



## Implied volatility directional forecasting: a machine learning approach

Spyridon D. Vrontos, John Galakis & Ioannis D. Vrontos

To cite this article: Spyridon D. Vrontos, John Galakis & Ioannis D. Vrontos (2021) Implied volatility directional forecasting: a machine learning approach, Quantitative Finance, 21:10, 1687-1706, DOI: [10.1080/14697688.2021.1905869](https://doi.org/10.1080/14697688.2021.1905869)

To link to this article: <https://doi.org/10.1080/14697688.2021.1905869>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 15 Jun 2021.



Submit your article to this journal [↗](#)



Article views: 4690



View related articles [↗](#)



View Crossmark data [↗](#)

# Implied volatility directional forecasting: a machine learning approach

SPYRIDON D. VRONTOS<sup>\*†</sup>, JOHN GALAKIS<sup>‡</sup> and IOANNIS D. VRONTOS<sup>§</sup>

<sup>†</sup>Department of Mathematical Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

<sup>‡</sup>Iniohos Advisory Services, Geneva, Switzerland

<sup>§</sup>Department of Statistics, Athens University of Economics and Business, Athens, Greece

(Received 14 August 2020; accepted 12 March 2021; published online 15 June 2021)

This study investigates whether the direction of U.S. implied volatility, VIX index, can be forecast. Multiple forecasts are generated based on standard econometric models, but, more importantly, on several machine learning techniques. Their statistical significance is assessed by a plethora of performance evaluation measures, while real-time investment strategies are devised to appraise the investment implications of the underlying modeling approaches. The main conclusion of the analysis is that the implementation of machine learning techniques in implied volatility forecasting can be more effective compared to mainstream econometric models and model selection techniques, as they are superior both in a statistical and an economic evaluation setting.

**Keywords:** Forecasting; Implied volatility; Binary logit; Machine learning; Penalized likelihood models; Investment strategies

## 1. Introduction

Over time, a great deal of attention has centered on the predictability of equity market volatility, be it realized and/or implied, as changes in market volatility have significant repercussions on investor preferences. History shows that asset performance and its associated volatility is distinct and asymmetric during different market volatility regimes, as the pricing of market risk and, as a result, investor sentiment is time-varying. As such, identifying, but, more importantly, predicting aggregate market volatility is of critical importance for the implementation of effective asset allocation programs, investment and/or trading strategies and for hedging purposes in particular.

As periods of transition from a low to a high market volatility regime can be very abrupt and relatively short-lived, the development of an effective modeling framework is of critical importance for the design and implementation of active portfolio immunization strategies in order to avoid sizeable drawdowns during periods of turmoil in particular.

In practice, a number of different metrics have been introduced for the estimation of market volatility; the most well-known and followed metrics, however, are realized and implied volatility. Realized volatility gauges the fluctuations of underlying securities or indices by measuring

price changes over predetermined time periods, while implied volatility is a forward-looking metric that represents future expectations of the market's uncertainty. This study focuses on the Chicago Board Options Exchange's (CBOE) VIX index (VIX) that can be considered as a model-free estimator of the equity market's implied volatility. More specifically, the VIX is designed to produce a measure of constant, 30-day expected volatility, derived from real-time, mid-quote prices of S&P 500 index call and put options. The VIX is one of the most recognized volatility measures globally.

A substantial amount of research by both academics and practitioners has focused on the investigation of volatility forecasting and, consequently, on identifying variables that have predictive ability for time-varying volatility dynamics. It is noteworthy that a large part of the empirical evidence on implied volatility predictability is mixed. Dumas *et al.* (1998) explore whether S&P 500 implied volatility dynamics are predictable across different strike prices and expiry dates over alternative periods. They conclude that there is too much instability in the underlying forecasts for them to be useful for pricing and hedging purposes. By contrast, Goncalves and Guidolin (2006) detect a statistically predictable pattern that cannot be meaningfully exploited due to the high transaction costs involved.

Part of the literature has concentrated on the predictability of short-term at-the-money volatility; more specifically,

\*Corresponding author. Email: [svrontos@essex.ac.uk](mailto:svrontos@essex.ac.uk)

Harvey and Whaley (1992) study at the money option volatility for the S&P 100 index, Guo (2000) for the Philadelphia Stock Exchange currency options, and Brooks and Oozer (2002) for LIFFE long Gilt futures and options markets. A set of macroeconomic predictors is used to conclude that changes in implied volatility are partially predictable, but, once again, the results are not economically significant. In a similar study, Gemmill and Kamiyama (2000) find that changes in index option implied volatilities in a certain market are impacted by lagged changes in other markets implied volatilities, i.e. there are large spillover effects. Moreover, Goyal and Saretto (2006) use information stemming from the cross-section of various stock option implied volatilities to find predictable patterns in implied volatility dynamics and they conclude that there is both significant statistical and economic predictability.

The empirical evidence on the predictability of implied volatility indices is less extensive. Early studies include Aboura (2003), Ahoniemi (2006) and they conclude that the evolution of implied volatility indices is statistically predictable. In addition, Ahoniemi (2006) investigates the economic significance of the underlying forecasts and finds that implementing a trading strategy based on S&P 500 options does not achieve abnormal profits. Konstantinidi *et al.* (2008) explore whether European and U.S. index implied volatility can be forecast. More specifically, four major U.S. (VIX, VXO, VXN, VXD) and three European (VDAX, VCAC, and VSTOXX) implied volatility indices are studied. A set of economic and financial market variables that exhibit asset return predictability, serve as potential predictors. Four different types of models are employed to generate forecasts; an economic variables model, uni- and multivariate (VAR) autoregressive models, a Principal Components (PCA) model, as well as ARIMA and ARFIMA models. In sample, the estimated models exhibit low explanatory power, with the VAR and PCA ones performing better than the others. Moreover, there is higher explanatory power associated with European indices. Out of sample, both point and interval forecasts are generated. Regarding point forecasts, the results show that in 28% of the cases one of the estimated models outperforms the random walk model. Predictability is, once again, higher within European indices, as in 41% of the cases there is out-performance over the random walk model, compared to just 18% in the case of their U.S. counterparts. The VAR and PCA models continue to generate the best forecasts related to the European indices, while the ARIMA and ARFIMA models perform best in the case of the U.S. indices. The overall results suggest that there are implied spillover effects between markets, as the information contained in all implied indices can be used to predict separately every European index. This is not true for the U.S. indices; their underlying autocorrelation structure must be considered in order to forecast their evolution. Regarding interval forecasts, predictability is present in 47% of the cases and is stronger for the U.S. indices. Once again, volatility spillovers could be helpful for predictive purposes. However, there is no economic significance associated with the statistically predictable patterns. Clements and Fuller (2012) employ a semi-parametric approach to create a forecasting model regarding the sign of the change in the VIX index. More specifically, VIX

forecasts are generated using a weighted average of past signed changes in the index, where the highest weight is assigned to periods with comparable conditions to the time for which the forecasts are being formed. The main focus of the study is to correctly predict increases in the VIX index. The underlying models are able to generate successful forecasts marginally higher than 50% of the time, while the proportion of increases being correctly predicted range from 49.86% to 54.95%. The generated forecasts are then included in a portfolio immunization context. An active volatility exposure is considered to hedge an underlying long position in the S&P 500 index. It is shown that a long volatility hedge improves the risk-adjusted performance of a long-only equity market position. This is especially true during periods of market turbulence. Fernandes *et al.* (2014) perform a statistical investigation of the time series properties of the daily VIX index. To comprehend the statistical behavior of this index, they employ a multivariate setting that controls for different macroeconomic and financial market conditions. Variants of the Heterogeneous Autoregressive (HAR) process, including a semi-parametric neural network model with Bayesian regularization, are estimated both in and out-of-sample. Out-of-sample forecasts are generated for different daily horizons which show that the simple HAR model seems to exhibit superior performance when compared to alternative HAR specifications and a random walk model, while the neural network model performs best at longer forecast horizons.

Full-blown application of Machine Learning (ML) algorithms in realized and implied volatility forecasting, however, has been relatively scarce. Vortelinos (2017) examines whether nonlinear models, such as Principal Components Combining, Neural Networks and GARCH, are better in forecasting realized volatility in numerous U.S. asset markets relative to the HAR model. Recently, Hosker *et al.* (2018) compare three established financial models that have been used in equity market volatility forecasting to various machine/deep learning supervised regression methods. More specifically, their analysis provides forecasts for the 1-month VIX futures contract three and five days ahead. They conclude that ML methods based on Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) generate improved results over linear regression, Principal Components Analysis (PCA), and ARIMA models. More recently, Audrino *et al.* (2020) analyze the impact of sentiment indicators and investor attention variables on realized equity market volatility. They apply an innovative sentiment classification system to explore whether sentiment measures contain additional information content for realized volatility, after controlling for numerous macroeconomic and financial predictive variables. Using a penalized regression framework, they identify the most relevant variables for volatility prediction. Their results reveal that although attention and sentiment indicators improve volatility forecasts, they are not significant from an economic perspective. Hirska *et al.* (2021) apply several machine learning techniques and neural networks to replicate the VIX index as well as VIX futures by using a small number of S&P options, and thus, are able to exploit potential arbitrage opportunities between the VIX index and its underlying derivatives.

The majority of the empirical research conducted on realized and implied volatility forecasting has been focused on point or interval forecasts rather than directional ones. The primary aim of this study is to create an effective forecasting framework to predict the direction of the VIX index. Forecasts are generated using standard binary econometric models, combination of forecasts methodologies and numerous machine learning techniques. Various statistical evaluation metrics are employed to assess the predictive ability of the underlying models and a number of investment strategies are devised to appraise their economic significance.

More specifically, the contributions of our work are several; first, a binary-valued dependent variable representing the directional patterns of the VIX index is introduced to explore whether the direction of the VIX can be forecast using a wide set of predictor variables. Second, following the trend of new variable detection, a number of less well studied predictors are considered in the underlying analysis. Third, a plethora of machine learning techniques is implemented in order to find out which of these methods improve predictive performance. Fourth, within the machine learning framework and for penalized likelihood techniques, automatic variable selection is performed to identify important predictors that are crucial in order to understand the evolution of market risk and the way that market returns change over time, and to build real-time investment strategies.

In general, the results reveal that machine learning techniques can lead to substantial improvement in predicting the directional patterns of the VIX index, as revealed by statistical and economic evaluation metrics.

The remainder of the paper is organized as follows. Section 2 provides an outline of the research methodology, i.e. introduces the different model specifications and the machine learning techniques; in Section 3 the data sets are described; Section 4 summarizes the empirical design and findings and finally, Section 5 concludes.

## 2. Research methodology: model specifications and machine learning techniques

The modeling approaches and machine learning techniques used to forecast VIX index related directional patterns are outlined in this section. More specifically, standard Logit regression models, penalized likelihood Logit models, as well as several machine learning techniques are briefly discussed.

### 2.1. Binary Logit regression models

A directional binary-valued dependent variable  $Y_t$  of the following form is considered throughout this paper:

$$Y_t = \begin{cases} 1, & \text{if } VIX_t > VIX_{t-1}, \\ & \text{i.e. there is an increase in the VIX volatility index} \\ & \text{at time } t \\ 0, & \text{otherwise, i.e. if } VIX_t \leq VIX_{t-1}, \\ & \text{there is a decrease in the VIX volatility index} \\ & \text{at time } t. \end{cases}$$

Generalized linear models (GLM) can be used to model the dependence of the binary directional variable  $Y_t$ , given a set of  $N$  (lagged) covariates or predictor variables,  $X_{j,t-l}$ ,  $j = 1, \dots, N$ ,  $l = 1, \dots, h$  and/or autoregressive components,  $Y_{t-i}$ ,  $i = 1, \dots, p$ . In the analysis, GLM are considered to model the mean  $\mu$  of an observation of  $Y$ , which is related to the predictor variables through a link function and a linear predictor model of the form  $g(\mu) = \eta$ , where  $g(\mu)$  is the link function, and  $\eta$  is the linear predictor. Regarding the Logit regression model, the link function  $g(\mu)$  is the Logit transformation,  $g(\mu_t) = \ln(\frac{p_t}{1-p_t})$ , which is a linear function of the predictor variables, the autoregressive components, and the unknown parameters.  $Z_{t-h} = (1, X_{1,t-1}, \dots, X_{1,t-h}, \dots, X_{N,t-1}, \dots, X_{N,t-h}, Y_{t-1}, \dots, Y_{t-p})'$  denotes the augmented vector of predictors that contain the lagged predictor variables and the autoregressive components,  $h$  being the maximum number of lags used in the model, and  $\theta = (\beta', \phi')'$  the total parameter vector with  $\beta = (\beta_0, \beta_{1,1}, \dots, \beta_{1,h}, \dots, \beta_{N,1}, \dots, \beta_{N,h})'$  and  $\phi = (\phi_1, \dots, \phi_p)'$ . Then, the Logit regression model can be written in the form:

$$\ln\left(\frac{p_t}{1-p_t}\right) = Z'_{t-h}\theta = \beta_0 + \sum_{j=1}^N \sum_{l=1}^h \beta_{j,l} X_{j,t-l} + \sum_{i=1}^p \phi_i Y_{t-i}, \quad (1)$$

where  $p_t = \text{Prob}(Y_t = 1|Z)$  is the conditional probability of an upward directional move.

