regression on the constant of $s_{i,t-1}$ and the stock volatility X_t.

$$s_{i,t-1} = \eta_{i,0} + \eta_i V_t + \mu_{i,t-1} \tag{19}$$

The loadings η_i reflect the sensitivity of each proxy $s_{i,t-1}$ to investors and are instrumented with future stock volatility V_t. Because the expected component of V_t is driven by X_{t-1}, the proxy variables are related to expected stock volatility and are uncorrelated with unpredictable volatility shocks, as shown in equations (17) and (18).

In the second step, we perform a cross-sectional regression. Specifically, for each time period t, we perform a cross-sectional regression on $s_{i,t}$. The corresponding loadings $\widehat{\eta}_i$ estimated in the time-series regression (19).

$$s_{i,t} = \varphi_t + F_t^{PLS}\widehat{\eta}_i + v_{i,t} \tag{20}$$

where F_t^{PLS} is the regression coefficient estimated by partial least squares approach using the first stage load and individual predictors.

In the last step, we regress the stock volatility series $\{V_t\}$ and the regression estimation coefficient F_t^{PLS} obtained in the second step to generate the forecast value of the stock volatility in the next period until we get the forecast series with the length of $T-t$.

$$V_t = \omega + \sum_{i=0}^{p-1}\alpha_i V_{t-i} + \beta_{PLS}\widehat{F}_{t-1}^{PLS} + \varepsilon_t \tag{21}$$

3.3. Predictive evaluation

Numerous studies (Inoue et al. 2005; Huang et al., 2015; Neely, Rapach, Tu, & Zhou, 2014) have shown that even if predictors have good in-sample forecasting performance, they do not necessarily have good out-of-sample forecasting performance. In fact, investors and market decision makers are more concerned with out-of-sample forecasting performance. Therefore, we compare the benchmark autoregressive model with the model of interest in terms of out-of-sample forecasting performance.

Based on the existing literature (Rapach et al., 2010, 2016; Neely et al., 2014; Huang et al., 2015; Lin, Wu, & Zhou, 2017; Jiang, Lee, Martin, & Zhou, 2019; Zhang et al., 2019; Liu and Pan, 2020), we employ R_{OS}^2 statistic put forward by Campbell and Thompson (2008). The R_{OS}^2 statistic measures the out-of-sample prediction accuracy of the model of interest relative to the benchmark model, which is computed as percent reduction of mean squared predictive error (MSPE) of the interest model ($MSPE_{model}$) relative to that of benchmark model ($MSPE_{bench}$), that is:

$$R_{OS}^2 = 1 - \frac{MSPE_{mod\ el}}{MSPE_{bench}} \tag{22}$$

where$MSPE_i = \frac{1}{T-M}\sum_{t=M+1}^{T}(V_t - \widehat{V}_{t,i})^2 (i = \ mod\ el, bench)$; vt vt model and $\widehat{V}_{t,bench}$ are the actual stock return volatility, stock return volatility forecast of interest model, and stock return volatility forecast of benchmark model, respectively, at k-th month. M and $T-M$ are the length of initial estimation period and forecast evaluation period, respectively.

According to the statistical test proposed by Clark and West (2007), we can test the predictive model of interest to see if it outperforms the baseline model in terms of MSFE. That is, the MSFE of the model of interest is less than or equal to the MSFE of the benchmark autoregressive model as the null hypothesis of the test, while the alternative hypothesis is the opposite of the null hypothesis (corresponding to $H_0 : R_O S^2 \leqslant 0 \text{against} H_A : R_O S^2 > 0$). Firstly, a theoretical definition of Clark and West (2007) statistics is given as:

$$f_t = \left(r_t - \widehat{r}_{t,bench}\right)^2 - \left(r_t - \widehat{r}_{t,\ mod\ el}\right)^2 + \left(\widehat{r}_{t,bench} - \widehat{r}_{t,\ mod\ el}\right)^2 \tag{23}$$

Then, we get the constant t-statistic by regression $\{f_s\}_{s=m+1}^{T}$ to a constant, which is Clark and West (2007) statistics. At last, we convert the boundary value of the standard normal distribution into the one-sided upper tail p-value.

4. In-sample analysis

According to Inoue and Kilian (2005), we know that having good in-sample forecasting ability does not necessarily mean having good out-of-sample forecasting ability. The null hypothesis of predictive regression model is that the sign of predictors related to stock volatility is not significant, then $H_0 : \beta = 0$. According to Inoue and Kilian (2005), to improve the validity of the hypothesis test, we use the one-sided alternative hypothesis that $H_A : \beta > 0$. When the predictor passes the test, it has predictive performance.

According to Huang et al. (2015), we know that there are three main problems with the practical application of PLS methods to stock volatility forecasting. Firstly, if the predictor is highly persistent, it can lead to spurious regressions (Ferson, 2003). Secondly, when the predictor is highly persistent and correlated with stock volatility, the inference from the above assumptions is wrong (Stambaugh, 1999). Finally, we introduce the expected bias due to sample information used in the first step of the PLS regression. It is important to note that the estimated coefficients from a limited sample are distorted, which is the reason for introducing the expected bias.

We apply the following methodology to address these questions. Firstly, we employ p-values from the wild bootstrap procedure (Goncalves & Kilian, 2004; Cavaliere, Pademetrio, Alvarez, & Blanco, 2010; Dai, Zhu, & Yin, 2022) to account for the correlation between stock volatility and predictors, the persistence of predictors and the general form of the volatility distribution. Secondly, we mimic Stambaugh (1999) to construct the adjusted deviations of Amihud, Hurvich, and Wang (2008) to re-estimate the regression coefficients. Thirdly, to solve the deficiency of look-forward bias in the PLS model, the available information used in first-step time series regression (Eq. (19)) is limited to time *t*, which generates the factor load η_i is used to Eq. (20), thereby the regression slope F_t^{PLS} is free from look-forward bias. Repeating the process, we will obtain the time series F_t^{PLS} that without prospective bias. Table 1.

Table 1 presents estimated coefficients for the predictive regressions, *t*-statistics based on covariance correction, p-values calculated based on the wild bootstrap procedure and R2 statistics for the regression models relative to the AR(6) baseline model. Panel A of Table 1 reports the predictive performance of indicators of investor sentiment, macroeconomic variables and technical indicators of stock volatility. In terms of p-values, the regression coefficients for all predictors are not significant. The results show that RIPO and EQTI in the investor sentiment index are significant at the 1 % level of significance, while the other indicators are not significant. Relative to the technical indicators, the predictive power of the individual predictors in the macroeconomic variables is weak, with only LTY being significant at the 5 % level. However, with the exception of the volume index (OBV), which is significant at the 1 % level, the other two types of technical indicators do not provide information for predicting stock volatility.

