A Black-Litterman Portfolio Selection Model with Investor Opinions Generating from Machine Learning Algorithms

Liangyu Min, Jiawei Dong, Dewen Liu, Xiangxi Kong

***Abstract*—Generation error is one of the shortcomings of the classical mean-variance portfolio selection model which would result in unstable performance in out-of-sample data sets. Machine learning provides several potential solutions for reducing the generalization error, such as adding penalty terms and random sampling. Due to some significant advantages of machine learning algorithms, we implement Black-Litterman (BL for short) series portfolio models integrated with quantitative opinions generating from machine learning algorithms in this paper. And, the fundamental factors in the Fama-French model form the basis of quantitative opinions. Considering the non-linearity among fundamental factors, we construct high-order terms and cross terms of basic features by the approach of dimension-increasing transformation. A multi-period portfolio strategy is designed in our work for the timeliness of quantitative opinions, the experimental results reveal that the optimal BL model with investor opinion generating from Random Forest gained over 20% average annual return and a Sharpe Ratio of 1.25. By comparison, S&P 500 index gained about 14.94% annual return and a Sharpe Ratio of 1.00, the 1/N strategy gained about 15.92% annual return and the Sharpe Ratio of 0.99. Moreover, the BL series models are more diversified and robust, about 30% to 60% of assets are selected to construct the portfolio. Even when the transaction cost is taken into account, our proposed models still obtained higher cumulative returns than S&P 500 if the transaction cost is lower than 30%%.**

***Index Terms*—Portfolio selection, Machine learning, Black-Litterman, Factor model, Non-linearity.**

I. Introduction

THE parameters in the mean-variance model proposed by Markowitz (1952) [1], $\hat{\mu}$ for expected return and $\hat{\Sigma}$ for risk, are estimated from historical data, and the portfolio weights can be calculated by the convex quadratic programming problem with the two parameters. This seminal theoretical model marked the inception of modern portfolio theory. However, scholars have found that several obvious shortcomings exist in the mean-variance model, such as high sensitivity for inputted parameters and unstable performance in out-of-sample data sets [2], [3], [4]. An enormous amount of researches have been done to overcome these drawbacks. Goldfarb & Iyengar (2003) [5] built the robust portfolio selection model considering the worst case of estimated parameters. The proposed worst-case model can be reformulated into a second-order cone programming (SOCP) for solutions. And, the worst-case portfolio model provides hard guarantees on the performance. Based on Goldfarb & Iyengar's model, Zhu & Fukushiama (2009) further proposed the worst-case CVaR robust portfolio selection model [6] dealing with the uncertainty of the probability distributions. However, the issue of conservatism influences the performance of robust portfolio models in numerical tests, especially in the scenario of high noisy data. From the point of Bayesian analysis, Black & Litterman (1990, 1992) proposed a portfolio selection model incorporating subjective views [7], [8], which overcomes the problems of the sensitivity of parameters, highly-concentrated portfolio, and estimation error maximization to some extent [9], [10]. The main contribution of the Black and Litterman model is that predictive opinions are integrated into the classical mean-variance framework. Practically, the performance of the BL model depends on the accuracy and predictability of quantitative opinions considering the complexity and dynamic nature of the financial market.

According to the review of related literature, a variety of intelligent forecasting methodologies have been employed for generating investor views. Didenko & Demicheva (2013) used ensemble learning algorithms to generate investor opinions in the Meucci portfolio model[11]. In their work, the random forest ensemble learning algorithm is the only predictor for producing investor views, which may be not comprehensive enough. Asad (2015) [12] devised an ensemble system for stock prediction, support vector machine (SVM), relevance vector machines, random forest, and k-nearest neighbor are employed in their framework. Although rather logical and comprehensive the ensemble system is, the binary classification of labels can not provide enough descriptions for the financial market. With the development of computing power and artificial intelligence, several hybrid models are developed and employed in financial forecasting. Kim et al (2018) [13] devised a hybrid model, integrating Long short-term memory networks (LSTM) with GARCH-type models for volatility forecasting. The empirical analysis illustrated that their hybrid model enhances the prediction performance in stock market volatility. Krisjanpoller & Minutolo

Manuscript received December 26, 2020; revised April 15, 2021. This work is supported by the Graduate Innovation Fund in SUFE (CXJJ-2018-401, CXJJ-2019-325, CXJJ-2019-342) and the Fundamental Research Funds for the Central Universities (No. 2020110552).