Maximum likelihood estimates of the Logit regression model parameters,  $\theta$ , can be obtained by assuming that each  $Y_t$  is the outcome of independent Bernoulli random variables with an upward directional move probability  $p_t$ . The log-likelihood function for a sample of  $T$  observations is given by

$$\begin{aligned} \ln f(Y|Z, \beta) &= \sum_{t=1}^T [y_t \ln p_t + (1 - y_t) \ln (1 - p_t)] \\ &= \sum_{t=1}^T \ln [1 - p_t] + \sum_{t=1}^T y_t \ln \left[ \frac{p_t}{1 - p_t} \right] \\ &= - \sum_{t=1}^T \ln \left[ 1 + \exp \left( \beta_0 + \sum_{j=1}^N \sum_{l=1}^h \beta_{j,l} X_{j,t-l} + \sum_{i=1}^p \phi_i Y_{t-i} \right) \right] \\ &\quad + \sum_{t=1}^T y_t \left( \beta_0 + \sum_{j=1}^N \sum_{l=1}^h \beta_{j,l} X_{j,t-l} + \sum_{i=1}^p \phi_i Y_{t-i} \right). \end{aligned}$$

### 2.2. Penalized likelihood binary logit regression models

The Ridge, the Least Absolute Shrinkage and Selection Operator (LASSO), and the Elastic Net regularization techniques are presented below, as they are implemented in the empirical analysis, using variants based on the Area Under the Curve (AUC), Deviance, and Class metrics. The specific techniques are penalized least squares methods that impose shrinkage in the regression coefficients and allow for automatic variable selection. The inclusion of a parameterization penalty in the likelihood function is strongly recommended

to overcome overfitting issues, especially in the presence of a large predictor set, and to improve out-of-sample predictive performance.

**2.2.1. The Ridge Logit model.** The Ridge Logit model (based on the Ridge regression model introduced by Hoerl and Kennard 1970) can result in better prediction accuracy, by shrinking the estimated coefficients towards zero. The specific process has numerous benefits, as it leads to lower variance and decreases the complexity of the underlying model, without reducing the number of predictors. Given the log-likelihood function of the ordinary Logit model, with parameter vector  $\theta$ , the Ridge log-likelihood Logit function introduces a shrinkage penalty that employs the  $\ell_2$ -norm of  $\theta$ ,  $\|\theta\|_2 = \sqrt{\sum_i \theta_i^2}$ , and a tuning parameter  $\lambda$ ,  $\lambda > 0$ , that controls the degree of regularization. Increasing the value of  $\lambda$  has the tendency to reduce the magnitude of the estimated coefficients, but does not result in the elimination of any of the predictors. The maximum likelihood estimator of the Ridge Logit model is given by

$$\hat{\theta}_R^{Logit} = \operatorname{argmax}_{\theta} \left\{ \ln f(Y|Z, \theta) - \lambda \sum_i \theta_i^2 \right\}.$$

**2.2.2. The LASSO Logit model.** An alternative modeling approach is the Least Absolute Shrinkage and Selection Operator (LASSO) model, introduced by Tibshirani (1996). By imposing an alternative type of shrinkage, the LASSO Logit model provides better interpretability relative to that of the Ridge regression and also performs variable selection. The LASSO Logit coefficients are estimated by maximizing the corresponding log-likelihood function, while imposing a shrinkage penalty based on the  $\ell_1$ -norm of  $\theta$ . The vector  $\hat{\theta}_L^{Logit}$  is obtained by

$$\hat{\theta}_L^{Logit} = \operatorname{argmax}_{\theta} \left\{ \ln f(Y|Z, \theta) - \lambda \sum_i |\theta_i| \right\}$$

where  $\|\theta\|_1 = \sum_i |\theta_i|$  is the  $\ell_1$ -vector norm of  $\theta$  and  $\lambda$ ,  $\lambda > 0$ , is the LASSO-related tuning parameter that controls the degree of shrinkage of  $\theta$ . The  $\ell_1$  penalty function shrinks the underlying coefficients towards zero, forcing some of the coefficient estimates to be exactly equal to zero, when the tuning parameter  $\lambda$  is sufficiently large. The specific modeling approach eliminates entirely a number of predictors and could, thus, result in a set of predictive variables that are the most important to accurately forecast the direction of the VIX index.

**2.2.3. The Elastic Net Logit model.** The Elastic Net is an alternative modeling approach, proposed by Zou and Hastie (2005) that has been extended to the Logit regression framework. The Elastic Net was introduced as an improved technique (to LASSO and Ridge), that is able to handle high correlated variables in the predictor set that is, typically, inherent in the analysis of high dimensional data. It combines the penalty terms of the Ridge and LASSO regression by using a

convex combination scheme. In practice, this can be achieved as the  $\ell_1$ -norm of the Elastic Net performs automatic variable selection, while the  $\ell_2$ -norm stabilizes the solution, and, thus, raises the out-of-sample predictive performance. The Elastic Net Logit coefficient estimates are obtained by maximizing the log-likelihood function that penalize the size of the model coefficients based on both the  $\ell_1$ -vector norm and  $\ell_2$ -vector norm of  $\theta$ . Thus, the parameter estimates are obtained by

$$\hat{\theta}_{EN}^{Logit} = \operatorname{argmax}_{\theta} \left\{ \ln f(Y|Z, \theta) - \lambda \left( (1 - \alpha) \sum_i \frac{\theta_i^2}{2} + \alpha \sum_i |\theta_i| \right) \right\}$$

where  $\lambda$  and  $\alpha$  are the Elastic Net tuning parameters. For tuning parameter selection  $\lambda$  in Ridge, LASSO and Elastic Net approaches, a validation set is needed in which the predictive value of the specific penalized Logit model can be compared for various values of the tuning parameter, and the optimal parameter should be chosen in such a way that the error rate is minimal. In the analysis, the optimal tuning parameter is chosen through cross-validation.

### 2.3. Discriminant analysis

Both Linear and Regularized Discriminant analysis is applied in this study. These techniques are alternative classification tools that overcome certain limitations linked to standard Probit/Logit type regression models, i.e. when the binary-valued dependent variable is well separated or when there are just a few observations in a certain state, increasing the probability that the Logit regression estimates become unstable.

**2.3.1. Linear discriminant analysis.** Linear Discriminant Analysis (LDA) is a multivariate statistical technique used to classify and generate forecasts with respect to a binary dependent variable based on a set of predictors. The main idea is to derive a linear combination of the predictors that ‘best’ differentiates the events related to a specific state. There are several ways to determine discriminant criteria. One is to use a Bayesian approach for classification and derivation of probabilities. Based on the Bayes’ theorem, the probability of a specific event given a set of predictors  $Z$  is estimated by:

$$Pr(Y_t = k | Z_t = z_t) = \frac{\pi_k f_k(z)}{\sum_{i=0}^1 \pi_i f_i(z)}, \quad \text{for } k = 0, 1$$

where  $k = 0$  and  $k = 1$  denote the two possible outcomes of the dependent variable,  $\pi_k$  denotes the prior probability,  $f_k(z) = Pr(Z = z | Y = k)$  denotes the conditional density of  $Z$  given  $Y = k$ , while  $Pr(Y = k | Z = z)$  denotes the posterior distribution of higher or lower implied volatility, given the set of predictors  $Z$ . Assuming that the conditional distribution of the set of predictors  $Z$  given  $Y = k$  has a multivariate normal density with mean vector  $\mu_k$  and a common covariance matrix  $S_k = S$  for  $k = 0, 1$ , i.e.  $f_k(z) \sim N(\mu_k, S)$ , LDA assigns an observation  $Z = z$  to each regime based on the largest value



of

$$z' S^{-1} \mu_k - \frac{1}{2} \mu_k' S^{-1} \mu_k + \log \pi_k$$

for  $k = 0, 1$ . For more details, please refer to James *et al.* (2013).

**2.3.2. Regularized discriminant analysis.** The Regularized Discriminant Analysis (RDA) technique replaces the class specific covariance estimates used in LDA by their average, pooled covariance matrix and utilizes regularization to enhance performance. The regularized covariance matrices have the following form:

$$S_k(\lambda) = \lambda S_k + (1 - \lambda)S$$

where  $S$  is the pooled covariance matrix used in LDA,  $S_k, k = 0, 1$ , is the class specific covariance matrix used in Quadratic Discriminant Analysis, and  $\lambda \in [0, 1]$  can be specified by employing cross-validation in practice. An alternative way of shrinkage is to use a convex combination that allows  $S_k$  to be shrunk towards a scaled identity matrix, using the shrinkage parameter  $\gamma$  as follows:

$$S_k(\lambda, \gamma) = (1 - \gamma)S_k(\lambda) + \gamma \frac{1}{d} \text{Tr}[S_k(\lambda)]I$$

where  $\frac{1}{d} \text{Tr}[S_k(\lambda)]$  is the mean of the diagonal elements of  $S_k(\lambda)$ , representing the mean variance of the class predictors. The RDA classifier is given by the following equation:

$$\Psi_k^{RDA}(z) = \left\{ (z - \bar{z})' S_k^{-1}(\lambda, \gamma) (z - \bar{z}_k) + \log |S_k(\lambda, \gamma)| \right\}$$

where  $\lambda$  is the cross-validation parameter that controls the degree of shrinkage of the individual class covariance matrix estimates towards the pooled estimates and  $\gamma$  is an additional regularization parameter that controls shrinkage toward a multiple of the identity matrix for a given value of  $\lambda$ . Please refer to Friedman (1989) for more details.

## 2.4. Classification and Regression Trees

A classification problem can also be addressed using tree-based classification methods. Classification and Regression Tree (CART) models, introduced by Breiman *et al.* (1984), offer a flexible way to analyze non-linear relationships between a dependent variable and a set of predictors. In tree-based models, the space of the predictor variables is partitioned into a number of simple regions created recursively employing a sequence of binary decisions. The terminal nodes of the tree correspond to distinct and non-overlapping regions of the partition, and the partition is determined by splitting rules associated with the set of predictors at each of the internal (splitting) nodes.

The estimation procedure of the tree structure is usually based on a top-down, greedy algorithm to grow a tree together with a pruning algorithm to avoid overfitting. Starting from a single node at the top of the tree, and then successively splitting the predictor space, the tree is growing by sequentially choosing the new partitions, based at each step on a

model fitting criterion, such as the Gini index or entropy. This sequential procedure produces a maximal binary tree that is corrected by pruning through a model selection criterion, such as cost complexity pruning, cross-validation, or an information criterion like AIC or BIC. Based on the estimated tree structure, prediction for a given observation can be generated by using the mode of the training observations in the partition to which it belongs.

Even though tree-structured models are simple and easy to interpret, they suffer, however, from reduced predictive accuracy and non-robustness (small changes in the data can lead to different tree-structures and, therefore, to different predictions). Ensemble methods can substantially enhance predictive performance by aggregating/combining numerous different decision trees; see, for example, Breiman (1998), Bauer and Kohavi (1999), Buhlmann and Yu (2002), Biau *et al.* (2008), Dietterich (2000). Ensemble methods include Bagging (Breiman 1996), Boosting (Freund and Schapire 1997, Friedman 2001), and Random Forest (Breiman 2001).

**2.4.1. Bagging.** Bagging (Breiman 1996) improves the predictive accuracy of a classifier by generating multiple versions of the classifier based on bootstrap replicates of the training data set, and then combines these classifiers in order to create a single one. The specific aggregated process produces a stable forecast/classifier with smaller variance and, thus, succeeds in achieving substantial gains in terms of accuracy.

The core idea of the algorithm is to create  $B$  bootstrap samples  $S_1^*, S_2^*, \dots, S_B^*$ . From each bootstrap sample  $S_i^*$ ,  $i = 1, \dots, B$ , the quantity of interest, i.e. the classifier, say  $\hat{C}_i^*$ , is estimated based on the same learning procedure. Then, the bagged estimator/classifier,  $\hat{C}_{bag}$ , can be obtained by aggregating the different bootstrap classifiers.

**2.4.2. Random Forest.** The Random Forest (Breiman 2001) represents an alternative machine learning technique that generates improved predictive accuracy compared to the standard classification and regression tree models. Random Forest can reduce the classifier variance by aggregating a number of low correlated trees obtained using different bootstrap training samples. This can be achieved by creating multiple decision trees, where the splitting variables of the internal nodes are based on a random process, i.e. by introducing randomness in the tree-growing procedure.

As in the case of bagging,  $B$  bootstrap samples  $S_1^*, S_2^*, \dots, S_B^*$  are generated from the training data set. For each bootstrap sample, a Random Forest classification tree is created,  $T_i^{rf}$ , and the classifier  $\hat{C}_i^{rf}$ ,  $i = 1, \dots, B$ , is estimated based on the same learning procedure. Then, the Random Forest classifier,  $\hat{C}_{rf}$ , can be obtained by aggregating the different bootstrap classifiers using the majority vote rule, i.e.  $\hat{C}_{rf} = \text{majority vote } \{\hat{C}_i^{rf}\}_1^B$ .

The Random Forest classifier found by aggregating different bootstrapped classifiers is similar to that generated by the Bagging technique. There is a significant modification in the way that Random Forest classification trees are created, however. The splitting variable is chosen to be the 'best' variable among a random subset of  $m$  candidate variables taken

from the full set of the  $N$  predictive variables. A new random sample of  $m$  candidate splitting variables is taken at each splitting node of the tree. Thus, the Random Forest selects at each node of each tree, a random subset of variables, and only these variables are used as candidates to find the best split for the node. The use of different bootstrap samples, i.e. create different classification trees, together with the introduction of this randomness at each splitting node of each tree, provides a number of uncorrelated trees and, as a consequence, the Random Forest forecast/classifier achieves reduced variance.

**2.4.3. Adaptive boosting.** Boosting is an ensemble technique that combines several weak classifiers to produce a more accurate and powerful classifier. The basic idea is to combine classifiers that are iteratively created through the resampling of the training data by assigning higher weight to observations that were misclassified, to produce a new classifier that could boost the underlying performance. This process is repeated, generating several classifiers, which are then combined into a final classifier by applying weighted majority vote.

A very interesting feature of the Adaptive Boosting algorithm was presented by Friedman *et al.* (2000) who showed that Adaptive Boosting can be interpreted as a forward stagewise additive model that minimizes exponential loss. In the analysis, the Adaptive Boosting (Adaboost) algorithm proposed by Freund and Schapire (1997) is employed.