The results in Panel C of Table 1 show that all three new indexes constructed using the PLS methodology are significant at the 1 % level of significance. Meanwhile, the new investor sentiment index (PLS_UMS) has an R_{OS}^2 of 0.42 %, which exceeds the R_{OS}^2 of the new macroeconomic variables index (PLS_ECON) and the new technical indicators index (PLS_TI). In addition, the PLS_UMS index has the largest R_{OS}^2 in Table 1. The PLS_ECON index outperforms the forecasting performance of all macroeconomic variables, including LTY. The PLS methodology eliminates the useless noise component of stock volatility forecasts, which means that PLS_TI has greater forecasting power than most popular technical indicators. In line with our expectations, the PLS_TI index produced a slope significant at the 1 % level and an R_{OS}^2 of 0.142 %, which has stronger predictive power than most technical indicators. Panels A to C show that the new indexes constructed using the PLS approach have the ability to predict stock return volatility and outperform other predictors. In addition, the results in Table 1 also show that the new indexes constructed using the PLS method have more valuable forecasting information than the other indexes.

5. Out-of-sample analysis

In this section, we use a recursive estimation window that allows researchers to use past information to forecast future trends in stock volatility and produce forecasts of stock volatility from January 1996 to December 2018. We also compare the out-of-sample forecasting performance of 37 individual forecasters (including 5 investor sentiment indexes, 14 macroeconomic variables and 18 technical indicators), portfolio forecasts and the PLS index. Table 2.

We apply a forecasting model to generate out-of-sample forecasts of stock volatility that uses data from January 1996 to December

Table 1
In-sample results.

Panel A: Individual indicator	β(%)	*t*-stat	R^2(%)		β(%)	*t*-stat	R^2(%)
pdnd	−0.792	−1.620	0.004	MA(1,9)	−0.443	−4.771	0.166
cefd	−0.294	−0.232	0.000	MA(2,9)	−0.455	−4.479	0.161
ripo	1.650	6.507***	0.051	MA(3,9)	−0.418	−4.345	0.143
nipo	−0.958	−3.265	0.028	MA(1,12)	−0.431	−4.346	0.148
eqti	569.741	4.876***	0.089	MA(2,12)	−0.392	−3.950	0.123
DP	−0.245	−4.505	0.026	MA(3,12)	−0.403	−4.033	0.129
DY	−0.250	−4.594	0.028	MOM(1)	−0.229	−3.529	0.056
EP	0.032	0.893	0.001	MOM(2)	−0.287	−4.128	0.084
DE	−0.163	−3.210	0.021	MOM(3)	−0.327	−4.342	0.106
SVAR	0.011	0.264	0.000	MOM(6)	−0.383	−4.346	0.128
BM	−0.640	−4.945	0.017	MOM(9)	−0.414	−4.103	0.135
NTIS	0.909	1.046	0.002	MOM(12)	−0.422	−4.031	0.136
TBL	−0.012	−2.496	0.004	OBV(1,9)	0.021	0.411	0.001
LTY	0.016	1.936**	0.004	OBV(2,9)	0.110	2.146***	0.014
LTR	−0.015	−1.480	0.008	OBV(3,9)	0.162	2.923***	0.030
TMS	−0.076	−4.293	0.046	OBV(1,12)	0.199	3.703***	0.046
DFY	−0.056	−1.614	0.002	OBV(2,12)	0.173	3.346***	0.035
DFR	−0.015	−1.763	0.003	OBV(3,12)	0.193	3.797***	0.043
INFL	−0.141	−1.195	0.010				
Panel B:PLS index							
PLS-UMS	42.295	6.284***	0.420				
PLS-ECON	−0.013	−1.657***	0.003				
PLS-TI	0.209	4.189***	0.142				

Notes: The initial sample of in-sample R^2 results is from January 1991 to December 2019. *, * *, and * * denote rejection of the null hypothesis at 10%, 5% and 1% significance levels, respectively.

2018 and assesses out-of-sample forecasting power over this period. Table 2 reports R^2_{OS} and the *p*-values of CW test for the equivalence of MSPE.

The results in Panel A of Table 2 are estimated using recursive windows of individual predictors (comprising five indexes of investor sentiment, 14 macroeconomic variables and 18 technical indicators). As we can see, most of the investor sentiment indexes and macroeconomic variables have negative R^2_{OS} values, failing to improve out-of-sample forecasting. From the perspective of the investor sentiment index, only the EQTI's out-of-sample forecasting performance exceeds that of the benchmark model, which has an R^2_{OS} of 2.182 %, significant at the 1 % level. From the perspective of macroeconomic variables, only DP, DY, SVAR, TMS and DFR are significant at the 10 % or even lower level among the 14 macroeconomic variables. Of these five variables, DFR has the largest R^2_{OS} of 9.825 % and is significant at the 1 % level of significance. Finally, we analyzed the out-of-sample forecasting performance of technical indicators. The technical indicators all achieve a positive R^2_{OS}, most of which are significant at the 1 % level. p-values from CW tests indicate that the predictive regression models with individual technical indicators mostly outperform the benchmark models. Of all the technical indicators, only MA(3,9), MA(2,12) and MA(3,12) failed to exceed the forecast performance of the benchmark model.

The out-of-sample forecasting results for the five combined forecasting methods and the four new indexes constructed with the PLS method (including PLS_UMS, PLS_ECON, PLS_TI and PLS_ALL) are reported in Panel B and Panel C of Table 2, respectively. The results of the CW tests for the five combination forecasting methods show that the statistics that achieve a positive R^2_{OS} are largely significant. It is noteworthy that the combined forecasts of technical indicators exhibit stronger forecasting performance than the combined forecasts of other types of indicators. In particular, the DMSPE(0.9) combination method for technical indicators calculates an R^2_{OS} of 7.897 %, the largest R^2_{OS} of all combination methods.

In terms of the R^2_{OS} statistic, models using PLS indexes produce more accurate forecasts than the benchmark model and are significant at the 1 % level. PLS indexes include the PLS_UMS index, the PLS_MF index, and the PLS_TI index, constructed from indicators

Table 2
Out-of-sample results.

Panel A: Individual indicator					
	R^2_{os}	MSPE-adjust		R^2_{os}	MSPE-adjust
pdnd	−1.013	0.658	MA(1,9)	8.472***	0.596
cefd	−0.068	0.652	MA(2,9)	4.786***	0.621
ripo	−1.652	0.662	MA(3,9)	1.610	0.641
nipo	−0.872	0.657	MA(1,12)	2.230***	0.637
eqti	2.182**	0.637	MA(2,12)	0.520	0.648
DP	−0.953*	0.658	MA(3,12)	0.531	0.648
DY	1.561***	0.642	MOM(1)	12.632***	0.569
EP	−0.798	0.657	MOM(2)	11.276***	0.578
DE	−0.630	0.656	MOM(3)	6.131***	0.612
SVAR	−3.271***	0.673	MOM(6)	3.178***	0.631
BM	−0.268	0.653	MOM(9)	2.445***	0.636
NTIS	−0.566	0.655	MOM(12)	1.321**	0.643
TBL	−0.522	0.655	OBV(1,9)	7.308***	0.604
LTY	−1.844	0.664	OBV(2,9)	9.301***	0.591
LTR	0.083	0.651	OBV(3,9)	9.817***	0.588
TMS	0.052**	0.651	OBV(1,12)	6.700***	0.608
DFY	−1.390	0.661	OBV(2,12)	3.090***	0.632
DFR	9.825***	0.588	OBV(3,12)	5.015***	0.619
INFL	−0.995	0.658			
Panel B: Combining method					
Mean_UMS	0.655**	1.505	Trimmed_TI	7.219***	6.128
Mean_MF	5.964***	3.961	Trimmed_ALL	5.491***	6.542
Mean_TI	7.618***	6.323	DMSPE1_UMS	0.654*	1.503
Mean_ALL	6.430***	5.918	DMSPE1_MF	6.196***	3.829
Median_UMS	0.221	0.843	DMSPE1_TI	7.897***	6.425
Median_MF	2.017***	3.428	DMSPE1_ALL	6.732***	5.874
Median_TI	6.158***	5.294	DMSPE2_UMS	0.666*	1.531
Median_ALL	3.758***	6.105	DMSPE2_MF	6.803**	3.370
Trimmed_UMS	0.132	0.738	DMSPE2_TI	7.795***	6.398
Trimmed_MF	3.479***	4.223	DMSPE2_ALL	6.991***	5.374
Panel C: PLS index					
PLS_UMS	1.486***	2.327	PLS_TI	11.129***	5.797
PLS_MF	8.915***	3.617	PLS_ALL	14.743***	5.286