Liangyu Min is a PhD candidate in the School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China (Corresponding author, e-mail: minux@163.sufe.edu.cn).

Jiawei Dong is a PhD candidate in the School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: hightempler@126.com).

Dewen Liu is a PhD candidate in the College of Business, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: dewen0731@126.com).

Xiangxi Kong is a PhD candidate in the College of Business, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: kongxiangxikxx@163.com).

(2018) [14] designed a hybrid model integrating GARCH, ANN, and PCA for volatility forecasting in the market of cryptocurrencies. Numerical tests on the price of bitcoin revealed the feasibility of this hybrid model. Kara et al (2019) constructed a BL portfolio model with investor views generating from a hybrid model [15]. In their model, the investor views are given by $\varepsilon-$SVM with inputted indicators predicted by GARCH. However, the significant advantages of ensemble models are not exerted in their portfolio selection model.

In this paper, for the purpose of overcoming the conservatism in the robust portfolio model as well as the drawbacks mentioned above in the mean-variance model, we are committed to proposing a Black-Litterman portfolio selection model with investor opinions generating from multiple machine learning algorithms. First of all, the factors developed by Fama & French (1993, 2012) [16], [17] are the fundamental predictors, which proved to be able to explain more than 90% of the market return rate. That is different from similar work used technical indicators [18]. According to the conclusions of Gu et al (2018) [19], the non-linearity of factors and the interaction between predictors should not be neglected, hence we construct high-order terms and cross terms at the stage of preprocessing. Secondly, several machine learning algorithms including Logistic Regression (LR), SVM, Random Forest (RF), XGBoost (XG), and Multi-layer perceptron (MLP) are utilized for generating investor opinions. Considering the issue of overfitting in machine learning algorithms, we apply some techniques such as L_1 regularization (Lasso), L_2 regularization (Ridge), elastic nets, and principle components analysis (PCA), et al for curbing overfitting. Finally, we construct a multi-period Black-Litterman portfolio selection model with regularly updated investor opinions. For comparison, market index (S&P 500 index) and benchmark portfolios including 1/N strategy [20], mean-variance (MV) portfolio, and worst-case robust portfolio are tested in the same data set. Moreover, to further simulate the real trading conditions, the transaction cost is considered.

The main contribution of our work is threefold and could be concluded as follows:

1) Slightly different from the classical Black-Litterman portfolio model, investor views derived from multiple machine learning algorithms are integrated into the portfolio framework. It can be found in a voluminous literature that learning algorithms, as well as shallow neural networks, are feasible on forecasting future directions of the various types of assets [21], [22], [11], [23], [24], [25], [26], [13], [14], [27], [28], [29], [30], which provides the empirical basis for our work.
2) A dynamic portfolio scheme is designed in virtue of the validity and timeliness of investor opinions. The data set containing 30 industry from Kenneth R. French is used for evaluating our model by comparing with other portfolio frameworks such as naive diversification (1/N) strategy, mean-variance portfolio, and worst-case robust model. Moreover, multiple risk-adjusted indicators are exhibited for further analysis.
3) We further consider the diversification of the proposed model, the results of weight analysis are presented in section 4.E. Also, we investigate the diversification of portfolios from two angles: the number of assets selected in a portfolio and the maximum weight allocated in a portfolio. Also, the transaction cost is taken into account as well to simulate the real trading conditions. The corresponding results of sensitivity analysis are revealed in section 4.F.

The paper is organized as follows: Section 2 introduces the Black-Litterman portfolio selection model and the worst-case robust portfolio selection model. Section 3 reveals the machine learning algorithms and the mechanism of forming and applying for investor views in BL series models. The numerical tests and analysis of results are presented in Section 4. Finally, the conclusions and discussions are in Section 5.

II. Portfolio selection models

In this section, we introduce the classical Black-Litterman model and the worst-case robust model. Both portfolio models overcome the drawbacks of the Markowitz model to some extent.