**2.4.4. Gradient boosting.** Friedman (2001, 2002) developed a framework for the creation of a new generation of boosting techniques, based on boosting's link to the statistical principles of additive modeling and maximum likelihood (Friedman *et al.* 2000). More specifically, Friedman (2001) presented a general method for estimating/approximating a function  $F(z)$ , mapping the predictor variables  $Z$  to the dependent variable, based on numerical optimization in the function space. The function  $F(z)$  can be expressed in terms of forward stagewise additive modeling and the application of steepest-descent minimization. A generic gradient descent 'boosting' algorithm is developed (Gradient Boosting) for additive expansions based on different fitting criteria. The core idea of the Gradient Boosting algorithm is to create additive models by sequentially fitting a base (weak) learner to current pseudo-residuals at each iteration. Thus, the learning process fits new models consecutively, to provide a more accurate estimate of the dependent variable.

Assuming that each  $Y_t$  ( $Y_t = 1$  denotes higher implied volatility at time  $t$ ) is the outcome of independent Bernoulli random variables with probability of higher implied volatility  $p_t$ , the quantity of interest is the log-odds, denoted by  $\lambda(z)$ , where  $\lambda(z) = \ln\left(\frac{p_t}{1-p_t}\right)$  that is a function of the set of predictors  $Z_t$ . Under Gradient Boosting,  $\lambda(z)$  is the function  $F(z)$  that can be expressed based on additive expansions of the form:

$$\lambda(z) = F(z) = \sum_{b=1}^B g_b(z; \gamma_b)$$

where  $g_b(z; \gamma_b)$  is the base learner (e.g. a tree), and  $\gamma_b$  is the parameter vector to be estimated (for trees, it contains

identification of splitting variables, their splitting values, and the constants in the terminal nodes). The steps of the Gradient Boosting algorithm can be found in Friedman (2001) and Efron and Hastie (2016).

**2.4.5. Recursive Partitioning Algorithm.** Recursive Partitioning Algorithms (rpart) are used to create classification and regression models; that is, the resulting models can be represented as binary trees. The tree structured model is created using a two-stage procedure; in the first step, a single predictor variable is found that best splits the data into two groups. There are several measures that can be employed to find the best split; possible choices can be the information index and the Gini index. The dataset is partitioned using this splitting variable into two child nodes, and this process is applied independently to each sub-group (child node). This is done recursively until the subgroups or the terminal nodes either reach a minimum size, or until no improvement can be achieved. The specific tree structure may be overly complicated. The second step of the procedure entails pruning of the full tree. A complexity parameter  $\alpha$ , measuring the 'cost' of adding another variable, can be used.  $\alpha$  is chosen using cross-validation. More details can be found at Therneau and Atkinson (2019).

**2.4.6. k-Nearest Neighbor.**  $k$ -Nearest Neighbor ( $k$ -nn) is a non-parametric technique used for classification. It is a distance-based classifier that takes into account the similarity of the predictive variables at a future time period with the observations of the training set. Then, the new observation is classified based on the majority vote of its neighbors, i.e. the new observation is assigned to the class that is most common among its  $k$  nearest neighbors. Implementation of this technique, requires the specification of a distance metric, that is needed to calculate the distance between the new observation and that of the training sample, and a positive integer  $k$ , which determines the number of neighbors used in the voting process. A Euclidean distance metric can be used for computing distances in a multi-dimensional predictor space with quantitative variables, as is the case in the underlying analysis. The choice of  $k$  is very important, as it affects the classifier obtained by the technique. A potential solution is to use cross-validation to select the value of  $k$ . The specific classification method does not require any model to be fitted, and does not impose any strong assumption regarding the underlying data.

**2.4.7. Naive Bayes.** The Naive Bayes learning scheme (see, e.g., Bauer and Kohavi 1999) applies the Bayes theorem and assumes conditional independence of the predictors given the classifier. The basic mechanism for classification can be derived as follows:

$$\begin{aligned} Pr(Y = k|Z = z) &= \frac{Pr(Z = z|Y = k)\pi(Y = k)}{Pr(Z = z)} \propto Pr(Z = z|Y = k)\pi(Y = k) \\ &= \prod_i Pr(Z_i = z_i|Y = k)\pi(Y = k) \end{aligned}$$

where  $Pr(Y = k|Z = z)$  denotes the posterior distribution of higher implied volatility ( $k = 1$ ), or lower implied volatility ( $k = 0$ ), given the set of predictors  $Z$ ,  $Pr(Z = z|Y = k)$  denotes the conditional density of  $Z$  given  $Y = k$ , and  $\pi(Y = k)$  denotes the prior distribution of the binary dependent variable. Thus the Naive Bayes classifier can be found by applying:

$$\Psi^{NB}(Z) = \operatorname{argmax}_k \left\{ \pi_k \prod_i Pr(Z_i = z_i|Y = k) \right\}$$

where  $\pi_k$  is estimated by the sample proportions. It is evident that the method assigns an observation  $Z = z$  to the class with the largest posterior probability, known as the maximum posterior decision rule.

The method is implemented in classification problems, especially when the dependent variable is qualitative, as the independence assumption is less restrictive than might be expected, and, usually, the method assigns maximum probability to the correct class. Its good performance can be attributed to the use of the zero-one loss function.

**2.4.8. Bayesian generalized linear models.** The Bayesian approach to inference regarding the estimation of stable Logit regression coefficients is briefly outlined. In Bayesian analysis, inference regarding the model parameters is based on the posterior distribution that is computed using the Bayes theorem:

$$\begin{aligned} f(\theta_m|D) &= \frac{f(D|\theta_m)f(\theta_m)}{f(D)} \\ &= \frac{f(D|\theta_m)f(\theta_m)}{\int f(D|\theta_m)f(\theta_m)d\theta_m} \propto f(D|\theta_m)f(\theta_m) \end{aligned}$$

where  $f(\theta_m|D)$  is the posterior distribution of a specific model's  $m$  parameters,  $\theta_m$ , given the data  $D$ ,  $f(D|\theta_m)$  is the likelihood function, and  $f(\theta_m)$  is the prior density of model  $m$  parameters. The posterior distribution summarizes all that is known about the model parameters after 'seeing' the data. To reflect the ex-ante opinion of the uncertainty regarding the model parameters, a prior distribution is assigned. Conditional on having observed the data, the prior opinion can then be updated to a posterior opinion on model parameters using the Bayes theorem.

To implement the Bayesian generalized linear model in the problem of interest, the approach proposed by Gelman *et al.* (2009) is applied, using the 'bayesglm' function, that finds the approximate posterior mode and variance using extensions of the classical generalized linear model computations. Gelman *et al.* (2009) propose the use of a new prior distribution, the Student-t family, focusing on the Cauchy sequence as a conservative choice, for the Logit regression model parameters by first scaling the predictor variables. The proposed prior distribution setting enables the production of stable, regularized estimates, even when there is separation in Logit regression, and allows for an automated process in applied data analysis.

### 3. Data

As the primary sources of equity market volatility stem from either exogenous and/or endogenous shocks in the economic, credit, and financial cycle, it is only rational to include a set of economic and financial variables as likely predictors. A relatively large number of predictive variables is considered, as empirical research has shown that their effectiveness is, to a large extent, time dependent. More specifically, the set of predictors contains 31 macroeconomic and financial market related indicators. Part of them is widely-followed by both policy makers and practitioners, and has been used in the existing literature for implied volatility prediction, while a number of less studied indicators is also included in the analysis. In general, the forecasting variables are representative of risk categories related to macroeconomic and financial market risk, momentum, as well as investor sentiment factors with monthly frequency.

The set of macroeconomic risk factors includes variables, such as the term spread, the oil price, the dollar index, the U.S. Economic Policy Uncertainty index and a proxy for macroeconomic volatility based on the annualized 12-month standard deviation of industrial production growth. Equity market risk is proxied by numerous equity related factor portfolios. Apart from the well-known Fama–French three factor portfolios, HML (Value), SMB (Size), and MKT (Market), additional factor portfolios are employed, representing alternative risk factors, such as QMJ (Quality), BAB (Low Volatility), and RVAR (Residual Variance). In addition, financial and credit risk is proxied by the default and the TED spread, and the Chicago Financial Conditions Leverage index, among others.

The use of medium and short-term equity momentum factors is supported by the large literature on the predictability of asset returns. As a result the Fama–French MOM (medium term price momentum) and STR (short-term reversal) portfolio returns are apparent candidates. However, momentum factors from various other asset classes (fixed income, commodity, and foreign exchange) are also employed, due to the existence of cross-asset or inter-market linkages. Last, a series of investor sentiment related factors based on the American Association of Individual Investor survey is included in the analysis.

The directional indicator is calculated from the Chicago Board Options Exchange (CBOE) Volatility index (VIX). The indicator is a binary variable that takes the value of 1, if the VIX trends higher in month  $t$ , and 0, if the VIX trends lower. Autoregressive components of the directional indicator are used as predictors, to capture the likely underlying persistence in the implied volatility series. A comprehensive list of the predictors, as well as their source, is presented in table A1 in the appendix.

The data cover the period between January 1990 to December 2019 (360 monthly observations). As the initial data point of the VIX index is January 1990, the directional indicator begins in February 1990. The full sample is divided into an in-sample and an out-of-sample period; the in-sample period ranges from February 1990 to December 2003 (167 monthly observations), while an out-of-sample period ranging from January 2004 to December 2019 (192 monthly observations) is used to evaluate predictive performance. Out-of-sample



forecasts are generated recursively, using a forecast horizon of 1-month.

When CART techniques are implemented, each classification model is trained; the training dataset is preprocessed both in a closed center and scale form, the parameter(s) of each model is tuned by cross-validation as well as resampling, and variable importance is determined before out-of-sample forecast generation. The resampling approach in specific, seeks to determine the values of each of the model parameters (if any) and uses the best tuning parameter(s) based on fitted (in-sample) accuracy measures to produce out-of-sample forecasts. In each model, the best tuning parameter(s) are used to run the out-of-sample forecasts recursively, and their respective performance evaluation measures are obtained.

#### 4. Empirical design and analysis

The empirical findings of the analysis are presented in this section. The main objective is to evaluate the out-of-sample statistical and economic performance of the different modeling approaches under consideration. First, the out-of-sample predictive performance of the underlying models will be investigated based on several statistical evaluation measures. Next, the economic significance of the generated forecasts will be assessed through the creation of real-time investment strategies.

Initially, the out of sample performance of various univariate Logit regression models is examined using the set of predictor variables, as well as autoregressive components. Motivated by the inferior performance of these basic models, the attention is directed to alternative modeling approaches to find out whether they are able to substantially improve the underlying predictive ability. In this sense, the implemented class of models is extended to a multiple Logit regression framework, and the use of penalized likelihood Logit regression models, as well as other machine learning techniques.

In general, under the multiple regression framework estimation of the model parameters is straightforward, the identification of the most important set of predictors, however, is not an easy task, due to the high dimensionality of the problem. More specifically, with a set of  $N = 31$  predictor variables, using a Logit autoregressive model of order three, i.e.  $p = 3$ , and individual predictor variables with one up to three lagged terms, i.e.  $h = 3$ , the maximum number of potential lagged predictors is 96, and the number of competing models vast. Different model/variable selection strategies have been used in the literature to address this issue. Stepwise regression (and/or its variants), standard information criteria such as Akaike's (AIC) 1973 or Schwarz's 1978 Bayesian Information Criterion (BIC), the Bayesian approach to model selection, or penalized likelihood methods, such as the Elastic Net, could be used to narrow down and identify important predictors. While the majority of these approaches is implemented, the Elastic Net regularization approach, that imposes shrinkage in the regression coefficients and allows for automatic variable selection, is the core approach followed in this study, and in a real-time out-of-sample framework in specific.

The total sample is divided into an in-sample and an out-of-sample period to assess the predictive performance of different modeling approaches and machine learning techniques. To obtain out-of-sample forecasts, the model selection algorithm that identifies important predictors is implemented, and the predictive models are estimated recursively using an expanding window scheme. That is, the model selection algorithm and the fitted models are first estimated using data from the beginning of the dataset to the end date of the in-sample period and a one month-ahead forecast is generated. Then, the estimation window is updated by adding one observation to the estimation sample at each step. As such, model selection is conducted and coefficients are re-estimated after each step of the recursion. Proceeding iteratively in this way through the end of the out-of-sample period, a series of  $P$  out-of-sample forecasts is generated.

The proposed algorithm is briefly presented below:

- (1) Set the out-of-sample period  $P$ .
- (2) For iteration  $= 1, \dots, P$ 
  - (a) Specify the estimation window, which is updated based on a recursive (expanding) scheme
  - (b) Apply the Elastic Net regularization approach and identify the important predictors
  - (c) Implement the proposed modeling approaches, i.e. the machine learning techniques and the standard predictive models
  - (d) Compute VIX directional one-step ahead forecasts
  - (e) Construct different investment strategies
- (3) Compute Statistical and Economic Evaluation measures

##### 4.1. Predictive performance evaluation based on statistical metrics

The results of the statistical performance evaluation of the analyzed modeling approaches are presented in this section. The out-of-sample performance of the directional VIX forecasts obtained by the univariate Logit models based on autoregressive (lagged) terms and individual predictor variables of the form  $\ln\left(\frac{p_t}{1-p_t}\right) = \phi_0 + \sum_{l=1}^p \phi_l y_{t-l}$ , and  $\ln\left(\frac{p_t}{1-p_t}\right) = \beta_0 + \sum_{j=1}^h \beta_{ij} x_{i,t-j}$ ,  $i = 1, \dots, N$ , respectively, are presented first, in order to illustrate the potential benefits and the motivation behind the use of machine learning techniques.

Table 1 (panel A) presents the statistical evaluation metrics for the VIX index directional forecasts obtained by the univariate Logit models. The majority of the performance evaluation metrics reported, are based on the confusion matrix, which summarizes the forecasting outcomes regarding the classification problem. A detailed description and analysis of several performance evaluation measures used in such problems can be found in Sokolova and Lapalme (2009), Hossin and Sulaiman (2015) and Tharwat (2020), among several others. In particular, metrics such as Area Under the ROC curve (AUC), Misclassification Error (MCE), Accuracy, Kappa statistic, Sensitivity and Specificity portion, Precision,  $F_1$ -Score, and Balanced Accuracy are presented. For all metrics, with the exception of MCE, high positive values indicate superior forecasting performance. In addition, the statistical

Table 1. Univariate and multiple Logit models: Statistical Performance Evaluation.