Notes: The initial sample of out-of-sample R^2_{OS} results is from January 1991 to December 1995. *, * *, and * * denote rejection of the null hypothesis at 10%, 5% and 1% significance levels, respectively.

of investor sentiment, macroeconomic variables, and technical indicators, respectively. When considering the PLS_ALL index constructed with all indicators, R^2_{OS} is as high as 14.743 % and significant at the 1 % level, the largest R^2_{OS} value in Table 2. In addition, for the portfolio forecasts, although the PLS_UMS index does not appear to have an ultra-high R^2_{OS} relative to the other PLS indexes, the R^2_{OS} increases from a maximum of 0.666 % in the portfolio approach to 1.486 %. The PLS_MF index has stronger forecasting power than the portfolio forecasting approach based on macroeconomic variables, as the model using the PLS_MF index produces an R^2_{OS} of 8.915 %, which is greater than the R^2_{OS} of the portfolio approach. In addition, the technical indicators produce a much lower R^2_{OS} than that produced by the PLS_TI index. The results suggest that the PLS approach takes into account the reduction of forecast variance by eliminating noise when using information from multiple forecasters, thus producing relatively reliable forecasts.

6. Asset allocation

For investors, they are more interested in the economic value of stock volatility forecasts. According to Fleming et al. (2001, 2003) and Guidolin and Timmermann (2005, 2007), we calculate the economic benefits of forecasting from the perspective of maximizing investor utility. Based on the extensive literature on stock returns and stock volatility forecasting (Guidolin, 2008; Rapach et al., 2010; Neely et al., 2014; Wang, Ma, Wei, & Wu, 2016, 2018; Ma, Li, Liu, & Zhang, 2017, 2018; Zhang et al., 2019; Dai & Zhu, 2020; Dai & Zhu, 2023; Dai, Zhu, & Zhang, 2022), we know that rational investors allocate their wealth wisely to equities and risk-free assets. The total utility obtained by mean–variance investors with an optimal allocation strategy is as follows:

$$L_t(r_t) = E_t(\omega_t R_t + R_{t,f}) - \frac{1}{2}\gamma Var_t(\omega_t R_t + R_{t,f}) \tag{24}$$

where ω_t is the weighting of stocks in an investor's portfolio, R_t and $R_{t,f}$ are the return in the stock market and the profits of risk-free

Table 3
Portfolio Performance.

Predictor	CER gain	Growth of CER gain	Predictor	CER gain	Growth of CER gain
Panel A: Individual indicator					
pdnd	−0.330	−6.475	MA(1,9)	1.328	26.022
cefd	−0.467	−9.143	MA(2,9)	0.714	13.983
ripo	0.503	9.864	MA(3,9)	0.521	10.206
nipo	−0.315	−6.177	MA(1,12)	0.528	10.355
eqti	0.151	2.954	MA(2,12)	0.123	2.401
DP	−0.057	−1.109	MA(3,12)	0.157	3.074
DY	−0.023	−0.454	MOM(1)	0.375	7.339
EP	0.162	3.177	MOM(2)	0.654	12.813
DE	−0.272	−5.328	MOM(3)	0.354	6.927
SVAR	0.564	11.059	MOM(6)	0.842	16.507
BM	−0.078	−1.528	MOM(9)	0.390	7.647
NTIS	0.196	3.845	MOM(12)	−0.115	−2.247
TBL	−0.318	−6.229	OBV(1,9)	0.249	4.879
LTY	−0.080	−1.567	OBV(2,9)	0.478	9.367
LTR	0.211	4.139	OBV(3,9)	−0.045	−0.879
TMS	−0.230	−4.508	OBV(1,12)	0.382	7.480
DFY	−0.299	−5.862	OBV(2,12)	−0.001	−0.023
DFR	1.408	27.587	OBV(3,12)	0.430	8.434
INFL	−0.037	−0.724			
Panel B: Combining method					
Mean_UMS	−0.088	−1.718	Trimmed_TI	0.468	9.164
Mean_MF	0.295	5.779	Trimmed_ALL	0.301	5.906
Mean_TI	0.464	9.095	DMSPE1_UMS	−0.096	−1.876
Mean_ALL	0.342	6.704	DMSPE1_MF	0.300	5.874
Median_UMS	−0.223	−4.366	DMSPE1_TI	0.461	9.040
Median_MF	0.206	4.036	DMSPE1_ALL	0.349	6.829
Median_TI	0.572	11.212	DMSPE2_UMS	−0.102	−1.999
Median_ALL	0.279	5.458	DMSPE2_MF	0.316	6.190
Trimmed_UMS	−0.247	−4.843	DMSPE2_TI	0.466	9.133
Trimmed_MF	0.247	4.843	DMSPE2_ALL	0.358	7.016
Panel C: PLS index					
PLS_UMS	0.718	14.062	PLS_TI	1.174	23.011
PLS_MF	0.368	7.217	PLS_ALL	0.826	16.177

Notes: The first column of results is the certainly equivalent return relative to the benchmark model from 1991 to 2018 based on $\gamma = 3$. The second column is the growth rate of certainly equivalent return.

investment, respectively. γ represents the investor's risk aversion coefficient. $E_t(\cdot)$ and Var_t are, respectively, the conditional expectation and the variance of investors' portfolio return.