A. Black-Litterman portfolio model

Assuming that there are N risky assets could be invested in the market, the $N\times 1$ vector r represents the return rate of assets following a multivariate normal distribution: $\mathbf{r}\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$. In the condition of market equilibrium, every investor holds the market equilibrium portfolio ω_{eq}, if the risk aversion coefficient is λ, then the market equilibrium return rate[1] $\boldsymbol{\Pi}=\lambda\boldsymbol{\Sigma}\omega_{eq}$. The expected return rate of assets μ could be split into two parts, the first part is $\boldsymbol{\Pi}$ mentioned above, and the second part is residual $\boldsymbol{\epsilon_{\mu}}\sim\mathcal{N}(0,\tau\boldsymbol{\Sigma})$, τ represents the uncertainty of the estimation Σ from samples. According to the property of Gaussian normal distribution, we have $\boldsymbol{\mu}\sim\mathcal{N}(\boldsymbol{\Pi},\tau\boldsymbol{\Sigma})$.

If experts have K investor opinions about N assets in the market, then the dimension of the opinions matrix $\mathbf{P}$ is $K\times N$, where $\mathbf{P}^T=[p_1,p_2,\dots,p_K]$. And $\mathbf{Q}$ is a $K\times 1$ vector that represents the expected return rates, where $\mathbf{Q}^T=[q_1,q_2,\dots q_K]$. The relationship between vector $\mathbf{P}$ and $\mathbf{Q}$ satisfies $\mathbf{P}\boldsymbol{\mu}=\mathbf{Q}+\boldsymbol{\epsilon_{\nu}}$, where $\boldsymbol{\epsilon_{\nu}}\sim\mathcal{N}(0,\boldsymbol{\Omega})$ is the measure of the investor opinions uncertainty. Under the condition of i.i.d, the matrix $\boldsymbol{\Omega}$ is a $K\times K$ diagonal. Also, we could derive the statistical property based on Gaussian distribution as follows:

$$\mathbf{P}\boldsymbol{\mu}\sim\mathcal{N}(\mathbf{Q},\boldsymbol{\Omega})$$

In order to solve the BL model in the Bayesian framework, we assume that the residual $\boldsymbol{\epsilon_{\mu}}$ and $\boldsymbol{\epsilon_{\nu}}$ satisfy the condition of independent distribution as follows:

$$\begin{pmatrix}\boldsymbol{\epsilon_{\mu}}\\ \boldsymbol{\epsilon_{\nu}}\end{pmatrix}\sim\mathcal{N}\left(0,\begin{bmatrix}\tau\boldsymbol{\Sigma} & 0\\ 0 & \boldsymbol{\Omega}\end{bmatrix}\right)$$

[1]In the condition of market equilibrium, each investor has the identical optimization problem $\max\omega^T\boldsymbol{\Pi}-\lambda\omega^T\boldsymbol{\Sigma}\omega$, the optimized $\boldsymbol{\Pi}^*$ could be reached through the first-order condition.

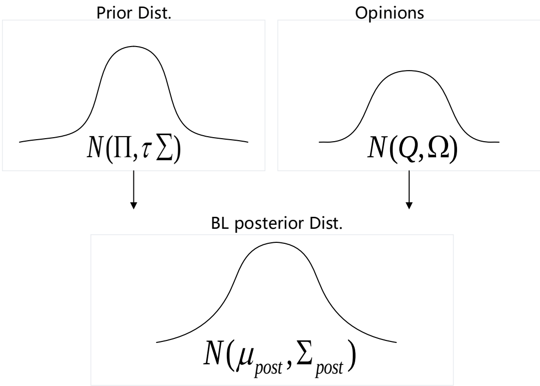


Fig. 1. The flow chart of Black-Litterman model

The posterior distribution of returns is derived according to the Bayesian formula:

$$f(\boldsymbol{\mu}|\mathbf{Q}) = \frac{f(\mathbf{Q},\boldsymbol{\mu})}{f(\mathbf{Q})} = \frac{f(\mathbf{Q}|\boldsymbol{\mu})f(\boldsymbol{\mu})}{\int f(\mathbf{Q}|\boldsymbol{\mu})f(\boldsymbol{\mu})d\boldsymbol{\mu}}$$