Method	AUC	MCE	Accuracy	Kappa	Sensitivity	Specificity	Precision	$F_1$ Score	Balanced Accuracy	Pesaran - Timmermann
Panel A: Univariate Logit Models										
VIX-lagged	0.617	0.401	0.599	0.192	0.570	0.623	0.551	0.560	0.596	2.659
SMB	0.513	0.500	0.500	-0.070	0.151	0.783	0.361	0.213	0.467	-1.162
HML	0.548	0.505	0.495	-0.067	0.209	0.726	0.383	0.271	0.468	-1.030
MKT	0.608	0.422	0.578	0.133	0.442	0.689	0.535	0.484	0.565	1.863
MOM	0.537	0.453	0.547	0.030	0.198	0.830	0.486	0.281	0.514	0.497
STR	0.592	0.453	0.547	-0.003	0.035	0.962	0.429	0.065	0.499	-0.105
RVAR	0.573	0.479	0.521	-0.013	0.233	0.755	0.435	0.303	0.494	-0.205
BAB	0.573	0.484	0.516	-0.062	0.047	0.896	0.267	0.079	0.471	-1.470
QMJ	0.485	0.500	0.500	-0.050	0.244	0.708	0.404	0.304	0.476	-0.748
Term Spread	0.559	0.531	0.469	-0.126	0.163	0.717	0.318	0.215	0.440	-1.971
Credit Spread	0.593	0.484	0.516	-0.050	0.105	0.849	0.360	0.162	0.477	-0.948
10-year Treas.	0.617	0.531	0.469	-0.144	0.081	0.783	0.233	0.121	0.432	-2.573
3-month T-bill	0.603	0.474	0.526	-0.047	0.023	0.934	0.222	0.042	0.479	-1.395
Oil Price	0.488	0.495	0.505	-0.026	0.314	0.660	0.429	0.362	0.487	-0.377
Dollar Index	0.579	0.505	0.495	-0.087	0.116	0.802	0.323	0.171	0.459	-1.532
TED Spread	0.588	0.474	0.526	-0.029	0.105	0.868	0.391	0.165	0.486	-0.582
Skew	0.494	0.505	0.495	-0.018	0.453	0.528	0.438	0.446	0.491	-0.252
CF Leverage	0.486	0.464	0.536	0.021	0.256	0.764	0.468	0.331	0.510	0.320
TS MOM	0.561	0.469	0.531	0.011	0.256	0.755	0.458	0.328	0.505	0.168
MOM CM	0.536	0.474	0.526	-0.020	0.151	0.830	0.419	0.222	0.491	-0.349
MOM EQ	0.570	0.458	0.542	0.059	0.407	0.651	0.486	0.443	0.529	0.824
MOM FI	0.488	0.484	0.516	-0.014	0.279	0.708	0.436	0.340	0.493	-0.204
MOM FX	0.532	0.469	0.531	0.018	0.291	0.726	0.463	0.357	0.509	0.262
HMLD	0.537	0.432	0.568	0.095	0.337	0.755	0.527	0.411	0.546	1.401
ECU	0.513	0.479	0.521	-0.004	0.279	0.717	0.444	0.343	0.498	-0.061
MACVOL	0.540	0.448	0.552	0.028	0.140	0.887	0.500	0.218	0.513	0.549
Bullish Aver.	0.551	0.474	0.526	-0.039	0.058	0.906	0.333	0.099	0.482	-0.929
Neutral Aver.	0.529	0.490	0.510	-0.050	0.151	0.802	0.382	0.217	0.477	-0.847
Bearish Aver.	0.523	0.453	0.547	0.004	0.070	0.934	0.462	0.121	0.502	0.102
Spread Aver.	0.494	0.469	0.531	-0.029	0.058	0.915	0.357	0.100	0.487	-0.709
BB Ratio Aver.	0.567	0.490	0.510	-0.080	0.012	0.915	0.100	0.021	0.463	-2.272
8-week MAVB	0.542	0.479	0.521	-0.040	0.105	0.858	0.375	0.164	0.482	-0.768
Panel B: Multiple Logit Models—Stepwise approach										
AIC	0.525	0.453	0.547	0.093	0.547	0.547	0.495	0.519	0.547	1.291
BIC	0.606	0.385	0.615	0.214	0.523	0.689	0.577	0.549	0.606	2.973
Deviance	0.597	0.391	0.609	0.206	0.535	0.670	0.568	0.551	0.602	2.856

significance of the corresponding forecasts is assessed by the Pesaran and Timmermann (1992) test statistic. Positive (negative) values of the Pesaran–Timmermann (PT) statistic indicate that the true or hit rate is larger (smaller) than the false rate, and, therefore, higher (than the critical value) positive test statistic values indicate significant forecasting ability. The results presented in table 1 suggest that a Logit autoregressive model of order three, i.e. with three VIX lagged terms, emerges as the best predictor, as it has the highest (and statistically significant) PT statistic (2.659). Moreover, the predictive accuracy of the specific model is relatively high based on the elevated values of AUC (61.7%), Accuracy (59.9%),  $F_1$ -Score (0.56), Balanced Accuracy (59.6%), and the lowest value of the MCE (40.1%). Not surprisingly, another useful predictor is the market variable, MKT, with high values in terms of AUC (60.8%), Accuracy (57.8%), Precision (53.5%), Balanced Accuracy (56.5%), and the PT statistic (1.863), as well as a relatively low (42.2%) MCE. It is evident that only a limited number of individual predictors are successful in forecasting the future direction of the VIX. The results are consistent and confirm previous studies, such as Konstantinidi *et al.* (2008), that point out that autoregressive

terms and the market variable are important predictors for the evolution of implied volatility.

The most rational move is to focus on multiple Logit regression models, where the basic set up remains the same, i.e. both autoregressive components and predictor variables up to lag three, i.e.  $p = 3$  and  $h = 3$  are considered. A real-time stepwise regression framework based on the Akaike Information Criterion (AIC), the Schwarz Bayesian information criterion (BIC), and the Deviance metrics was implemented. Table 1 (panel B) reports the relevant statistical measures that indicate that the stepwise Logit model based on BIC and Deviance outperform the corresponding model based on the AIC criterion, and are slightly better than the univariate Logit model with three autoregressive terms, but clearly superior compared to all the rest of the univariate models, as they produce higher AUC (around 60%), lower MCE (around 39%), higher Accuracy (around 61%), kappa statistic (around 0.21), Precision (around 57%, 58%), and a significant statistical predictive performance based on the PT statistic (2.973 and 2.856, respectively). It is noteworthy that the inferior results of the AIC based model might be due to its tendency to overestimate the number of predictors (possible overfitting as it

Table 2. Elastic Net—Deviance Model ( $\alpha = 0.50$ )—Selected predictor variables and frequencies.

Variable	Lag	Inclusion frequency (# of times)	Inclusion frequency (%)
VIX	1	192	100.00%
MOM EQ	1	192	100.00%
MKT	1	190	98.96%
RVAR	1	135	70.31%
Oil Price	1	49	25.52%
MOM	1	14	7.29%
VIX	3	11	5.73%
QMJ	2	10	5.21%
HMLD	3	10	5.21%
Skew	3	9	4.69%
MOM FI	3	8	4.17%
TS MOM	1	6	3.13%
MOM FX	2	5	2.60%
QMJ	3	4	2.08%
ECU	2	2	1.04%
MOM	2	2	1.04%
Term Spread	1	2	1.04%
TS MOM	3	1	0.52%
MOM FI	2	1	0.52%

selects a larger number of predictors in each out-of-sample iteration) compared to the BIC and Deviance criteria.

The next step is to use a penalized likelihood method for variable selection. More specifically, the Elastic Net

Table 3. Elastic Net—Deviance Model ( $\alpha = 0.75$ )—Selected predictor variables and frequencies.

Variable	Lag	Inclusion frequency (# of times)	Inclusion frequency (%)
VIX	1	192	100.00%
MOM EQ	1	192	100.00%
MKT	1	178	92.71%
RVAR	1	121	63.02%
Oil Price	1	38	19.79%
QMJ	2	8	4.17%
Skew	3	8	4.17%
VIX	3	7	3.65%
HMLD	3	6	3.13%
MOM	1	5	2.60%
MOM FX	2	4	2.08%
QMJ	3	3	1.56%
MOM FI	3	3	1.56%
Term Spread	1	1	0.52%
ECU	2	1	0.52%

regularization method based on the Deviance metric is used for predictor selection and inclusion. Table 2 presents the predictive variables that are included at each out-of-sample iteration in the Elastic Net Deviance model using an  $\alpha = 0.50$ , ranked on their underlying frequency of inclusion. The results indicate that the first order autoregressive component, lagged binary variable, ( $VIX_{t-1}$ ), and the medium term equity market

Table 4. Machine learning techniques: Statistical Performance Evaluation – Preselection with Elastic Net ( $\alpha = 0.50$ ).

Method	AUC	MCE	Accuracy	Kappa	Sensitivity	Specificity	Precision	$F_1$ Score	Balanced Accuracy	Pesaran - Timmermann
Panel A: Ridge Logit Models										
Ridge AUC	0.589	0.438	0.562	0.066	0.233	0.830	0.526	0.323	0.531	1.085
Ridge Class	0.624	0.391	0.609	0.190	0.430	0.755	0.587	0.497	0.592	2.714
Ridge Dev.	0.627	0.349	0.651	0.292	0.593	0.698	0.614	0.604	0.646	4.049
Panel B: LASSO Logit Models										
LASSO AUC	0.621	0.365	0.635	0.261	0.581	0.679	0.595	0.588	0.630	3.620
LASSO Class	0.628	0.370	0.630	0.245	0.535	0.708	0.597	0.564	0.621	3.408
LASSO Dev.	0.626	0.359	0.641	0.273	0.593	0.679	0.600	0.596	0.636	3.777
Panel C: Elastic Net Logit Models										
EN AUC	0.622	0.359	0.641	0.273	0.593	0.679	0.600	0.596	0.636	3.777
EN Class	0.629	0.375	0.625	0.235	0.535	0.698	0.590	0.561	0.616	3.269
EN Dev.	0.626	0.359	0.641	0.273	0.593	0.679	0.600	0.596	0.636	3.777
Panel D: Discriminant Analysis Models										
LDA	0.623	0.354	0.646	0.284	0.605	0.679	0.605	0.605	0.642	3.934
RDA	0.636	0.370	0.630	0.264	0.674	0.594	0.574	0.620	0.634	3.709
Panel E: Bayesian Models										
Naive Bayes	0.679	0.344	0.656	0.318	0.721	0.604	0.596	0.653	0.662	4.490
Bayes GLM	0.625	0.365	0.635	0.263	0.593	0.670	0.593	0.593	0.631	3.642
Panel F: Trees, Forest and Boosting Techniques										
Bagging	0.613	0.406	0.594	0.182	0.570	0.613	0.544	0.557	0.591	2.527
Random Forest	0.642	0.396	0.604	0.200	0.558	0.642	0.558	0.558	0.600	2.766
Ada. Boosting	0.662	0.344	0.656	0.306	0.628	0.679	0.614	0.621	0.654	4.248
Grad. Boosting	0.640	0.396	0.604	0.207	0.605	0.604	0.553	0.578	0.604	2.873
Panel G: K-nn and Recursive Partitioning										
knn	0.630	0.411	0.589	0.169	0.547	0.623	0.540	0.543	0.585	2.341
rpart	0.621	0.385	0.615	0.219	0.558	0.660	0.571	0.565	0.609	3.035
Panel H: Combination of forecasts										
Mean (k = 1)	0.639	0.375	0.625	0.221	0.442	0.774	0.613	0.514	0.608	3.175
Mean (k = 2)	0.643	0.359	0.641	0.268	0.558	0.708	0.608	0.582	0.633	3.720
Median (k = 1)	0.645	0.375	0.625	0.221	0.442	0.774	0.613	0.514	0.608	3.175
Median (k = 2)	0.629	0.370	0.630	0.251	0.581	0.670	0.588	0.585	0.626	3.485
BMA	0.621	0.370	0.630	0.247	0.547	0.698	0.595	0.570	0.622	3.425

momentum factor,  $MOMEQ_{t-1}$ , are selected by the algorithm at all the out-of-sample iterations. Another important predictor is the equity market factor,  $MKT_{t-1}$ , with frequency of inclusion of around 99%. In addition, the residual variance factor,  $RVAR_{t-1}$ , and the oil price,  $Oil_{t-1}$  are part of the top ranked list, with frequency of inclusion about 70% and 25%, respectively. Other predictors that affect the VIX direction, but, to a smaller extent, and at different lags, are the Fama-French momentum factor,  $MOM_{t-1}$ , the lagged autoregressive components of VIX,  $VIX_{t-3}$ , the quality minus junk factor,  $QMJ_{t-2}$ , and the high minus low devil factor,  $HMLD_{t-3}$ .

The list of relevant predictors is not surprising; the lagged binary variable and the market factor have already exhibited their significance in the univariate Logit model setting. In addition, the presence of medium-term price momentum factors is consistent with empirical evidence that show that they are significant for asset return predictability. The low volatility, quality, and value-related factor portfolios approximate equity market risk factors. Last, but not least, it has been shown that commodity markets tend to be forward-looking and exhibit leading properties; as such, the inclusion of the oil price is not unforeseen.

To assess the sensitivity of the selected predictors, the variable selection exercise is repeated using the Elastic Net Deviance model, but with an  $\alpha = 0.75$ . Table 3 presents the

selected predictors together with their inclusion frequency across the out-of-sample iterations. The results are robust, as they are similar to those of table 2, as the first autoregressive component, the equity market momentum factor (MOM EQ), the equity market factor (MKT), the residual variance factor (RVAR) and the oil price (Oil price) are, once again, identified as important predictors.