To maximize the total utility return of investors' assets, we go through the essential steps of determining ω_t:

$$\omega_t^* = \frac{1}{\gamma}\left(\frac{\widehat{r}_{t+1}}{\widehat{\sigma}_{t+1}^2}\right) \tag{25}$$

where $\widehat{r}_{t+1}$ is the forecast of stock return at $(t+1)$-th month, and $\widehat{\sigma}_{t+1}^2$ is the forecast of its variance. Following Neely et al. (2014), we apply the same historical average forecast to obtain $\widehat{r}_{t+1}$ by using a five-year rolling window of stock return. Obviously, the value of the optimal weight is closely related to risk aversion coefficient, that is, the smaller the value of γ, the greater the weight. We consider γ taking three according to Rapach et al. (2010). According to the literature (e.g., Rapach et al., 2010; Neely et al., 2014), we set the weight of stocks in the portfolio range from 0 to 1.5, which is to prevent the effect of short selling on the asset return and to take precautions against exceeding 50 % leverage. The $(t+1)$-month portfolio return is computed as:

$$P_{t+1} = \omega_t^* R_{t+1} + R_{t+1,f} \tag{26}$$

The calculated portfolio is evaluated using the certainty equivalent return (CER):

$$CER_p = \widehat{\mu}_p - \frac{\gamma}{2}\widehat{\sigma}_p^2 \tag{27}$$

where $\widehat{\mu}_p$ and $\widehat{\sigma}_p^2$ present mean and variance of portfolio returns over the forecast evaluation period, respectively. The CER gain is the difference between the CER of investors using the predictive regressions with different models (including the predictive regression with individual indicators, combination methods and predictive regression with PLS indexes) and the CER of investors using the AR(6) benchmark model. We multiply this difference by 1200, which can be defined as the annual percentage portfolio management fee that an investor is willing to pay to use the forecasts of given models instead of the AR(6) benchmark model forecasts. Table 3.

Using the autoregressive model as a benchmark, the second column of Table 3 shows the CER gains using different models. The third column shows the growth in CER gains. We calculate the growth in CER gains by the percentage difference between CER gains using the model of interest and CER gains using the benchmark model. As can be seen from the second column of Panel A in Table 3, of the five investor sentiment indexes, RIPO and EQTI have positive CER gains of 50.3 basis points and 15.1 basis points respectively. Among the 14 macroeconomic variables, only EP, SVAR, NTIS, LTR and DFR have positive CER gains, with DFR generating the largest CER return of 140.8 basis points. The technical indicators all produce positive CER gains except for MOM(12), OBV(3.9) and OBV (2,12), the largest of which is MA(1,9) at 132.8 bps. The third column of panel A shows the growth of CER gains relative to the benchmark model. We can see that for investors using a forecasting model with a single predictor, CER gains do not improve significantly.

As can be seen from Panel B of Table 3, when using the portfolio approach based on investor sentiment indicators, the CER gains are all negative, with the closest positive value being −8.8 basis points. Across the different portfolio approaches, the technical indicators provide greater CER gains relative to other types of forecasting indicators. As can be seen from Panel C of Table 3, the predictive power of the PLS indexes is outstanding. The CER gains for the PLS_UMS, PLS_MF, PLS_TI and PLS_ALL indexes are 71.8 bps, 36.8 bps, 117.4 bps and 82.6 bps respectively. The CER gains for the PLS_UMS indexes are better than all investor sentiment in Panel B results and the results of most investor sentiment indicators in Panel A. Notably, in line with the results in Panels A and B, the PLS_TI index outperforms the other indexes in terms of predictive performance (117.4 basis points), with the PLS_TI index achieving a CER return increase of 23.011 %. From a portfolio utility perspective, the PLS index is more significant in improving the stock volatility forecasting model. In conclusion, the PLS Index has greater economic value to investors than the individual indicators and the five portfolio approaches.

7. Robustness tests

7.1. Business cycle

It is important to examine the impact of the business cycle on stock volatility forecasts. We further analyze the forecasting performance of various predictors in expansionary and recessionary periods in this section. Following Neely et al. (2014), Huang et al. (2015), Ma et al. (2018) and Dai, Zhuo, Kang, and Wen (2021), R_{OS}^2 is calculated based on expansionary and recessionary periods of the business cycle as determined by the National Bureau of Economic Research (NBER). We decompose R_{OS}^2 statistic into $R_{OS,EXP}^2$ and $R_{OS,REC}^2$ statistic based on business cycle, and compute $R_{OS,EXP}^2$ and $R_{OS,REC}^2$ statistic as follow:

$$R_{OS,c}^2 = 1 - \frac{\sum_{t=1}^{M}\left(V_t - \widehat{V}_{t,\ mod\ el}\right)^2 I_t^c}{\sum_{t=1}^{M}\left(V_t - \widehat{V}_{t,bench}\right)^2 I_t^c},\ forc = EXP, REC \tag{28}$$

where $I_t^E XP(I_t^R EC)$is an indicator variable. When the expansionary (recession) period appears in t month, I_t^{EXP} (I_t^{REC})takes one and zero otherwise. Table 4.

Panel A of Table 4 shows the forecast results for stock volatility across business cycles. From the R_{OS}^2 perspective, we find that R_{OS}^2 for RIPO and EQTI are significant at the 10 % level during expansionary periods, while only R_{OS}^2 for EQTI is significant at the 5 % level during recessionary periods. Furthermore, R^2_{OS} for macroeconomic variables and technical indicators is more significant in periods of economic expansion compared to economic recession. However, in all indicators, the maximum $R^2_{OS,REC}$ is greater than the maximum $R^2_{OS,EXP}$. For example, the maximum $R^2_{OS,REC}$ for the investor sentiment indicator is 11.221 %, which is higher than the maximum $R^2_{OS,EXP}$ by only 0.743 %. Similar results are obtained for macroeconomic variables and technical indicators. In short, the predictive performance of a single predictor of stock volatility varies across the business cycle.

Similar to Panel A, Panel B reveals that R^2_{OS} is more significant in expansionary periods than in recessionary periods. Notably, when $R^2_{OS,REC}$ and $R^2_{OS,EXP}$ are based on the same model and both are significant, the value of $R^2_{OS,REC}$ is greater. As can be seen from Panel C, the regression model based on the PLS approach produces a larger R^2_{OS} than most other models over the business cycle. In other words, the forecasting models using PLS indexes produced stock volatility forecasts that are closer to the true values than the other models during expansion and recession periods. In addition, PLS indexes constructed using the PLS methodology generally outperformed forecasts in recessionary periods than in expansionary periods. Considering the impact of all predictors on stock volatility, the PLS_ALL index has high $R^2_{OS,REC}$ and $R^2_{OS,EXP}$ of 25.658 % and 12.994 % respectively, indicating that a PLS approach combining information from all predictors will significantly improve forecasting performance.

7.2. Alternative risk aversion coefficients

In this section, we examine the robustness of stock volatility forecasting results to risk aversion factors and transaction costs. Table 5 sets the transaction cost to 50 basis points and explores the impact of transaction costs on portfolio return growth. Table 6 sets the risk aversion coefficients to 5 and 7 from the original 3.

The results in Tables 5 and 6 show that the forecasting model with PLS indexes based on investor sentiment indexes,

Table 4
Out-of-sample results during expansion and recession period.