where the second equation could be interpreted from the perspective of machine learning: $f(\boldsymbol{\mu})$ is the prior distribution of return-on-assets, $f(\mathbf{Q}|\boldsymbol{\mu})$ is the likelihood estimation, and $\int f(\mathbf{Q}|\boldsymbol{\mu})f(\boldsymbol{\mu})d\boldsymbol{\mu}$ is the normalized term. We can infer the relationship among variables according to the analysis above:

$$\boldsymbol{\mu} - \boldsymbol{\Pi} = \boldsymbol{\epsilon_\mu} \sim \mathcal{N}(0,\tau\boldsymbol{\Sigma})$$
$$\mathbf{P}\boldsymbol{\mu} - \mathbf{Q} = \boldsymbol{\epsilon_\nu} \sim \mathcal{N}(0,\boldsymbol{\Omega})$$

Overall, the formula of posterior distribution about μ is as follows, and the flow chart of Black-Litterman model is shown in Fig. 1.

$$\boldsymbol{\mu}|\mathbf{Q} \sim \mathcal{N}([(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\Pi} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{Q}], [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}]^{-1}) \tag{1}$$

In the Black-Litterman framework, subjective opinions are integrated the into portfolio model, which would affect the parameters $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ in the classical mean-variance model. It can be found that the actual performance of the Black-Litterman model depends on the accuracy and unbiasedness of investor opinions. However, subjective views derived from human-beings may hard to be absolutely objective and fair. That is the main reason for substituting subjective views with opinions generating from machine learning algorithms.

B. Worst case robust portfolio model

On the basis of robust optimization [31], [32] and factor model [33], [16], Goldfarb et al constructed the robust portfolio model. The relationship between return rates and factors can be built up with a linear equation: $\mathbf{r} = \boldsymbol{\mu} + \mathbf{V}^T\mathbf{f} + \boldsymbol{\epsilon}$, where $\boldsymbol{\mu} \in \mathbb{R}^n$ is the vector representing the mean values; $\mathbf{V} \in \mathbb{R}^{m\times n}$ is the factor loading matrix; $\mathbf{f} \in \mathbb{R}^m$ is the returns vector of market-driving factors; $\boldsymbol{\epsilon} \sim \mathcal{N}(0,\boldsymbol{\Omega})$ is the vector of residuals and the variance-covariance matrix $\boldsymbol{\Omega} \succeq \mathbf{0}$.

The factor model assumes that the residual vector $\boldsymbol{\epsilon}$ and the factor return vector $\mathbf{f}$ are independent, the variance-covariance matrix of factors is symmetric and positive semi-definitive, that is, $\mathbf{F} \succeq \mathbf{0}$. Accordingly, $\mathbf{r} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V}^T\mathbf{F}\mathbf{V}+\boldsymbol{\Omega})$, where $\mathbf{V}^T\mathbf{F}\mathbf{V}+\boldsymbol{\Omega}$ is the approximation of the variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$ estimated from sample data. In the robust portfolio model, the uncertainty structure of parameters is set as follows:

$$S_\nu = \{\mathbf{V} : \mathbf{V} = \mathbf{V}_0 + \mathbf{W}, \||W_i\||_g \le \rho_i, i = 1, 2, \ldots, n\}$$
$$S_\omega = \{\boldsymbol{\Omega} : \boldsymbol{\Omega} = diag(\omega), \omega_i \in [\underline{\omega}_i, \overline{\omega}_i], i = 1, 2, \ldots, n\}$$
$$S_\mu = \{\boldsymbol{\mu} : \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\xi}, |\xi_i| \le \gamma_i, i = 1, 2, \ldots, n\}$$

Let $\boldsymbol{\Phi}$ represents the vector of portfolio weights and consider the worst-case parameters, we can construct the Min-Max portfolio model as follows:

$$\min \max_{\mathbf{V}\in S_v} \|\mathbf{V}\boldsymbol{\Phi}\|_f^2 + \boldsymbol{\Phi}^T\overline{\boldsymbol{\Omega}}\boldsymbol{\Phi}$$
$$s.t. \quad \begin{cases} \min\limits_{\boldsymbol{\mu}\in S_\mu} \boldsymbol{\mu}^T\boldsymbol{\Phi} \ge \alpha \\ \mathbf{1}^T\boldsymbol{\Phi} = 1 \end{cases} \tag{2}$$

where $\|\mathbf{x}\|_f = \sqrt{\mathbf{x}^T\mathbf{F}\mathbf{x}}$, α is the lower bound of target return.