Having preselected the set of predictors, next, the ability of the machine learning techniques is evaluated with the aim to provide even more accurate VIX directional forecasts. Table 4 reports the statistical evaluation measures for different modeling approaches (panels A to H). More specifically, the analysis involves the application of the Ridge, LASSO and Elastic Net regularization techniques, Discriminant analysis and Bayesian methods, as well as Classification Tree related techniques,  $k$ -nn and partitioning methods. Useful insights emerge from the reported results. It is evident that the majority of the machine learning techniques produces significant predictive performance based on the high positive values of the PT statistic and the different confusion matrix-related measures. In particular, the Naive Bayes and the Adaptive boosting methods seem to be the top performing models, which clearly outperform the univariate and the multiple stepwise Logit regression models, as well as the other machine learning techniques. Naive Bayes and Adaptive

Table 5. Machine Learning Techniques: Statistical Performance Evaluation—Preselection with Elastic Net ( $\alpha = 0.75$ ).

Method	AUC	MCE	Accuracy	Kappa	Sensitivity	Specificity	Precision	$F_1$ Score	Balanced Accuracy	Pesaran - Timmermann
Panel A: Ridge Logit Models										
Ridge AUC	0.614	0.422	0.578	0.099	0.244	0.849	0.568	0.341	0.547	1.629
Ridge Class	0.620	0.391	0.609	0.192	0.442	0.745	0.585	0.503	0.594	2.725
Ridge Dev.	0.631	0.344	0.656	0.303	0.605	0.698	0.619	0.612	0.651	4.205
Panel B: LASSO Logit Models										
LASSO AUC	0.628	0.365	0.635	0.263	0.593	0.670	0.593	0.593	0.631	3.642
LASSO Class	0.635	0.370	0.630	0.247	0.547	0.698	0.595	0.570	0.622	3.425
LASSO Dev.	0.632	0.359	0.641	0.274	0.605	0.670	0.598	0.601	0.637	3.799
Panel C: Elastic Net Logit Models										
EN AUC	0.629	0.359	0.641	0.274	0.605	0.670	0.598	0.601	0.637	3.799
EN Class	0.638	0.365	0.635	0.258	0.558	0.698	0.600	0.578	0.628	3.582
EN Dev.	0.630	0.359	0.641	0.274	0.605	0.670	0.598	0.601	0.637	3.799
Panel D: Discriminant Analysis Models										
LDA	0.630	0.354	0.646	0.284	0.605	0.679	0.605	0.605	0.642	3.934
RDA	0.642	0.365	0.635	0.272	0.663	0.613	0.582	0.620	0.638	3.804
Panel E: Bayesian Models										
Naive Bayes	0.682	0.339	0.661	0.327	0.709	0.623	0.604	0.652	0.666	4.581
Bayes GLM	0.630	0.359	0.641	0.274	0.605	0.670	0.598	0.601	0.637	3.799
Panel F: Trees, Forest and Boosting Techniques										
Bagging	0.620	0.406	0.594	0.179	0.547	0.632	0.547	0.547	0.589	2.475
Random Forest	0.644	0.396	0.604	0.198	0.547	0.651	0.560	0.553	0.599	2.743
Ada. Boosting	0.673	0.359	0.641	0.277	0.628	0.651	0.593	0.610	0.639	3.848
Grad. Boosting	0.639	0.375	0.625	0.238	0.558	0.679	0.585	0.571	0.619	3.307
Panel G: K-nn and Recursive Partitioning										
knn	0.626	0.417	0.583	0.158	0.535	0.623	0.535	0.535	0.579	2.183
rpart	0.626	0.365	0.635	0.261	0.581	0.679	0.595	0.588	0.630	3.620
Panel H: Combination of forecasts										
Mean (k = 1)	0.633	0.391	0.609	0.193	0.453	0.736	0.582	0.510	0.595	2.737
Mean (k = 2)	0.638	0.359	0.641	0.269	0.570	0.698	0.605	0.587	0.634	3.738
Median (k = 1)	0.635	0.391	0.609	0.193	0.453	0.736	0.582	0.510	0.595	2.737
Median (k = 2)	0.621	0.365	0.635	0.263	0.593	0.670	0.593	0.593	0.631	3.642
BMA	0.621	0.370	0.630	0.247	0.547	0.698	0.595	0.570	0.622	3.425



boosting achieve top rankings for almost all the statistical measures reported in table 4. They produce the smallest MCE of 34.4% (versus 40.1% of the univariate Logit autoregressive model, and 38.5% of the multiple stepwise BIC model), the highest AUC (67.9% and 66.2%), Accuracy (65.6%), kappa statistic (0.318 and 0.306), Sensitivity (72.1% and 62.8%), Balanced Accuracy (66.2% and 65.4%) as well as the highest PT statistic (4.49 and 4.248, respectively). This is a sizeable and non-negligible improvement of 5.7% over the univariate Logit autoregressive model, based on the Accuracy metric. Having said that, notable improvements emerge for nearly all measures. Additional modeling approaches that rank among the most accurate are the Ridge-Deviance and the Linear Discriminant Analysis (LDA), followed by the Elastic Net (AUC and Deviance), the LASSO, the Regularized Discriminant Analysis (RDA), and the Bayes GLM.

To assess the robustness of the results the real-time out-of-sample forecasting exercise is repeated using the Elastic Net Deviance model with an  $\alpha = 0.75$ . Table 5 presents the statistical evaluation measures for the different models and confirms the superior predictive performance of the Naive Bayes method over the competing techniques, since it ranks first with respect to almost all statistical measures. The predictive ability of the Naive Bayes method has been reported in the literature; see, for example Domingos and Pazzani (1997) and the references therein. Additional modeling approaches that produce high predictive ability and rank among the best performers are the Ridge Deviance model and the Adaptive Boosting method. In general, the results are robust, and similar in spirit with those of table 4, as the majority of the machine learning techniques produces significant out-of-sample performance.

Table 6. Univariate and Multiple Logit models: Economic Performance Evaluation.

Method	Strategy	Annual. Return	Annual. Risk	Sharpe Ratio	Sortino Ratio	Downside Risk	Average Drawdown	Alpha (Equity & Cash)	Beta (Equity & Cash)
Panel A: VIX Long									
	VIX Long	− 1.76	70.54	− 0.04	− 0.08	38.61	− 71.33	2.2	− 7.23
Panel B: Univariate Logit Models									
VIX Lagged	Strategy 1	25.33	45.65	0.53	0.82	29.18	− 26.34	2.84	− 2.86
	Strategy 2	59.11	69.20	0.84	1.44	40.07	− 40.78	3.45	1.50
	Strategy 3	29.10	46.87	0.59	1.11	25.07	− 25.70	2.66	− 1.53
	Strategy Con. 2	44.82	61.99	0.70	0.97	45.09	− 51.53	2.52	1.85
	Strategy Con. 3	16.41	39.82	0.38	0.55	27.33	− 45.23	1.57	− 0.91
MKT	Strategy 1	14.09	31.43	0.41	0.53	24.01	− 23.75	1.59	− 1.47
	Strategy 2	32.25	70.07	0.44	0.58	53.17	− 63.37	0.94	4.29
	Strategy 3	17.73	33.60	0.49	0.81	20.37	− 35.77	1.33	0.11
	Strategy Con. 2	40.69	50.04	0.79	0.77	51.27	− 42.67	1.83	3.20
	Strategy Con. 3	6.69	20.87	0.26	0.30	18.35	− 24.24	0.31	0.71
HMLD	Strategy 1	15.05	46.25	0.30	0.29	46.84	− 31.78	2.21	− 3.14
	Strategy 2	34.46	70.01	0.47	0.75	44.48	− 37.79	2.18	0.95
	Strategy 3	24.94	47.24	0.50	0.65	36.41	− 23.76	2.51	− 1.93
	Strategy Con. 2	9.57	46.60	0.18	0.16	52.73	− 41.87	0.48	0.87
	Strategy Con. 3	3.77	30.99	0.08	0.06	45.02	− 18.81	0.50	− 0.57
MOM EQ	Strategy 1	3.67	44.24	0.05	0.06	42.79	− 41.42	1.21	− 2.74
	Strategy 2	9.37	70.50	0.11	0.17	48.44	− 74.44	0.17	1.75
	Strategy 3	8.17	45.41	0.15	0.20	35.29	− 39.65	1.19	− 1.60
	Strategy Con. 2	38.06	57.57	0.64	0.77	47.96	− 49.13	2.16	1.70
	Strategy Con. 3	14.57	36.65	0.36	0.44	30.36	− 19.65	1.41	− 0.81
Panel C: Multiple Logit Models—Stepwise approach									
AIC	Strategy 1	11.37	46.68	0.22	0.29	35.17	− 54.91	1.73	− 2.50
	Strategy 2	26.09	70.22	0.35	0.54	45.55	− 73.98	1.21	2.23
	Strategy 3	13.74	48.01	0.26	0.40	30.90	− 57.78	1.45	− 1.12
	Strategy Con. 2	32.70	65.49	0.48	0.66	47.40	− 69.03	1.76	1.88
	Strategy Con. 3	14.67	45.82	0.29	0.43	31.48	− 58.08	1.49	− 1.04
BIC	Strategy 1	30.20	43.81	0.66	1.00	28.82	− 26.69	3.14	− 2.76
	Strategy 2	71.46	68.72	1.02	1.66	42.30	− 47.17	4.03	1.70
	Strategy 3	38.03	44.77	0.82	1.52	24.20	− 19.74	3.24	− 1.57
	Strategy Con. 2	53.08	60.12	0.86	1.13	45.68	− 45.80	3.02	1.80
	Strategy Con. 3	23.81	38.01	0.59	1.01	22.27	− 16.68	2.10	− 0.91
Deviance	Strategy 1	28.89	44.35	0.62	0.97	28.62	− 29.09	3.06	− 2.78
	Strategy 2	68.10	68.85	0.97	1.59	42.13	− 48.42	3.87	1.67
	Strategy 3	36.44	45.32	0.78	1.43	24.54	− 22.23	3.15	− 1.59
	Strategy Con. 2	52.49	60.44	0.85	1.14	44.98	− 46.60	3.02	1.69
	Strategy Con. 3	25.01	38.79	0.61	1.08	21.93	− 16.07	2.21	− 1.01

As a final check, a series of combined<sup>†</sup> implied volatility probability forecasts is generated based on linear weighting schemes, as well as Bayesian methods (Bayesian Model Averaging). Even though the accuracy of the forecasts generated by the combination methods is, in general, higher compared to the univariate Logit and stepwise Logit regression models, in most cases they fail to outperform the higher ranked machine learning techniques.

In conclusion, the evidence clearly supports the implementation of machine learning techniques in implied volatility prediction, as it significantly improves the accuracy of out-of-sample forecasts, based on numerous statistical evaluation measures.

#### 4.2. Economic performance evaluation based on real-time trading strategies

To evaluate the economic significance of the forecasts generated by each of the estimated models, a number of investment strategies based on the VIX index and the S&P 500 index have been devised. The strategies are implemented also for models with inferior statistical performance, as statistical evidence is not always confirmed by the underlying financial performance.

A good starting point would be to focus on the performance of the most naive strategy. That would result in being constantly long the VIX index. It is not surprising that the specific strategy would constitute a suboptimal investment choice, as the long periods of VIX range trading combined with the relatively short periods of sizeable rise are clearly against long positioning. This is reflected in the performance evaluation metrics (see table 6, panel A) with negative Sharpe and Sortino ratio readings, as well as high average draw downs. Table 6 presents the various metrics, including an estimated alpha and a beta coefficient versus a reference benchmark that consists of 50% cash (the return on 1-month Treasury bill) and 50% equities (the return of the S&P 500 index). The specific composition tries to reflect an investor that is long an equity portfolio, but at the same time aims to protect his downside risk.

The first strategy (Strategy 1) goes long the VIX index when implied volatility is expected to rise, and remains on the sidelines when implied volatility is expected to fall. The second strategy (Strategy 2) is leveraged, as it invests in the VIX index when implied volatility is forecast to rise and shorts the VIX index when it is forecast to decline. In the third strategy (Strategy 3), the S&P 500 index is bought when implied

volatility is expected to trend lower. For robustness and confidence reasons variants of Strategies 2 and 3 are also formed. More specifically, when the underlying forecast lies between the 45%-55% probability interval, no position is taken; if the estimated probability is above 55% the strategy goes long the VIX index, alternatively, when the probability is below 45% the strategy goes short the VIX index (similar to Strategy 2) and long the S&P 500 index (similar to Strategy 3). The specific strategies are tagged Confidence Strategy 2 (Strategy Con. 2) and Confidence Strategy 3 (Strategy Con. 3), respectively. All the outlined strategies are fully implementable, as the recent rise in passive investing or indexing has enabled investors gain exposure to previously inaccessible market segments through innovative index-linked financial products. More specifically, all strategies can be executed through the use of Exchange Traded Funds (ETFs), as nowadays there are available long and short VIX-related Short-Term ETFs. No transaction costs are taken into account in the calculations.

Table 6 (panel B) shows the economic performance evaluation metrics of a selected number of univariate Logit models. More specifically, the performance of the four top ranked models, i.e. of the VIX lagged autoregressive, the equity market factor (MKT), the High minus Low Devil factor (HMLD), and the equity market momentum factor (MOM EQ) model, based on the statistical evaluation measures, such as the PT statistic, Balanced Accuracy, Precision, is presented. The results reaffirm the predictive ability of the autoregressive model, as well as the market factor one, as they achieve relatively high risk adjusted performance as reflected in the Sharpe and Sortino ratios, as well as the high estimated alphas versus the reference benchmark. In general, the rest of the univariate models produce suboptimal performance.

Next, the economic performance of the multiple Logit models based on the real-time stepwise regression framework (AIC, BIC and Deviance metrics) is reported in table 6 (panel C) and suggest that an investor that generates forecasts through the stepwise BIC approach will achieve higher positive annualized returns, ranging from 23.81% (Strategy Con. 3) to 71.46% (Strategy 2), compared to the corresponding returns based on the univariate Logit models, as well as those of the stepwise AIC and Deviance approach. Moreover, the BIC-related forecasts attain higher risk adjusted performance, as the Sharpe and Sortino ratios record values ranging from 0.59 (Strategy Con. 3) to 1.02 (Strategy 2), and from 1.00 (Strategy 1) to 1.66 (Strategy 2), respectively. Strategy 2 provides the investor with the highest annualized returns (71.46%) and also higher risk adjusted performance metrics, i.e. Sharpe (1.02) and Sortino (1.66) ratios. The stepwise Deviance approach ranks second with relatively high economic performance metrics.