Panel A: Individual indicator					
Predictor	$R^2_{OS,REC}$	$R^2_{OS,EXP}$	Predictor	$R^2_{OS,REC}$	$R^2_{OS,EXP}$
pdnd	−3.671	−0.587	MA(1,9)	9.373**	8.328***
cefd	−0.654	0.026	MA(2,9)	6.465*	4.517***
ripo	−7.277	−0.751*	MA(3,9)	4.071	1.216*
nipo	−1.797	−0.724	MA(1,12)	5.650*	1.682**
eqti	11.221**	0.734*	MA(2,12)	1.216	0.408
DP	−12.109	0.834***	MA(3,12)	4.610*	−0.122
DY	−12.099	3.749***	MOM(1)	20.584***	11.358***
EP	−10.329	0.729**	MOM(2)	13.131**	10.979***
DE	−1.974	−0.414	MOM(3)	11.711*	5.237***
SVAR	−84.461*	9.738***	MOM(6)	7.500*	2.486***
BM	−8.721	1.086**	MOM(9)	5.959*	1.882**
NTIS	10.199*	−2.290	MOM(12)	3.579**	0.959*
TBL	−3.172	−0.097	OBV(1,9)	8.906**	7.052***
LTY	3.869*	−2.759	OBV(2,9)	11.840*	8.894***
LTR	0.500	0.016	OBV(3,9)	15.916**	8.840***
TMS	0.236	0.023*	OBV(1,12)	9.123*	6.311***
DFY	−5.726	−0.696	OBV(2,12)	−2.166	3.932***
DFR	31.744**	6.313***	OBV(3,12)	4.640	5.075***
INFL	−3.543	−0.586			
Panel B: Combining method					
Mean_UMS	1.208	0.567*	Trimmed_TI	8.955**	6.941***
Mean_MF	16.602**	4.260***	Trimmed_ALL	7.260**	5.208***
Mean_TI	9.497**	7.318***	DMSPE1_UMS	1.278	0.555
Mean_ALL	11.566**	5.607***	DMSPE1_MF	18.234**	4.267***
Median_UMS	−0.468	0.331	DMSPE1_TI	9.704**	7.608***
Median_MF	−0.939	2.491***	DMSPE1_ALL	12.398**	5.824***
Median_TI	7.321**	5.972***	DMSPE2_UMS	1.257	0.571*
Median_ALL	5.230**	3.522***	DMSPE2_MF	21.843**	4.394***
Trimmed_UMS	−1.304	0.362	DMSPE2_TI	9.397**	7.538***
Trimmed_MF	5.393*	3.172***	DMSPE2_ALL	14.294**	5.820**
Panel C: PLS index					
PLS_UMS	0.222	1.689***	PLS_TI	10.539**	11.223***
PLS_MF	14.223*	8.064***	PLS_ALL	25.658**	12.994***

Notes: The results show that the initial sample is R^2_{OS} of decline period and expansion period from January 1991 to December 1995. ** *, * *, and * * denote rejection of the original hypothesis at 10%, 5% and 1% significance levels, respectively.

macroeconomic variables and technical indicators consistently yields greater economic returns in the asset allocation process, which has passed the robustness test. The economic significance of the model with PLS indexes is not directly related to the type of investor and the investment market environment, and higher CER returns are generated when using this forecasting model.

7.3. Different prediction windows

Rossi and Inoue (2011) argue that estimation window size has an impact on out-of-sample prediction performance. Therefore, in this section, we apply-two forecasting windows, January 2001 to December 2018 and January 2006 to December 2018, to forecast stock return volatility. Our out-of-sample results are presented in Tables 7 and 8.

The results in Tables 7 and 8 are very similar to those obtained when the estimation window of January 1991 to December 2018 was chosen. The technical indicators show significant forecasting performance compared to the other two categories of predictors, with EQTI, DFR and MOM(1) having stronger forecasting power in their corresponding variable categories. Combined forecasts are mostly significant, with technical indicators based on combined forecasting models having stronger forecasting power than any other variables. Consistent with the previous results, forecasting models using PLS indicators outperformed any other models, with those using PLS_ALL indicators producing the largest R^2_{OS}. In addition, stock volatility is much more predictable in recessionary periods than in expansionary periods. Therefore, the out-of-sample forecasts are robust to different estimation windows.

8. Conclusion

In this article, we construct a new PLS-index to predict the volatility of stock returns using the PLS methodology, which is based on a range of predictors including macroeconomic variables, an index of investor sentiment and technical indicators. Our empirical analysis

Table 5
Portfolio Performance for risk aversion $\gamma = 3$ with cost.

Predictor	CER gain	Growth of CER gain	Predictor	CER gain	Growth of CER gain
Panel A: Individual indicator					
pdnd	−0.359	−8.246	MA(1,9)	1.392	31.965
cefd	−0.473	−10.849	MA(2,9)	0.766	17.583
ripo	0.499	11.449	MA(3,9)	0.536	12.299
nipo	−0.334	−7.661	MA(1,12)	0.542	12.453
eqti	0.188	4.308	MA(2,12)	0.129	2.964
DP	−0.103	−2.366	MA(3,12)	0.162	3.730
DY	−0.066	−1.517	MOM(1)	0.232	5.325
EP	0.161	3.690	MOM(2)	0.613	14.066
DE	−0.310	−7.109	MOM(3)	0.333	7.655
SVAR	0.515	11.825	MOM(6)	0.863	19.818
BM	−0.099	−2.273	MOM(9)	0.430	9.882
NTIS	0.200	4.600	MOM(12)	−0.100	−2.293
TBL	−0.321	−7.364	OBV(1,9)	0.239	5.486
LTY	−0.119	−2.735	OBV(2,9)	0.439	10.083
LTR	0.208	4.768	OBV(3,9)	−0.179	−4.108
TMS	−0.242	−5.551	OBV(1,12)	0.336	7.713
DFY	−0.318	−7.303	OBV(2,12)	−0.011	−0.257
DFR	1.159	26.609	OBV(3,12)	0.453	10.396
INFL	−0.032	−0.741			
Panel B: Combining method					
Mean_UMS	−0.094	−2.163	Trimmed_TI	0.547	12.558
Mean_MF	0.315	7.229	Trimmed_ALL	0.354	8.137
Mean_TI	0.548	12.573	DMSPE1_UMS	−0.102	−2.353
Mean_ALL	0.397	9.108	DMSPE1_MF	0.317	7.289
Median_UMS	−0.234	−5.362	DMSPE1_TI	0.545	12.519
Median_MF	0.212	4.868	DMSPE1_ALL	0.404	9.279
Median_TI	0.640	14.697	DMSPE2_UMS	−0.111	−2.543
Median_ALL	0.315	7.224	DMSPE2_MF	0.334	7.667
Trimmed_UMS	−0.263	−6.050	DMSPE2_TI	0.549	12.609
Trimmed_MF	0.265	6.078	DMSPE2_ALL	0.412	9.462
Panel C: PLS index					
PLS_UMS	0.732	16.797	PLS_TI	1.286	29.521
PLS_MF	0.473	10.863	PLS_ALL	0.943	21.660

Notes: The first column of results is the certainly equivalent return relative to the benchmark model from 1991 to 2018 based on $\gamma = 3$ with transaction cost of 50bps. The second column is the growth rate of certainly equivalent return with transaction cost of 50bps.

Table 6
Portfolio Performance for risk aversion $\gamma = 5, 7$ with cost.