Assuming that the set of parameters in the worst-case model is finite, then portfolio model (2) is convex quadratic programming, which can be reformulated into second-order cone programming (SOCP) by introducing some auxiliary variables, refer to [5] for more details.

Due to the only worst case is taken into account, the robust portfolio model is quite conservative. How to overcome the conservatism in a robust portfolio model has become an academic and industrial issue. For the purpose of illustrating the advantages of the proposed portfolio model, we employ the worst-case robust portfolio model as one of the benchmarks.

III. Machine learning algorithms and investor opinions incorporating

In this section, several machine learning algorithms and the mechanism for the synthesis of investor opinions generating from these artificial intelligence algorithms are introduced.

Generally, the quantitative view from investors is in a form of a numerical interval with a certain confidence. In order to build the relationship between predictive views and market tendency, a method of label discretization [34] is employed at the stage of preprocessing. Also, the confidence level of investor opinions is calibrated by a variance-covariance matrix Ω.

A. Logistic Regression

Logistic Regression is developed for predictive analysis, which is used to describe data and to explain the relationship between one dependent binary variable and several nominal, ordinal, or interval variables. More than binary classification, multi-class classification can be done through the one-vs-rest scheme or cross-entropy loss function.

Sigmoid is the function which could map the result of linear regression $\hat{y} = \boldsymbol{\theta}^T\mathbf{x}$ to $(0,1)$, define the probability of label y is as follows:

$$\mathbf{P}(y) = g(\hat{y}) = g(\boldsymbol{\theta}^T\mathbf{x}) = \frac{1}{1+e^{-\boldsymbol{\theta}^T\mathbf{x}}} = \frac{e^{\boldsymbol{\theta}^T\mathbf{x}}}{1+e^{-\boldsymbol{\theta}^T\mathbf{x}}}$$

Assuming that the range of label y is $\{1, 2, \ldots, N\}$, the equation of multi-classification could be derived by the one-vs-one scheme as follows:

$$\mathbf{P}(y = n|\boldsymbol{\theta}, \mathbf{x}) = \frac{e^{\boldsymbol{\theta}_n^T x}}{1 + \sum\limits_{n=1}^{N-1} e^{\boldsymbol{\theta}_n^T \mathbf{x}}}, \quad n = 1, 2, \ldots, N-1$$

$$\mathbf{P}(y = N|\boldsymbol{\theta}, \mathbf{x}) = \frac{1}{1 + \sum\limits_{n=1}^{N-1} e^{\boldsymbol{\theta}_n^T \mathbf{x}}}$$

Estimation of parameters in Logistic Regression could be solved resort to Maximum Likelihood Estimation (MLE), in which the logarithmic likelihood function could be efficiently solved by newton method, quasi-newton method, or gradient descent method.

B. Support Vector Machine

Support Vector Machine (SVM) is a kind of supervised learning algorithm [35] originating from statistical learning theory (SLT). In SLT, due to the scarcity of samples, the rule of Expected Risk Minimization is substituted with Empirical Risk Minimization (ERM). Nonetheless, if the complexity of the model is high while the number of samples is limited, the learning model would be overfitting under the rule of ERM [36]. To avoid such a dilemma, Structure Risk Minimization (SRM) is chosen as the objective function in SVM that means approaching the true risk by empirical risk and confidence interval. In the condition of a limited number of samples, SVM would learn the model with appropriate complexity[37], [38].