The economic performance evaluation measures related to the machine learning techniques is depicted in table 7 ( $\alpha = 0.5$  in the Elastic Net variable selection algorithm). As in the case of the univariate and multiple Logit regression models, the results mostly confirm the statistical evaluation rankings. Investment strategies constructed based on forecasts generated by the Naive Bayes, Ridge Deviance, Discriminant Analysis models, and combination of forecasts produce the best annualized returns and risk adjusted performance measures. In particular, the Naive Bayes and the Ridge Deviance

<sup>†</sup> Elliott *et al.* (2013) introduced a new method for combining forecasts based on complete subset regressions. For a given set of potential predictors, the authors propose combining forecasts from all possible linear regressions that keep the number of predictors fixed. For  $K$  possible predictors, there are  $K$  univariate models and  $N_{k,K} = K!/((K-k)!k!)$  different  $k$ -variate models for  $k \leq K$ . The set of models for a fixed value of  $k$  is referred to as a complete subset and the authors propose to use equal-weighted combinations of the forecasts from all models within these subsets indexed by  $k$ . See, also, Meligkotsidou *et al.* (2019) and Meligkotsidou *et al.* (2021), who developed quantile forecast combination schemes for realized volatility prediction and a complete subset quantile regression approach for equity premium prediction.

Table 7. Machine Learning Techniques: Economic Performance Evaluation—Preselection with Elastic Net ( $\alpha = 0.50$ ).

Method	Strategy	Annual. Return	Annual. Risk	Sharpe Ratio	Sortino Ratio	Downside Risk	Average Drawdown	Alpha (Equity & Cash)	Beta (Equity & Cash)
Ridge AUC	VIX Long	− 1.76	70.54	− 0.04	− 0.08	38.61	− 71.33	2.2	− 7.23
	Strategy 1	8.57	21.41	0.34	0.33	22.41	− 17.38	0.90	− 0.63
	Strategy 2	19.88	70.35	0.26	0.34	54.55	− 79.37	− 0.45	5.97
	Strategy 3	13.69	24.76	0.50	0.78	16.00	− 20.41	0.72	1.09
	Strategy Con. 2	16.90	41.19	0.38	0.30	52.61	− 58.81	0.34	2.92
Ridge Class	Strategy Con. 3	4.12	14.10	0.20	0.18	15.54	− 24.51	0.04	0.90
	Strategy 1	25.29	37.20	0.65	1.05	22.92	− 16.09	2.56	− 2.01
	Strategy 2	59.01	69.21	0.83	1.27	45.44	− 50.55	2.88	3.20
	Strategy 3	32.65	38.78	0.81	1.65	18.98	− 19.29	2.55	− 0.49
	Strategy Con. 2	33.99	46.38	0.71	0.75	43.65	− 31.97	1.49	2.96
Ridge Dev.	Strategy Con. 3	3.22	19.20	0.10	0.11	18.26	− 19.54	0.03	0.72
	Strategy 1	42.32	42.34	0.97	1.87	21.99	− 21.39	3.79	− 2.44
	Strategy 2	103.91	67.27	1.53	2.33	43.98	− 43.44	5.34	2.35
	Strategy 3	49.46	43.50	1.11	2.45	19.66	− 25.55	3.76	− 1.06
	Strategy Con. 2	46.26	53.53	0.84	0.92	49.16	− 43.37	2.26	2.92
LASSO AUC	Strategy Con. 3	17.29	28.25	0.57	0.86	18.59	− 20.78	1.27	0.22
	Strategy 1	38.22	42.78	0.86	1.66	22.30	− 20.29	3.51	− 2.36
	Strategy 2	92.65	67.80	1.35	2.12	43.07	− 42.57	4.79	2.51
	Strategy 3	44.77	44.02	0.99	2.16	20.15	− 26.31	3.44	− 0.93
	Strategy Con. 2	59.17	58.25	0.99	1.21	47.77	− 42.61	3.13	2.48
LASSO Class	Strategy Con. 3	28.30	36.51	0.74	1.46	18.48	− 21.32	2.18	− 0.24
	Strategy 1	29.65	38.31	0.74	1.22	23.19	− 13.93	2.83	− 1.96
	Strategy 2	70.03	68.77	1.00	1.42	48.42	− 46.09	3.43	3.31
	Strategy 3	35.68	39.83	0.86	1.71	20.17	− 20.87	2.73	− 0.48
	Strategy Con. 2	47.85	55.64	0.84	0.95	49.19	− 36.91	2.33	2.96
LASSO Dev.	Strategy Con. 3	16.37	31.43	0.48	0.81	18.54	− 16.23	1.21	0.20
	Strategy 1	39.96	42.76	0.90	1.76	21.95	− 23.10	3.65	− 2.45
	Strategy 2	97.38	67.58	1.42	2.22	43.38	− 45.24	5.06	2.33
	Strategy 3	46.79	43.93	1.04	2.28	19.99	− 27.00	3.60	− 1.07
	Strategy Con. 2	61.51	58.42	1.03	1.26	47.77	− 42.32	3.25	2.50
Elastic Net AUC	Strategy Con. 3	28.45	36.51	0.74	1.47	18.48	− 21.26	2.19	− 0.24
	Strategy 1	39.51	42.88	0.89	1.71	22.30	− 22.71	3.62	− 2.45
	Strategy 2	96.17	67.63	1.40	2.19	43.37	− 44.54	5.01	2.32
	Strategy 3	46.30	44.05	1.02	2.22	20.29	− 26.70	3.58	− 1.08
	Strategy Con. 2	60.38	58.38	1.01	1.24	47.77	− 42.03	3.19	2.50
Elastic Net Class	Strategy Con. 3	27.93	36.49	0.73	1.44	18.48	− 21.24	2.15	− 0.24
	Strategy 1	29.64	39.27	0.72	1.24	22.90	− 19.31	2.84	− 1.97
	Strategy 2	70.01	68.78	1.00	1.47	46.89	− 49.80	3.43	3.29
	Strategy 3	35.22	40.77	0.83	1.69	20.14	− 25.43	2.71	− 0.49
	Strategy Con. 2	46.99	51.18	0.89	0.96	47.53	− 29.45	2.37	2.70
Elastic Net Dev.	Strategy Con. 3	13.05	24.40	0.48	0.67	17.55	− 13.08	0.94	0.27
	Strategy 1	39.96	42.76	0.90	1.76	21.95	− 23.10	3.65	− 2.45
	Strategy 2	97.38	67.58	1.42	2.22	43.38	− 45.24	5.06	2.33
	Strategy 3	46.79	43.93	1.04	2.28	19.99	− 27.00	3.60	− 1.07
	Strategy Con. 2	55.81	57.70	0.95	1.14	47.77	− 40.64	2.92	2.55
knn	Strategy Con. 3	25.13	35.31	0.68	1.29	18.48	− 21.33	1.94	− 0.18
	Strategy 1	24.54	39.37	0.59	0.93	24.91	− 24.66	2.75	− 2.73
	Strategy 2	57.15	69.27	0.81	1.11	50.27	− 51.85	3.25	1.77
	Strategy 3	30.80	40.39	0.73	1.33	22.23	− 22.21	2.81	− 1.66
	Strategy Con. 2	52.90	60.55	0.85	0.98	52.87	− 41.80	3.19	1.24
rpart	Strategy Con. 3	23.07	35.30	0.62	0.96	22.73	− 20.26	2.19	− 1.34
	Strategy 1	27.43	42.12	0.62	0.84	30.99	− 23.56	3.06	− 3.08
	Strategy 2	64.38	69.00	0.91	1.35	46.86	− 39.61	3.88	1.06
	Strategy 3	35.08	42.91	0.79	1.33	25.34	− 20.24	3.23	− 2.11
	Strategy Con. 2	69.13	67.77	1.00	1.46	46.53	− 34.00	4.15	0.98
Mean (k = 1)	Strategy Con. 3	34.79	42.91	0.78	1.32	25.46	− 20.69	3.22	− 2.12
	Strategy 1	20.40	29.10	0.66	0.84	22.64	− 21.92	1.94	− 1.15
	Strategy 2	47.06	69.63	0.66	0.82	55.68	− 66.05	1.64	4.92
	Strategy 3	25.88	31.39	0.78	1.33	18.46	− 28.19	1.78	0.48
	Strategy Con. 2	40.42	46.25	0.85	0.82	47.61	− 34.65	1.77	3.33
Mean (k = 2)	Strategy Con. 3	5.66	16.32	0.27	0.26	16.86	− 13.12	0.16	0.91
	Strategy 1	39.98	41.60	0.93	1.86	20.85	− 22.66	3.61	− 2.32
	Strategy 2	97.45	67.57	1.42	2.22	43.24	− 46.93	4.98	2.59
	Strategy 3	47.32	42.82	1.08	2.47	18.62	− 25.88	3.58	− 0.89
	Strategy Con. 2	43.58	53.68	0.79	0.83	50.78	− 52.81	1.89	3.55
	Strategy Con. 3	9.75	23.82	0.36	0.43	19.55	− 28.54	0.60	0.55

(Continued).

Table 7. Continued.

Method	Strategy	Annual. Return	Annual. Risk	Sharpe Ratio	Sortino Ratio	Downside Risk	Average Drawdown	Alpha (Equity & Cash)	Beta (Equity & Cash)
Median (k = 1)	Strategy 1	21.81	33.42	0.61	0.96	21.38	−19.55	2.10	−1.35
	Strategy 2	50.46	69.51	0.71	1.00	49.01	−62.82	1.97	4.52
	Strategy 3	27.24	35.50	0.73	1.46	17.73	−27.89	1.92	0.33
	Strategy Con. 2	41.31	48.97	0.82	0.69	58.08	−54.84	1.74	3.59
	Strategy Con. 3	3.14	15.95	0.12	0.09	21.22	−31.49	−0.04	0.91
Median (k = 2)	Strategy 1	35.69	43.29	0.80	1.32	26.17	−27.78	3.39	−2.46
	Strategy 2	85.85	68.10	1.24	1.92	43.99	−51.67	4.54	2.30
	Strategy 3	42.05	44.49	0.92	1.80	22.71	−30.22	3.32	−1.08
	Strategy Con. 2	45.20	54.25	0.81	0.89	49.61	−54.03	2.19	2.93
	Strategy Con. 3	14.10	27.08	0.47	0.64	20.00	−34.45	1.07	0.09
BMA	Strategy 1	27.73	40.93	0.65	0.92	28.86	−19.45	2.85	−2.38
	Strategy 2	65.15	68.97	0.93	1.36	47.11	−38.07	3.45	2.47
	Strategy 3	34.41	42.07	0.79	1.39	23.76	−18.26	2.87	−1.14
	Strategy Con. 2	71.71	58.76	1.20	1.57	44.85	−24.74	4.07	1.64
	Strategy Con. 3	30.93	38.12	0.78	1.38	21.49	−10.23	2.57	−0.90
LDA	Strategy 1	40.94	43.02	0.92	1.78	22.30	−21.74	3.72	−2.48
	Strategy 2	100.09	67.45	1.47	2.26	43.67	−43.35	5.20	2.26
	Strategy 3	47.97	44.16	1.06	2.29	20.36	−26.18	3.69	−1.11
	Strategy Con. 2	60.61	58.46	1.01	1.26	47.17	−42.05	3.21	2.48
	Strategy Con. 3	28.32	36.65	0.74	1.49	18.16	−20.15	2.18	−0.25
RDA	Strategy 1	39.83	47.15	0.82	1.35	28.49	−19.67	3.92	−3.29
	Strategy 2	97.04	67.59	1.42	2.21	43.32	−32.39	5.60	0.64
	Strategy 3	46.90	47.85	0.95	1.80	25.32	−17.58	4.01	−2.29
	Strategy Con. 2	81.93	61.65	1.31	1.71	47.15	−29.47	4.70	1.25
	Strategy Con. 3	43.54	40.94	1.03	2.16	19.54	−12.15	3.54	−1.47
Naïve Bayes	Strategy 1	42.37	40.97	1.00	1.70	24.21	−22.38	3.94	−2.90
	Strategy 2	104.03	67.26	1.53	1.92	53.60	−29.48	5.65	1.43
	Strategy 3	49.89	41.76	1.16	2.26	21.53	−17.94	4.05	−1.86
	Strategy Con. 2	92.92	64.09	1.43	1.66	55.28	−33.64	5.28	1.05
	Strategy Con. 3	47.29	39.78	1.16	2.32	19.86	−15.46	3.88	−1.82
Bayes GLM	Strategy 1	38.63	43.00	0.87	1.70	21.99	−22.79	3.57	−2.46
	Strategy 2	93.77	67.74	1.37	2.15	43.07	−44.67	4.91	2.30
	Strategy 3	45.26	44.18	1.00	2.18	20.21	−26.77	3.52	−1.08
	Strategy Con. 2	61.68	58.41	1.03	1.27	47.51	−41.77	3.26	2.49
	Strategy Con. 3	28.98	36.59	0.76	1.51	18.33	−20.00	2.22	−0.25
Bagging	Strategy 1	14.53	41.76	0.32	0.40	33.31	−38.71	2.03	−2.71
	Strategy 2	33.26	70.04	0.46	0.61	52.80	−69.23	1.82	1.81
	Strategy 3	18.99	42.88	0.41	0.60	29.57	−39.07	1.99	−1.61
	Strategy Con. 2	37.76	65.66	0.56	0.70	52.28	−62.70	2.22	1.48
	Strategy Con. 3	18.72	40.31	0.43	0.60	29.09	−44.00	1.93	−1.48
Random Forest	Strategy 1	12.92	38.24	0.30	0.35	32.96	−38.06	1.82	−2.44
	Strategy 2	29.58	70.14	0.40	0.52	54.53	−62.21	1.41	2.35
	Strategy 3	17.69	39.45	0.42	0.58	28.45	−34.93	1.81	−1.35
	Strategy Con. 2	39.08	64.32	0.59	0.70	53.90	−52.70	2.26	1.61
	Strategy Con. 3	21.18	37.29	0.53	0.72	27.56	−29.73	2.07	−1.40
Ada. Boosting	Strategy 1	35.10	39.73	0.85	1.10	30.88	−19.34	3.47	−2.81
	Strategy 2	84.26	68.17	1.22	1.50	55.49	−38.40	4.69	1.62
	Strategy 3	43.25	40.68	1.03	1.64	25.55	−18.42	3.59	−1.66
	Strategy Con. 2	85.00	63.48	1.32	1.45	57.61	−32.12	4.78	1.47
	Strategy Con. 3	42.93	36.92	1.13	2.01	20.76	−13.98	3.53	−1.53
Grad. Boosting	Strategy 1	19.62	39.62	0.46	0.57	31.94	−27.74	2.47	−2.92
	Strategy 2	45.19	69.69	0.63	0.81	54.27	−50.81	2.70	1.39
	Strategy 3	25.69	40.55	0.60	0.88	27.75	−24.61	2.57	−1.97
	Strategy Con. 2	48.98	60.83	0.78	0.88	54.01	−39.82	2.96	1.26
	Strategy Con. 3	22.83	34.73	0.62	0.90	24.08	−18.76	2.18	−1.37

model based forecasts rank as the top performing, as they produce the highest annualized returns (104.03% and 103.91%), the highest Sharpe ratio (1.53) and estimated alphas (5.65 and 5.34), as well as high Sortino values (1.92 and 2.33, respectively). Additional modeling approaches that rank among the most successful are the Linear Discriminant analysis (LDA) and the mean combination scheme of forecasts based on all

possible Logit regression models with  $k = 2$  predictors. The latter achieve relatively high annualized returns of 100.09% and 97.45%, Sharpe ratio of 1.47 and 1.42, estimated alphas of 5.20 and 4.98, and Sortino values of 2.26 and 2.22, respectively. These results indicate that the machine learning techniques clearly outperform the univariate and multiple stepwise Logit regression models.