$\gamma = 5$			$\gamma = 7$	
	Growth of CER gain	Growth of CER gain (cost = 50 bps)	Growth of CER gain	Growth of CER gain (cost = 50 bps)
Panel A: Individual indicator				
pdnd	−7.002	−10.170	−14.503	−21.719
cefd	−5.408	−6.980	−4.696	−7.391
ripo	12.605	16.535	−3.122	−5.495
nipo	−2.851	−3.406	−0.432	−0.801
eqti	5.724	10.143	5.815	9.659
DP	8.977	13.135	19.700	31.626
DY	10.419	15.520	22.338	36.283
EP	5.237	7.774	5.578	7.873
DE	−4.895	−6.242	3.592	7.111
SVAR	8.573	12.462	11.638	19.606
BM	3.430	5.356	14.341	21.911
NTIS	3.846	6.076	1.056	2.101
TBL	−3.327	−5.457	−1.837	−3.657
LTY	−5.801	−6.682	6.040	11.805
LTR	0.651	−0.422	−4.854	−9.864
TMS	−10.553	−16.261	4.256	6.540
DFY	−1.365	−1.472	5.667	8.800
DFR	15.358	9.796	7.395	−1.007
INFL	−1.002	−1.264	0.113	0.474
MA(1,9)	23.059	32.898	20.537	33.124
MA(2,9)	14.383	21.010	13.128	22.323
MA(3,9)	11.904	16.865	13.848	21.353
MA(1,12)	7.179	9.894	6.220	9.197
MA(2,12)	1.594	2.255	0.835	1.288
MA(3,12)	3.589	4.884	2.776	3.904
MOM(1)	−2.730	−12.094	4.203	0.044
MOM(2)	12.632	14.441	9.937	13.089
MOM(3)	5.915	7.213	7.694	11.962
MOM(6)	14.948	20.304	13.465	19.888
MOM(9)	6.696	9.773	5.062	9.045
MOM(12)	−2.387	−3.140	−2.727	−4.174
OBV(1,9)	1.732	0.526	10.039	13.590
OBV(2,9)	7.513	7.091	12.963	15.512
OBV(3,9)	14.535	10.796	17.520	14.580
OBV(1,12)	4.156	3.662	2.202	1.953
OBV(2,12)	0.134	0.376	−11.030	−16.059
OBV(3,12)	1.389	1.376	3.826	5.624
Panel B: Combining method				
Mean_UMS	1.601	2.615	−2.424	−3.819
Mean_MF	2.342	4.159	10.388	17.181
Mean_TI	6.385	12.577	8.767	18.403
Mean_ALL	4.205	8.220	8.153	15.474
Median_UMS	−0.265	−0.082	−0.399	−0.721
Median_MF	2.904	4.496	8.968	13.950
Median_TI	10.151	15.965	9.620	17.230
Median_ALL	5.279	8.344	5.722	9.600
Trimmed_UMS	−1.473	−1.731	−3.654	−5.625
Trimmed_MF	2.702	4.719	10.123	16.571
Trimmed_TI	6.567	12.478	8.578	17.640
Trimmed_ALL	4.056	7.902	8.062	15.096
DMSPE1_UMS	1.458	2.432	−2.660	−4.171
DMSPE1_MF	2.474	4.322	10.385	17.178
DMSPE1_TI	6.230	12.432	8.861	18.611
DMSPE1_ALL	4.201	8.301	8.123	15.554
DMSPE2_UMS	1.205	2.052	−2.739	−4.324
DMSPE2_MF	2.773	4.769	10.535	17.352
DMSPE2_TI	6.221	12.273	8.654	18.157
DMSPE2_ALL	4.263	8.286	8.090	15.345
Panel C: PLS index				
PLS_UMS	16.770	22.868	13.233	18.114
PLS_MF	4.768	9.560	12.333	23.908
PLS_TI	22.525	35.087	16.658	29.986
PLS_ALL	15.295	24.027	18.411	31.807

Notes: The first column of results is the growth rate of certainly equivalent return relative to the benchmark model from 1991 to 2018 based on $\gamma = 5$, 7. The second column is the growth rate of certainly equivalent return with transaction cost of 50bps.

Table 7
Out-of-sample results for 2001:01–2018:12.

Panel A: Individual indicator							
	R^2_{os}	$R^2_{OS,REC}$	$R^2_{OS,EXP}$		R^2_{os}	$R^2_{OS,REC}$	$R^2_{OS,EXP}$
pdnd	−0.770	−3.671	−0.178	MA(1,9)	10.247***	9.373**	10.425***
cefd	−1.062	−0.654	−1.146	MA(2,9)	6.278***	6.466*	6.240***
ripo	−0.116	−7.277	1.348**	MA(3,9)	2.536***	4.071	2.223**
nipo	−1.318	−1.797	−1.219	MA(1,12)	3.014***	5.650**	2.475**
eqti	2.439**	11.221**	0.644	MA(2,12)	0.882*	1.216	0.814*
DP	−2.335	−12.109	−0.337	MA(3,12)	0.944*	4.610*	0.194
DY	−0.355*	−12.100	2.045**	MOM(1)	14.807***	20.584***	13.627***
EP	−1.320	−10.329	0.521	MOM(2)	13.848***	13.131**	13.995***
DE	−2.573	−1.974	−2.695	MOM(3)	8.043***	11.711***	7.294***
SVAR	−3.877**	−84.462*	12.59***	MOM(6)	4.322***	7.500**	3.673***
BM	−1.413	−8.721	0.081	MOM(9)	3.344***	5.959*	2.809**
NTIS	−0.669	10.199**	−2.890	MOM(12)	1.562**	3.580*	1.149*
TBL	−1.169	−3.172	−0.759	OBV(1,9)	6.898***	8.906*	6.488***
LTY	−3.044	3.869*	−4.456	OBV(2,9)	7.365***	11.840**	6.451***
LTR	0.589*	0.500	0.608	OBV(3,9)	7.558***	15.916**	5.850***
TMS	−2.407	0.236	−2.947	OBV(1,12)	6.404***	9.123*	5.849***
DFY	−1.965	−5.726**	−1.196	OBV(2,12)	1.914**	−2.166	2.747**
DFR	11.392***	31.744	7.233***	OBV(3,12)	5.074***	4.640	5.163***
INFL	−0.738	−3.543	−0.164				
Panel B: Combining method							
Mean_UMS	0.452	1.208	0.297	Trimmed_TI	7.884***	8.955**	7.665***
Mean_MF	5.186***	16.602**	2.853***	Trimmed_ALL	5.320***	7.260**	4.924***
Mean_TI	8.155***	9.497**	7.881***	DMSPE1_UMS	0.471	1.077	0.347
Mean_ALL	6.331***	11.566**	5.261***	DMSPE1_MF	5.917***	19.838**	3.072***
Median_UMS	−0.332	−0.468	−0.304	DMSPE1_TI	8.372***	9.541**	8.134***
Median_MF	0.952**	−0.939	1.339***	DMSPE1_ALL	6.837***	12.990**	5.580***
Median_TI	7.125***	7.321**	7.085***	DMSPE2_UMS	0.468***	1.200	0.319***
Median_ALL	3.870***	5.230**	3.592***	DMSPE2_MF	6.247***	21.773**	3.074***
Trimmed_UMS	−0.393	−1.304	−0.207	DMSPE2_TI	8.248***	9.332**	8.027***
Trimmed_MF	2.492***	5.393*	1.899***	DMSPE2_ALL	6.943***	14.214**	5.457***
Panel C: PLS index							
PLS_UMS	2.431***	0.222	2.882***	PLS_TI	12.444***	10.539**	12.833***
PLS_MF	8.949***	14.223*	7.871***	PLS_ALL	16.901***	25.658**	15.111***

Notes: The initial sample of out-of-sample R^2_{OS} results are from January 1991 to December 2000. *, * *, and * * denote rejection of the null hypothesis at 10%, 5% and 1% significance levels, respectively.

provides evidence for the economic and statistical significance of the PLS-Index and confirms that the index is a powerful predictor.