The principle could be illustrated through a binary classification demo. Assuming that the train set is $\{(x_i, y_i)\}, i = 1, 2, \ldots, m$ and $x_i \in \mathbb{R}^n, y_i \in \{0, 1\}$, the objective of SVM is to optimize the following constrained equation with regularizer:

$$\min_{\omega,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i$$

$$s.t. \quad \begin{cases} y_i(w^T x_i + b) + \xi_i \geq 1 \\ \xi_i \geq 0 \end{cases}$$

Kernel function $\Phi(\cdot)$ is introduced in the case of nonlinear separated data. The primal problem could be converted into the dual form with kernel function as follows:

$$\min_{\alpha} \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \Phi(x_i, x_j) - \sum_i \alpha_i$$

$$s.t. \quad \begin{cases} \sum\limits_{i=1}^{N} \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C \end{cases}$$

where the dual problem is convex quadratic programming if $\Phi(\cdot)$ is the positive definite kernel (such as Gaussian, Polynomial, and Sigmoid) by Mercer theorem.

C. Random Forest

Random Forest (RF) [39] is an ensemble algorithm for classification or regression, by constructing N base estimators then synthesizing the outputs in a manner of bagging. Assuming that there are M features, each base estimator would randomly sample at most $k \approx \sqrt{M}$ features for decision. Define information gain as the criterion for tree node splitting as follows:

$$Gain(D, k_j) = Entropy(D) - \sum_i \frac{|D_i|}{D} Entropy(D_i)$$

where the parent node is D, the children nodes after splitting are D_i, k_j represents the features to be split. In general, we select the decision tree as the base estimator and split tree nodes according to the rule of the largest information gain about which feature could bring. When the leaves reach the threshold of impurity, Random Forest then vote on the results calculated by N decision trees, the final output, $R(x)$, for classification satisfies the equation as follows:

$$R(x) = argmax \sum_{i=1}^{n} \mathbf{I}(r_i(x) = R(x))$$

where $r_i(x)$ is the result of base estimator i and $\mathbf{I}(\cdot)$ is the indicator function.

In Random Forest, the pruning method is used to overcome the issue that a single classification model (e.g. decision tree) is inclined to overfitting. To improve the generalization and reduce the conservatism of the Black-Litterman portfolio model, it is reasonable to adopt the opinions generating from Random Forest.

D. XGBoost

XGBoost (Extreme Gradient Boosting) [40] is another ensemble learning algorithm. Different from bagging in Random Forest, boosting [41] is used in XGBoost for synthesizing outputs. In XGBoost, base estimators do not make the decision independently and parallelly but try to minimize objective loss function $L(\theta)$ by the scheme of greedy. Similar to SVM, the two objectives of the precision of the model (bias), and the stability of the model (variance) are taken into account together.

$$L(\theta) = \sum_{i=1}^{m} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

the structure of $\Omega(f_k)$ can be illustrated in the following equation with L_1 and L_2 regularizers:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\alpha|w_j| + \frac{1}{2}\lambda\|w\|_2^2$$

where the $l(y_i, \hat{y}_i)$ is the deviation of predictive results, $\Omega(f_k)$ represent the complexity of k_{th} tree and T is the number of leaves. Due to the Classification and Regression Trees (CART) [42] is the base estimator of XGBoost, the complexity could be defined through two hyper-parameters: γ and T. Take the second-order form of Taylor expansion to approximate the objective function in the Kth iteration. Define g_i and h_i as the first-order gradient statistics and the second-order one, we could get the form of objective function without any constant term in iterations as follows:

$$\begin{cases} \hat{L}^{(K)} \approx \sum\limits_{i=1}^{m} [g_i f_K(x_i) + \frac{1}{2} h_i f_K^2(x_i)] + \Omega(f_K) \\ g_i = \frac{\partial l(y_i^K, \hat{y}_i^{(K-1)})}{\partial \hat{y}_i^{(K-1)}} \\ h_i = \frac{\partial^2 l(y_i^K, \hat{y}_i^{(K-1)})}{\partial (\hat{y}_i^{(K-1)})^2} \end{cases}$$