Table 8. Machine Learning Techniques: Economic Performance Evaluation—Preselection with Elastic Net ( $\alpha = 0.75$ )

Method	Strategy	Annual. Return	Annual. Risk	Sharpe Ratio	Sortino Ratio	Downside Risk	Average Drawdown	Alpha (Equity & Cash)	Beta (Equity & Cash)
Ridge AUC	VIX Long	− 1.76	70.54	− 0.04	− 0.08	38.61	− 71.33	2.24	− 7.23
	Strategy 1	11.95	22.53	0.47	0.46	23.26	− 12.30	1.13	− 0.56
	Strategy 2	27.39	70.19	0.37	0.48	54.94	− 75.04	0.02	6.10
	Strategy 3	17.57	25.77	0.63	1.02	16.06	− 23.55	0.96	1.21
	Strategy Con. 2	19.89	44.07	0.42	0.34	55.29	− 45.74	0.49	3.13
Ridge Class	Strategy Con. 3	5.88	14.91	0.31	0.29	15.71	− 12.52	0.17	0.92
	Strategy 1	27.47	37.28	0.70	1.28	20.54	− 23.43	2.67	− 1.89
	Strategy 2	64.49	69.00	0.92	1.41	44.94	− 61.48	3.09	3.45
	Strategy 3	34.37	38.91	0.85	1.88	17.59	− 29.92	2.60	− 0.33
	Strategy Con. 2	37.83	48.54	0.75	0.83	44.25	− 35.50	1.70	3.05
Ridge Dev.	Strategy Con. 3	5.52	20.21	0.21	0.25	17.31	− 20.08	0.23	0.67
	Strategy 1	42.97	42.35	0.98	1.90	21.99	− 21.39	3.82	− 2.42
	Strategy 2	105.71	67.18	1.55	2.36	44.25	− 43.72	5.41	2.38
	Strategy 3	49.81	43.51	1.12	2.47	19.66	− 25.55	3.77	− 1.05
	Strategy Con. 2	50.91	55.59	0.89	0.99	50.10	− 43.39	2.43	3.22
LASSO AUC	Strategy Con. 3	18.73	27.85	0.63	1.02	17.09	− 20.82	1.35	0.28
	Strategy 1	36.42	43.65	0.81	1.34	26.27	− 26.73	3.43	− 2.44
	Strategy 2	87.80	68.01	1.27	1.97	43.87	− 49.34	4.62	2.35
	Strategy 3	41.90	44.89	0.91	1.78	22.80	− 32.35	3.31	− 1.05
	Strategy Con. 2	68.32	60.18	1.11	1.35	49.50	− 44.33	3.50	2.83
LASSO Class	Strategy Con. 3	29.76	36.11	0.79	1.66	17.18	− 21.02	2.25	− 0.17
	Strategy 1	27.80	39.31	0.67	0.97	27.38	− 19.73	2.73	− 2.01
	Strategy 2	65.32	68.96	0.93	1.31	49.01	− 53.57	3.22	3.20
	Strategy 3	33.14	40.82	0.78	1.39	22.94	− 25.65	2.60	− 0.55
	Strategy Con. 2	59.28	58.56	0.99	1.16	50.05	− 39.07	2.93	3.11
LASSO Dev.	Strategy Con. 3	21.95	32.90	0.63	1.20	17.24	− 18.04	1.63	0.11
	Strategy 1	37.69	43.76	0.83	1.39	26.27	− 26.73	3.53	− 2.51
	Strategy 2	91.22	67.86	1.33	2.04	44.19	− 49.57	4.82	2.21
	Strategy 3	43.77	44.93	0.95	1.85	22.97	− 30.37	3.45	− 1.15
	Strategy Con. 2	67.35	60.21	1.10	1.34	49.14	− 44.59	3.45	2.82
Elastic Net AUC	Strategy Con. 3	29.73	36.18	0.79	1.67	17.02	− 21.26	2.25	− 0.18
	Strategy 1	37.69	43.76	0.83	1.39	26.27	− 26.73	3.53	− 2.51
	Strategy 2	91.22	67.86	1.33	2.04	44.19	− 49.57	4.82	2.21
	Strategy 3	43.77	44.93	0.95	1.85	22.97	− 30.37	3.45	− 1.15
	Strategy Con. 2	71.96	59.41	1.19	1.46	48.46	− 44.07	3.72	2.74
Elastic Net Class	Strategy Con. 3	30.08	36.15	0.80	1.68	17.13	− 20.88	2.28	− 0.19
	Strategy 1	29.86	40.19	0.71	1.04	27.38	− 22.24	2.87	− 2.03
	Strategy 2	70.58	68.75	1.01	1.44	48.17	− 53.96	3.51	3.16
	Strategy 3	35.10	41.65	0.81	1.47	22.95	− 27.81	2.73	− 0.57
	Strategy Con. 2	55.09	53.61	1.00	1.10	48.93	− 29.66	2.75	2.96
Elastic Net Dev.	Strategy Con. 3	16.29	24.87	0.60	0.93	16.15	− 13.33	1.17	0.28
	Strategy 1	37.69	43.76	0.83	1.39	26.27	− 26.73	3.53	− 2.51
	Strategy 2	91.22	67.86	1.33	2.04	44.19	− 49.57	4.82	2.21
	Strategy 3	43.77	44.93	0.95	1.85	22.97	− 30.37	3.45	− 1.15
	Strategy Con. 2	62.96	59.59	1.04	1.26	49.14	− 43.59	3.20	2.88
LDA	Strategy Con. 3	26.47	34.97	0.72	1.48	17.02	− 21.29	2.02	− 0.12
	Strategy 1	38.57	43.64	0.85	1.40	26.66	− 26.65	3.58	− 2.50
	Strategy 2	93.58	67.75	1.36	2.07	44.49	− 49.43	4.92	2.22
	Strategy 3	44.80	44.81	0.97	1.89	23.08	− 30.29	3.51	− 1.14
	Strategy Con. 2	68.55	60.26	1.12	1.38	48.88	− 42.35	3.52	2.81
RDA	Strategy Con. 3	30.16	36.26	0.80	1.71	16.88	− 19.62	2.28	− 0.19
	Strategy 1	42.77	46.10	0.90	1.55	26.83	− 19.52	4.05	− 3.15
	Strategy 2	105.15	67.21	1.55	2.41	43.04	− 31.23	5.87	0.92
	Strategy 3	50.45	46.78	1.05	2.08	23.60	− 16.45	4.17	− 2.13
	Strategy Con. 2	93.46	61.55	1.50	1.97	46.88	− 27.83	5.17	1.45
Naïve Bayes	Strategy Con. 3	47.46	39.94	1.16	3.00	15.39	− 10.77	3.75	− 1.38
	Strategy 1	41.91	40.99	0.99	1.60	25.45	− 23.38	3.92	− 2.91
	Strategy 2	102.76	67.33	1.51	1.88	53.95	− 31.52	5.60	1.41
	Strategy 3	49.33	41.79	1.15	2.15	22.36	− 17.90	4.02	− 1.87
	Strategy Con. 2	93.43	62.99	1.46	1.68	54.77	− 35.62	5.27	1.15
	Strategy Con. 3	45.40	39.35	1.12	2.12	20.84	− 15.78	3.72	− 1.66

(Continued).

Table 8. Continued.

Method	Strategy	Annual. Return	Annual. Risk	Sharpe Ratio	Sortino Ratio	Downside Risk	Average Drawdown	Alpha (Equity & Cash)	Beta (Equity & Cash)
Bayes GLM	Strategy 1	37.69	43.76	0.83	1.39	26.27	−26.73	3.53	−2.51
	Strategy 2	91.22	67.86	1.33	2.04	44.19	−49.57	4.82	2.21
	Strategy 3	43.77	44.93	0.95	1.85	22.97	−30.37	3.45	−1.15
	Strategy Con. 2	69.07	60.25	1.13	1.39	48.88	−42.26	3.54	2.81
	Strategy Con. 3	30.35	36.26	0.80	1.72	16.88	−19.62	2.30	−0.19
Bagging	Strategy 1	13.52	40.28	0.30	0.40	30.69	−37.44	1.91	−2.57
	Strategy 2	30.94	70.10	0.42	0.57	52.47	−68.42	1.58	2.08
	Strategy 3	18.81	41.48	0.42	0.63	27.91	−36.11	1.91	−1.42
	Strategy Con. 2	34.30	65.50	0.50	0.64	51.67	−57.69	1.88	1.84
	Strategy Con. 3	16.97	38.02	0.41	0.60	26.30	−38.98	1.72	−1.24
Random Forest	Strategy 1	15.64	38.85	0.37	0.47	30.63	−33.16	2.01	−2.41
	Strategy 2	35.84	69.97	0.49	0.65	53.18	−59.06	1.79	2.41
	Strategy 3	20.75	40.07	0.49	0.72	26.94	−30.43	2.00	−1.27
	Strategy Con. 2	36.54	66.37	0.53	0.63	56.12	−58.25	1.92	2.14
	Strategy Con. 3	20.37	37.50	0.51	0.70	27.42	−31.39	1.95	−1.19
Ada. Boosting	Strategy 1	36.32	39.20	0.89	1.55	22.61	−20.59	3.54	−2.79
	Strategy 2	87.52	68.03	1.27	1.64	52.66	−35.50	4.83	1.65
	Strategy 3	44.43	40.17	1.07	2.12	20.32	−19.06	3.65	−1.63
	Strategy Con. 2	86.97	63.54	1.35	1.51	56.65	−36.77	4.76	1.78
	Strategy Con. 3	39.52	37.04	1.03	1.84	20.80	−15.70	3.29	−1.43
Grad. Boosting	Strategy 1	31.73	37.27	0.82	1.33	22.84	−14.44	3.12	−2.40
	Strategy 2	75.41	68.55	1.08	1.40	53.11	−33.46	3.99	2.43
	Strategy 3	39.52	38.47	0.99	1.92	19.92	−13.39	3.19	−1.14
	Strategy Con. 2	57.49	63.31	0.89	1.00	56.09	−40.45	3.24	1.88
	Strategy Con. 3	31.14	35.72	0.84	1.50	19.92	−13.40	2.66	−1.14
knn	Strategy 1	22.27	39.14	0.54	0.86	24.29	−22.63	2.60	−2.75
	Strategy 2	51.58	69.47	0.72	1.01	49.78	−62.07	2.96	1.72
	Strategy 3	29.03	40.17	0.69	1.25	22.13	−21.21	2.70	−1.68
	Strategy Con. 2	48.42	61.42	0.77	0.93	50.96	−49.22	2.89	1.39
	Strategy Con. 3	20.93	34.90	0.56	0.89	21.99	−19.84	2.03	−1.33
rpart	Strategy 1	35.69	41.74	0.82	1.49	23.03	−20.80	3.57	−3.00
	Strategy 2	85.84	68.10	1.24	1.89	44.83	−34.18	4.90	1.22
	Strategy 3	45.48	42.44	1.04	2.21	19.98	−16.82	3.82	−1.97
	Strategy Con. 2	89.54	66.83	1.32	1.99	44.31	−27.37	5.10	1.14
	Strategy Con. 3	44.31	42.45	1.01	2.15	20.06	−19.08	3.76	−1.97
Mean (k = 1)	Strategy 1	21.47	29.86	0.68	1.02	19.89	−20.14	2.07	−1.32
	Strategy 2	49.65	69.54	0.70	0.88	54.84	−61.90	1.90	4.58
	Strategy 3	27.31	31.98	0.81	1.51	17.26	−22.41	1.95	0.24
	Strategy Con. 2	42.51	48.49	0.85	0.85	48.52	−33.88	1.84	3.49
	Strategy Con. 3	6.07	16.60	0.29	0.29	16.61	−17.56	0.19	0.90
Mean (k = 2)	Strategy 1	40.11	41.66	0.93	1.89	20.55	−22.48	3.61	−2.31
	Strategy 2	97.80	67.56	1.43	2.23	43.28	−47.82	4.99	2.61
	Strategy 3	46.97	42.90	1.07	2.47	18.48	−26.07	3.56	−0.89
	Strategy Con. 2	43.44	55.60	0.76	0.83	50.84	−51.76	1.90	3.49
	Strategy Con. 3	11.18	25.88	0.38	0.51	19.42	−31.34	0.74	0.44
Median (k = 1)	Strategy 1	21.48	34.55	0.58	1.00	20.31	−17.46	2.16	−1.58
	Strategy 2	49.66	69.54	0.70	1.01	48.10	−58.57	2.07	4.06
	Strategy 3	27.01	36.46	0.71	1.45	17.70	−22.53	2.00	0.02
	Strategy Con. 2	45.36	51.39	0.86	0.78	56.75	−53.54	2.01	3.50
	Strategy Con. 3	6.65	19.63	0.27	0.27	20.05	−26.39	0.29	0.74
Median (k = 2)	Strategy 1	41.23	41.75	0.96	1.98	20.19	−21.92	3.69	−2.33
	Strategy 2	100.89	67.41	1.48	2.28	43.76	−49.07	5.14	2.56
	Strategy 3	47.56	42.95	1.08	2.53	18.30	−27.70	3.61	−0.97
	Strategy Con. 2	43.48	55.02	0.77	0.83	50.78	−51.32	2.00	3.19
	Strategy Con. 3	12.40	25.26	0.44	0.57	19.65	−27.67	0.90	0.24
BMA	Strategy 1	27.73	40.93	0.65	0.92	28.86	−19.45	2.85	−2.38
	Strategy 2	65.15	68.97	0.93	1.36	47.11	−38.07	3.45	2.47
	Strategy 3	34.41	42.07	0.79	1.39	23.76	−18.26	2.87	−1.14
	Strategy Con. 2	64.81	60.30	1.05	1.36	46.57	−29.74	3.63	1.90
	Strategy Con. 3	29.92	38.18	0.75	1.33	21.47	−10.64	2.49	−0.86