The in-sample analysis shows that the PLS-Index has the ability to predict stock volatility and exceeds the predictive power of any individual predictor in this paper. Therefore, the out-of-sample forecasting performance of the PLS-Index exceeds that of competing models. the PLS-Index has forecasting power both in-sample and out-of-sample and is robust to different forecast assessment periods, levels of risk aversion and business cycles. In addition, the PLS Index can also achieve positive portfolio returns for investors and the PLS Index tracks the business cycle well.

In conclusion, the PLS index constructed in a data-rich environment greatly improves the prediction of stock volatility, while it is also important in financial markets and investment decisions. the strength of the PLS approach is that it eliminates errors unrelated to stock return volatility, but does not discard it completely, but effectively combines valuable information from the index to obtain high prediction results. From a statistical and economics perspective, previous literature has not tested the data-rich PLS index and we could introduce more data to study this area.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table 8
Out-of-sample results for 2006:01–2018:12.

Predictor	R^2_{os}	$R^2_{OS,REC}$	$R^2_{OS,EXP}$	Predictor	R^2_{os}	$R^2_{OS,REC}$	$R^2_{OS,EXP}$
Panel A: Individual indicator							
pdnd	0.108	0.412	0.048	MA(1,9)	10.533***	14.065***	9.832***
cefd	−0.096	0.226	−0.127	MA(2,9)	6.229**	9.673**	5.545***
ripo	−0.540	−4.337	0.214	MA(3,9)	2.387***	5.994**	1.671*
nipo	−0.875	−1.566	−0.738	MA(1,12)	3.096**	8.099*	2.103**
eqti	2.581**	12.981**	0.517	MA(2,12)	1.216	2.653**	0.930
DP	−1.053	−16.194	1.953***	MA(3,12)	0.740***	6.930***	−0.489
DY	0.529	−17.724	4.152***	MOM(1)	14.212***	20.091***	13.045***
EP	−2.371	−11.668	−0.526	MOM(2)	13.097***	12.665**	13.182***
DE	−0.772	−3.213	−0.288	MOM(3)	8.348***	13.077***	7.408***
SVAR	−6.477**	−101.28*	12.34***	MOM(6)	4.068***	11.114***	2.670**
BM	−0.812	−10.686	1.148*	MOM(9)	3.002**	9.203**	1.772*
NTIS	−0.590	12.226*	−3.134	MOM(12)	2.349***	5.756**	1.673**
TBL	−0.891	−4.201	−0.234	OBV(1,9)	7.688***	11.793*	6.874***
LTY	−1.272	4.178	−2.354	OBV(2,9)	7.452***	9.532	7.040***
LTR	0.091	0.800*	−0.049	OBV(3,9)	6.735***	17.623**	4.574***
TMS	−1.460	−3.225	−1.109	OBV(1,12)	6.816***	10.810	6.023***
DFY	−0.597	−6.519	0.578	OBV(2,12)	3.512***	−1.452	4.498***
DFR	10.559***	36.040**	5.502***	OBV(3,12)	4.812***	−0.017	5.771***
INFL	−0.606	−3.857	0.040				
Panel B: Combining method							
Mean_UMS	0.588*	2.343**	0.240	Trimmed_TI	7.662***	10.044**	7.189***
Mean_MF	5.234***	17.950**	2.711***	Trimmed_ALL	5.091***	7.631**	4.587***
Mean_TI	7.928***	10.474**	7.422***	DMSPE1_UMS	0.594*	2.383**	0.239
Mean_ALL	6.187***	12.690**	4.896***	DMSPE1_MF	5.976***	22.228**	2.750***
Median_UMS	−0.188	−0.034	−0.218	DMSPE1_TI	8.105***	10.515**	7.627***
Median_MF	0.588*	−2.339	1.168***	DMSPE1_ALL	6.700***	14.713**	5.109***
Median_TI	7.074***	8.834**	6.724***	DMSPE2_UMS	0.602*	2.532**	0.219
Median_ALL	3.338***	5.556**	2.897***	DMSPE2_MF	6.269***	24.203**	2.710***
Trimmed_UMS	−0.089	−0.179	−0.071	DMSPE2_TI	8.020***	10.399**	7.548***
Trimmed_MF	2.358***	4.937	1.846***	DMSPE2_ALL	6.808***	15.956**	4.993***
Panel C: PLS index							
PLS_UMS	2.006**	3.385*	1.733**	PLS_TI	11.159***	11.906**	11.011***
PLS_MF	9.353***	14.924	8.247***	PLS_ALL	17.399***	27.477**	15.400***

Notes: The initial sample of out-of-sample R^2_{OS} results are from January 1991 to December 2005. *, * *, and * * denote rejection of the null hypothesis at 10%, 5% and 1% significance levels, respectively.

Data availability

Data will be made available on request.

Acknowledgement

This work was supported by the National Natural Science Foundation of China granted [71771030, 72131011], Natural Science Foundation of Hunan Province granted: 2021JJ30025, Ministry of Education Humanities and Social Sciences Project granted: 22YJA790011.

References

Amihud, Y., Hurvich, C. M., & Wang, Y. (2008). Multiple-Predictor Regressions: Hypothesis Testing. *Review of Financial Studies, 22*(1), 413–434.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics, 61*(1), 43–76.
Asgharian, H., Hou, A. J., & Javed, F. (2013). The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance: A GARCH-MIDAS Approach. *Journal of Forecasting, 32*(7), 600–612.
Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Economics, 146*, 304–317.
Baker, M. P., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance, 61*(4), 1645–1680.
Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives, 21*(2), 129–152.
Bates, J. M., & Granger, C. W. J. (2001). The combination of forecasts. *Journal of the Operational Research Society, 20*, 391–410.
Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Economics, 132*, 169–194.
Campbell, J., & Thompson, S. P. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Scholarly Articles, 21*(4), 1509–1531.
Cavaliere, V., Papademetrio, D. L., Alvarez, E. M. C., & Blanco, G. A. (2010). Haemostatic and immune role of cellular clotting in the sipunculan Themiste petricola. *Cell and Tissue Research, 339*(3), 597–611.