E. Multi-layer Perceptron Neural Network

Multi-layer perceptron (MLP) is a class of feed-forward neural networks with hidden layers. In MLP, the calculation results of layer i would be forward propagated to layer $i+1$ by the full connection between layers. The nonlinear activation function used in a neuron is $\sigma(\cdot)$, and the weight matrix is W, and the bias is b, the number of total layers is I, the input vector is x, we could get the output of layer $i+1$ as follows:

$$y^{i+1} = \sigma(z^{i+1}) = \sigma(W^{i+1}y^i + b^{i+1})$$

Define the loss function between true label y and predict label y^I is $L(W, b)$. The loss is propagated backward to each layer according to the back-propagation algorithm, as well as update the weight matrix W and the bias b based on the layer's contribution, $\delta^{j,i}$, to the loss function, $L(W, b)$, and the learning rate is η.

$$\begin{aligned}
L(W, b) &= \frac{1}{2}\|y^I - y\|_2^2 \\
\delta^{j,i} &= (W^{i+1})^T \delta^{j,i+1} \odot \delta'(z^{j,i}) \\
W^i &= W^i - \eta \sum_{j=1}^{m} \delta^{j,i} (y^{j,i-1})^T \\
b^l &= b^l - \eta \sum_{j=1}^{m} \delta^{j,i}
\end{aligned}$$

Several techniques are integrated into MLP neural networks for reducing generalization error, such as learning rate, batch, and early-stopping. Throughout the present research, neural networks have been widely used in many academic fields[43], [44], [45], among which the shallow neural network with about three hidden layers is suitable for fitting financial data with a low Signal Noise Ratio (SNR) [19].

F. Mechanism of incorporating investor opinions

The generation error of model, $E(f; D)$, is defined as follows:

$$E(f; D) = \underbrace{bias^2(x)}_{\textbf{Bias}} + \underbrace{var(x)}_{\textbf{Variance}} + \epsilon^2$$

where $bias^2(x)$ measures the difference between the predicted values and the true values of the model, which represents the precision of the model. $var(x)$ measures the difference between each predicted value and the average of predicted values, which represents the stability of the model. ϵ^2 is the idiosyncratic term that can not be interpreted by bias and variance.

Generally, the bias would be decreasing and the variance would be increasing as the complexity of the model increasing; the bias would be increasing and the variance would be decreasing as the complexity of the model decreasing. According to the VC theory [46], the complexity of the model is highly related to the number of features, and therefore we should select the features of higher explanatory for modeling.

In order to reduce the generalization error in the classical mean-variance model, these investor opinions generating from machine learning algorithms are incorporated into the Black-Litterman model through the matrix of opinions P and the vector of the expected returns Q. As a comparison, we also examine the performance of benchmark portfolios widely used in academic and industrial. The pseudo-code of BL series models is revealed in Algorithm 1 and the pseudo-code of backtesting is revealed in Algorithm 2.

Algorithm 1 Logic of BL series modeling

Input: The variance-covariance matrix Σ of sample data; The coefficient of uncertainty for sample data moments τ; The uncertainty matrix of opinions Ω; One of machine learning algorithm M;
Output: The optimal portfolio weights ω;
  Training machine learning models by M and Setting rules for predicting;
  Generating opinions for assets, pick matrix P and predicting return rate vector Q;
  Calculating the posterior moments $\hat{\mu}$ and $\hat{\Sigma}$ by equation (2.1);
  Making objective function with corresponding constraints;
  Solve objective function, get portfolio weights ω;
  return ω;

Algorithm 2 Logic of back testing

Input: The data frame of industry D; The length of training window Tr; The length of testing window Te;
Output: The cumulative return curve R;
  In training window, get portfolio weights ω from algorithm 1;
  In testing window, calculating the cumulative returns by portfolio weights ω;
  Rolling forward along the time axis;
  return R;

According to related theories in economics and finance [47] [16] [17], the nature of the excess return is to bearing the extra risk that can not be interpreted by common factors. From the perspective of risk sources, the risk could be classified as market risk, size risk, book-to-market ratio risk, profitability risk, and reinvestment risk. We select Fama French five factors (see TABLE I) as the basic features because more than 90% of the return rate could be explained from the point of the five fundamental factors. Scholars like Gu et al (2018) [19] pointed out that linear relationship among factors is given due attention in traditional econometric models, but there are relatively few researches on non-linear relationship and cross effect between factors. Actually, Some machine learning algorithms such as SVM, tree models, and neural networks are more suitable for non-linear modeling, but logistic regression, as well as classical econometric models, is good at calibrating linear relationships. In order to capture the non-linear relationship between factors, we construct high-order terms and cross terms of basic features at the stage of preprocessing.