Table 8 presents the evaluation measures for the machine learning techniques ( $\alpha = 0.75$  in the Elastic Net variable selection algorithm) and confirms the superior economic performance of these techniques. Once again, the Ridge Deviance, the Regularized Discriminant Analysis (RDA), the Naive Bayes based forecasts, as well as the median forecast combination scheme with a subset of  $k = 2$  predictors over the competing modeling approaches are the best performers.

In conclusion, the findings support the implementation of machine learning techniques in implied volatility prediction, as there is strong evidence regarding their superior accuracy for directional VIX forecasts that is also reflected in higher risk adjusted returns when real-time investment strategies are implemented.

## 5. Conclusion

The primary target of this study was to explore whether the application of innovative modeling approaches can contribute to the literature on the predictability of implied volatility indices. To this end, numerous alternative modeling approaches (Logit, Stepwise Logit, Combination of forecasts, and machine learning techniques) and model specifications have been employed to generate directional forecasts of the CBOE VIX index. The predictive accuracy of the generated out-of-sample forecasts was assessed both in a statistical and an economic setting. The economic significance was evaluated using various investment strategies based on the VIX index and the S&P 500 index.

The main conclusion is that the application of certain machine learning techniques in implied volatility prediction is strongly encouraged, as it significantly improves the accuracy of out-of-sample forecasts, based on numerous statistical evaluation measures over more mainstream econometric methods. Moreover, the use of penalized likelihood methods, such as the Elastic Net regularization technique, is strongly advised, as it can be used to narrow down the number and identify the important predictors.

As anticipated, there is a certain degree of divergence in the predictive accuracy of the underlying machine learning models. Model accuracy is not consistent across all models, some techniques, however, exhibit consistent superior relative performance (Naive Bayes, Ridge Deviance, Adaptive Boosting, Discriminant Analysis).

The analysis employs an elaborate set of predictive variables containing 31 macroeconomic and financial market related indicators, the majority of which have been widely followed and used in the existing literature on the predictability of asset returns. It also introduces a number of less studied indicators in the area of implied volatility predictability.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## References

- Aboura, S., International transmission of volatility: A study on the volatility indices VX1, VDAX and VIX, 2003. Available at SSRN: <https://ssrn.com/abstract=514282>.
- Audrino, F., Sigrist, F. and Ballinari, D., The impact of sentiment and attention measures on stock market volatility. *Int. J. Forecast.*, 2020, **36**, 334–357.
- Ahoniemi, K., Modeling and forecasting implied volatility: An econometric analysis of the VIX index. Working paper, Helsinki School of Economics, 2006.
- Akaike, H., Information theory and an extension of the maximum likelihood principle. In *Proceedings of Second International Symposium on Information Theory*, edited by B. N. Petrox and F. Caski, pp. 67–281, 1973 (Akademiai Kiado: Budapest).
- Bauer, E. and Kohavi, R., An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Mach. Learn.*, 1999, **36**, 105–139.
- Biau, G., Devroye, L. and Lugosi, G., Consistency of random forests and other averaging classifiers. *J. Mach. Learn. Res.*, 2008, **9**, 2015–2033.
- Breiman, L., Bagging predictors. *Mach. Learn.*, 1996, **24**(2), 123–140.
- Breiman, L., Arcing classifiers. *Ann. Stat.*, 1998, **26**(3), 801–849.
- Breiman, L., Random forests. *Mach. Learn.*, 2001, **45**, 5–32.
- Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A., *Classification and Regression Trees*, 1984 (CRC Press: Boca Raton).
- Brooks, C. and Oozer, M.C., Modeling the implied volatility of options on long gilt futures. *J. Bus. Financ. Account.*, 2002, **29**(1–2), 111–137.
- Buhlmann, P. and Yu, B., Analyzing bagging. *Ann. Stat.*, 2002, **30**(4), 927–961.
- Clements, A. and Fuller, J., Forecasting increases in the VIX: A time-varying long volatility hedge for equities, NCER Working Paper Series 88, National Centre for Econometric Research, 2012.
- Dietterich, T.G., An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learn.*, 2000, **40**(2), 139–157.
- Domingos, P. and Pazzani, M., Beyond independence: Conditions for the optimality of the simple Bayesian classifier. *Mach. Learn.*, 1997, **29**, 103–130.
- Dumas, B., Fleming, J. and Whaley, R.E., Implied volatility functions: Empirical tests. *J. Financ.*, 1998, **53**, 2059–2016.
- Efron, B. and Hastie, T., *Computer Age Statistical Inference, Algorithms, Evidence, and Data Science*, 2016 (Cambridge University Press: Cambridge).
- Elliott, G., Gargano, A. and Timmermann, A., Complete subset regressions. *J. Econom.*, 2013, **177**(2), 357–373.
- Fernandes, M., Medeiros, M.C. and Scharth, M., Modeling and predicting the CBOE market volatility index. *J. Bank. Financ.*, 2014, **40**, 1–10.
- Freund, Y. and Schapire, R.E., A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 1997, **55**(1), 119–139.
- Friedman, J.H., Regularized discriminant analysis. *J. Am. Stat. Assoc.*, 1989, **84**(405), 165–175.
- Friedman, J.H., Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, 2001, **29**(5), 1189–1232.
- Friedman, J.H., Stochastic gradient boosting. *Comput. Stat. Data Anal.*, 2002, **38**(4), 367–378.
- Friedman, J.H., Hastie, T. and Tibshirani, R., Additive logistic regression: A statistical view of boosting. *Ann. Stat.*, 2000, **28**(2), 337–407.
- Gelman, A., Jakulin, A., Pittau, M.G. and Su, Y., A weekly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.*, 2009, **2**(4), 1360–1383.
- Gemmell, G. and Kamiyama, N., International Transmission of Option Volatility and Skewness: When you're smiling, does the Whole World Smile? Working Paper, City University Cass Business School, 2000.

- Goncalves, S. and Guidolin, M., Predictable dynamics in the S&P 500 index options implied volatility surface. *J. Bus.*, 2006, **79**(3), 1591–1635.
- Goyal, A. and Saretto, A., Option returns and the cross-sectional predictability of implied volatility. Working Paper, Purdue University, 2006.
- Guo, D., Dynamic volatility trading strategies in the currency option market. *Rev. Deriv. Res.*, 2000, **4**(2), 133–154.
- Harvey, C.R. and Whaley, R.E., Market volatility prediction and the efficiency of the S&P 100 index option market. *J. Financ. Econ.*, 1992, **31**, 43–73.
- Hirsa, A., Osterrieder, J., Hadji Misheva, B., Cao, W., Fu, Y., Sun, H. and Wong, K.W., The VIX index under scrutiny of machine learning techniques and neural networks. Working paper, 2021.
- Hoerl, A.E. and Kennard, R.W., Ridge regression: Applications to nonorthogonal problems. *Technometrics*, 1970, **12**(1), 69–82.
- Hossin, M. and Sulaiman, M.N., A review on evaluation metrics for data classification evaluations. *Int. J. Data Mining Knowl. Manag. Process (IJDKP)*, 2015, **5**(2), 1–11.
- Hosker, J., Djurdjevic, S., Nguyen, H. and Slater, R., Improving VIX futures forecasts using machine learning methods. *SMU Data Sci. Rev.*, 2018, **1**(4), Article 6.
- James, G., Witten, D., Hastie, T. and Tibshirani, R., *An Introduction to Statistical Learning*, 2013 (Springer: New York).
- Konstantinidi, E., Skiadopoulos, G. and Tzagkaraki, E., Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *J. Bank. Financ.*, 2008, **32**(11), 2401–2411.
- Meligkotsidou, L., Panopoulou, E., Vrontos, I.D. and Vrontos, S.D., Quantile forecast combinations in realised volatility prediction. *J. Oper. Res. Soc.*, 2019, **70**(10), 1720–1733.
- Meligkotsidou, L., Panopoulou, E., Vrontos, I.D. and Vrontos, S.D., Out-of-sample equity premium prediction: A complete subset quantile regression approach. *Eur. J. Finance*, 2021, **27**(1–2), 110–135.
- Pesaran, M.H. and Timmermann, A., A simple nonparametric test of predictive performance. *J. Bus. Econ. Stat.*, 1992, **10**(4), 461–465.
- Schwarz, G., Estimating the dimension of a model. *Ann. Stat.*, 1978, **6**(2), 461–464.
- Sokolova, M. and Lapalme, G., A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.*, 2009, **45**, 427–437.
- Tharwat, A., Classification assessment methods. *Appl. Comput. Inform.*, 2020, **17**(1), 168–192.
- Therneau, T.M. and Atkinson, E.J., *An Introduction to Recursive Partitioning Using the RPART Routines*, 2019 (Mayo Foundation: Rochester).
- Tibshirani, R., Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc., Ser. B*, 1996, 267–288.
- Vortelinos, D., Forecasting realized volatility: HAR against principal components combining, neural networks and GARCH. *Res. Int. Bus. Financ.*, 2017, **39**, 824–839.
- Zou, H. and Hastie, T., Regularization and variable selection via the elastic net. *J. R. Stat. Soc., Ser. B*, 2005, **67**(2), 301–320.



## Appendix

Table A1. Set of Predictors – Overview.

Code	Predictive Variable	Unit	Transformation	Source
1	VIX Index Level – End of period	(%)	–	CBOE
2	SMB Factor Monthly Portfolio Returns – Small Minus Big – Size Factor	(%)	–	Fama – French
3	HML Factor Monthly Portfolio Returns – High Minus Low -Value Factor	(%)	–	Fama – French
4	MKT Factor Portfolio Returns – Equity Market Factor	(%)	–	Fama – French
5	MOM Factor Portfolio Returns – Medium Price Momentum Factor	(%)	–	Fama – French
6	STR Factor Portfolio Returns – Short Term Reversal Factor	(%)	–	Fama – French
7	RVAR Factor Portfolio Returns – Residual Variance Factor	(%)	–	Fama – French
8	BAB Factor Portfolio Returns – Betting Against Beta Factor	(%)	–	AQR
9	QMJ Factor Portfolio Returns – Quality Minus Junk Factor	(%)	–	AQR
10	Term spread – 10-year Treasury yield -3-month Tbill rate	(%)	–	FRED Database
11	Credit Spread – Moodys BAA Yield over 10-year Treasury Yield	(%)	–	FRED Database
12	10-year Treasury Yield	(%)	–	FRED Database
13	3-month T-bill Rate	(%)	–	FRED Database
14	Oil Price Index – WTI Crude Oil	(%)	month-on-month % change	FRED Database
15	Dollar Index	(%)	month-on-month % change	FRED Database
16	TED Spread	(%)	–	FRED Database
17	Skew Index	(%)	–	CBOE
18	Chicago Fed National Financial Conditions Leverage SubIndex	Z-score	Normalized	FRED Database
19	TSMOM Factor Portfolio Returns – Diversified Asset Medium-Term Price Momentum	(%)	–	AQR
20	MOM CM Factor Portfolio Returns – Commodities Medium-Term Price Momentum	(%)	–	AQR
21	MOM EQ Factor Portfolio Returns – Equity Market Medium-Term Price Momentum	(%)	–	AQR
22	MOM FI Factor Portfolio Returns – Fixed Income Market Medium-Term Price Momentum	(%)	–	AQR
23	MOM FX Factor Portfolio Returns – Foreign Exchange Market Medium-Term Price Momentum	(%)	–	AQR
24	HMLD Factor Portfolio Returns – High Minus Low Devil Factor	(%)	–	AQR
25	Economic Policy Uncertainty Index (ECU)	Index Level	–	Economic Policy Uncertainty
26	Macroeconomic Volatility (MacVol) – Industrial Production Growth	(%)	Annualized 12-month rolling SDEV	FRED Database
27	Bullish Sentiment – Average	(%)	Monthly Average	AAII
28	Neutral Sentiment – Average	(%)	Monthly Average	AAII
29	Bearish Sentiment – Average	(%)	Monthly Average	AAII
30	Bull-Bear Spread – Average	(%)	Monthly Average	AAII
31	Bull – Bear Ratio – Average	(%)	Monthly Average	AAII
32	8-week Moving Average Bullish Sentiment	(%)	–	AAII