Choudhry, T., Papadimitriou, F. I., & Shabi, S. (2016). Stock market volatility and business cycle: Evidence from linear and nonlinear causality tests. *Journal of Banking and Finance, 66*(5), 89–101.
Christiansen, C., Schmeling, M., & Schrimpf, A. (2012). A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics, 27*(6), 956–977.
Clark, T. E., & West, K. D. (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics, 138*(1), 291–311.
Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics, 7*(2), 174–196.
Conrad, C., & Loch, K. (2015). Anticipating long-term stock market volatility. *Journal of Applied Econometrics, 30*(7), 1090–1114.
Dai, Z., & Zhu, H. (2023). Dynamic risk spillover among crude oil, economic policy uncertainty and Chinese financial sectors. *International Review of Economics and Finance, 83*, 421–450.
Dai, Z., Zhu, H., & Yin, H. (2022). Research on the system risk spillover effects among crude oil, gold, estate and financial sectors in China. *Systems Engineering-Theory& Practice, 42*, 2603–2616.
Dai, Z., Zhu, H., & Zhang, X. (2022). Dynamic spillover effects and portfolio strategies between crude oil, gold and Chinese stock markets related to new energy vehicle. *Energy Economics, 109*, Article 105959.
Dai, Z. F., & Zhu, H. (2020). Stock return predictability from a mixed model perspective. *Pacific-Basin Finance Journal, 60*, Article 101267.
Dai, Z. F., Zhuo, H., Kang, J., & Wen, F. (2021). The skewness of oil price returns and stock premium predictability. *Energy Economics, 94*, Article 105069.
Ferson, R. B. W. E. (2003). A non-random walk down wall street by a. w. lo; a. c. mackinlay. *Journal of Economics, 80*(2), 198–203.
Fleming, J., Kirby, C., & Ostdiek, B. (2001). The Economic Value of Volatility Timing. *Journal of Finance, 56*(1), 329–352.
Fleming, J., Kirby, C., & Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics, 67*(3), 473–509.
Fuertes, A. M., Miffre, J., & Rallis, G. (2010). Tactical Allocation in Commodity Futures Markets: Combining Momentum and Term Structure Signals. *Journal of Banking and Finance, 34*(10), 2530–2548.
Goncalves, S., & Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics, 123*(1), 89–120.
Guidolin, M. (2008). The economic and statistical value of forecast combinations under regime switching: an application to predictable u.s. returns. Working Papers, 595-655.
Guidolin, M., & Timmermann, A. (2005). Economic Implications of Bull and Bear Regimes in UK Stock and Bond Returns. *The Economic Journal, 115*(500), 111–143.
Guidolin, M., & Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics Control, 31*(11), 3503–3544.
Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies, 28*(3), 791–837.
Inoue, A., & Kilian, L. (2005). In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? *Econometric Reviews, 23*(4), 371–402.
Jiang, F., Lee, J. A., Martin, X., & Zhou, G. (2019). Manager Sentiment and Stock Returns. *Journal of Financial Economics, 132*(1), 126–149.
Kelly, B., & Pruitt, S. (2013). Market Expectations in the Cross-Section of Present Values. *Journal of Finance, 68*(5), 1721–1756.
Kelly, B., & Pruitt, S. (2015). The Three-Pass Regression Filter: A New Approach to Forecasting Using Many Predictors. *Journal of Econometrics, 186*(2), 294–316.
Lin, H., Wu, C., & Zhou, G. (2017). Forecasting Corporate Bond Returns with a Large Set of Predictors: An Iterated Combination Approach. *Management Science, 64*, 4218–4238.
Liu, L., & Pan, Z. Y. (2020). Forecasting stock market volatility: The role of technical variables. *Economic Modelling, 84*, 55–65.
Lustig, H. N., Roussanov, N. L., & Verdelhan, A. (2014). Countercyclical Currency Risk Premia. *Journal of Financial Economics, 111*(3), 527–553.
Ma, F., Li, Y., Liu, L., & Zhang, Y. (2017). Are low-frequency data really uninformative? A forecasting combination perspective. *The North American Journal of Economics and Finance, 44*, 92–108.
Ma, F., Liu, J., Wahab, M. I. M., & Zhang, Y. (2018). Forecasting the aggregate oil price volatility in a data-rich environment. *Economic Modelling, 72*, 320–332.
Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time Series Momentum. *Journal of Financial Economics, 104*(2), 228–250.
Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the Stock Risk Premium: The Role of Technical Indicators. *Management Science, 60*(7), 1772–1791.
Nonejad, N. (2017). Forecasting aggregate stock market volatility using financial and macroeconomic predictors: Which models forecast best, when and why? *Journal of Empirical Finance, 42*, 131–154.
Paye, B. S. (2012). Deja Vol: Predictive Regressions for Aggregate Stock Market Volatility Using Macroeconomic Variables. *Journal of Financial Economics, 106*(3), 527–546.
Pettenuzzo, D., Timmermann, A., & Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics, 114*(3), 517–553.
Rapach, D. E., Ringgenberg, M. C., & Zhou, G. (2016). Short Interest and Aggregate Stock Returns. *Journal of Financial Economics, 121*(1), 46–65.
Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-Sample Stock Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies, 23*(2), 821–862.
Rossi, B., & Inoue, A. (2011). Out-of-Sample Forecast Tests Robust to the Choice of Window Size. *Journal of Business & Economic Statistics, 30*(3), 432–453.
Schwert, G. W. (1989). Why Does Stock Market Volatility Change over Time. *Journal of Finance, 44*(5), 1115–1153.
So, E. C. (2013). A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts? *Journal of Financial Economics, 108*(3), 615–640.
Stambaugh, R. F. (1999). Predictive regressions - a comment. *Journal of Financial Economics, 54*, 375–421.
Sullivan, R., Timmermann, A. G., & White, H. L. (1999). Data-Snooping, Technical Trading Rule Performance, and the Bootstrap. *Journal of Finance, 54*(5), 1647–1691.
Szakmary, A. C., Shen, Q., & Sharma, S. C. (2010). Trend-following trading strategies in commodity futures: A re-examination. *Journal of Banking and Finance, 34*(2), 409–426.
Wang, Y., Ma, F., Wei, Y., & Wu, C. (2016). Forecasting realized volatility in a changing world: A dynamic model averaging approach. *Journal of Banking and Finance, 64*(3), 136–149.
Wang, Y., Wei, Y., Wu, C., & Yin, L. (2018). Oil and the short-term predictability of stock return volatility. *Journal of Empirical Finance, 47*, 90–104.
Welch, I., & Goyal, A. (2008). A Comprehensive Look at the Empirical Performance of Stock Premium Prediction. *Review of Financial Studies, 21*(4), 1455–1508.
Wold, H. (1966). Estimation of principal components and related models by iterative least squares. *Multivariate Analysis*, 391–420.
Wold, H. (1975). Path Models with Latent Variables: The NIPALS Approach. *In Quantitative Sociology International Perspectives on Mathematical and Statistical Modeling*, 307–357.
Zhang, Y., Zeng, Q., Ma, F., & Shi, B. (2019). Forecasting stock returns: Do less powerful predictors help? *Economic Modelling, 78*, 32–39.