TABLE I
FUNDAMENTAL FACTORS IN FAMA FRENCH MODEL

Item	Formula	Description
Mkt-Rf	$r_M - r_f$	the excess return rate of market, market risk
SMB	$r_S - r_B$	the return spread of small minus large stocks, size risk
HML	$r_H - r_L$	the return spread of cheap minus expensive stocks, value risk
RMW	$r_R - r_W$	the return spread of the most profitable firms minus the least profitable
CMA	$r_C - r_A$	the return spread of firms that invest conservatively minus aggressively

IV. EMPIRICAL RESEARCH

A. Data

In this paper, we select USA 30 industry portfolio daily data from Jan. 2010 to May. 2019 for numerical testing, all of the data could be downloaded from Ken. French website. According to the present academic results, the USA financial market is relatively mature and is approaching a weak efficient market. Moreover, the SNR is higher in industry data than individual stock data, which is more fit for model comparison. To examine the fitting and generation of the proposed portfolio model, we split time series data into the training set and testing set for parameters fitting and model validating respectively.

B. Models and benchmarks

To test the validity of proposed models, we select 1/N, mean-variance, worst-case robust portfolio, and S&P 500 index as benchmarks. The reasons for selecting these benchmarks are as follows: DeMiguel et al (2007) [20] have illustrated that 1/N is one of the significant references as benchmark portfolio. Also, the performance of 1/N in out-of-sample is rather stable. The mean-variance portfolio is the most classical model widely referenced in academics. And the worst-case robust portfolio shows high stable performance in out-of-sample with certain probability guarantees [5], [6], but there exists theoretical conservatism in the worst-case robust model. The three benchmarks provide references from different theoretical angles, and S&P 500 index provides a reference from a practical point of view. We would look at the performance of portfolio models under the no-shorting constraint, the concrete description of BL series models is listed in TABLE II.

TABLE II
DESCRIPTION OF BL SERIES MODELS AND BENCHMARK PORTFOLIOS

Models	Description
BL-LR	BL model integrated with investor opinions generating from Logistic Regression
BL-SVM	BL model integrated with investor opinions generating from SVM
BL-RF	BL model integrated with investor opinions generating from Random Forest
BL-Xg	BL model integrated with investor opinions generating from XGBoost
BL-MLP	BL model integrated with investor opinions generating from MLP
1/N	equally weighted portfolio
MV	mean-variance portfolio
Worst case	robust portfolio considering the worst case of parameters

C. Accuracy and stability

Quantile discretizer is the method employed to classify the rate of returns, and five levels are defined to form predictive opinions. We set 50% as the reference for accuracy which was tested by the one-sided t-test, statistical indicators contain accuracy and stability of machine algorithms used are shown in TABLE III. It can be found that the mean accuracy of Logistic Regression, Random Forest, and XGBoost are significantly higher than 50% and the range of accuracy is between 20% and 35%. The mean accuracy in descending order is XGBoost, Random Forest, Logistic Regression, SVM, and MLP. In terms of numeric stability, from the points of range and standard deviation, the order of stability from high to low is XGBoost, MLP, Random Forest, Logistic Regression, and SVM. Overall, XGBoost reveals the best comprehensive performance.

TABLE III
TABLE 3: ACCURACY OF MACHINE LEARNING ALGORITHMS

Algorithm	LR	SVM	RF	XGBoost	MLP
Lowest Acc.	39.78%	37.75%	43.49%	61.52%	37.92%
Mean Acc.	53.03%	52.32%	54.53%	70.47%	49.04%
Highest Acc.	69.76%	69.59%	68.79%	84.46%	62.28%
Range of Acc.	29.98%	31.84%	25.29%	22.93%	23.46%
Std. of Acc.	6.95%	7.53%	6.26%	5.02%	5.89%
P-value	0.01***	0.05**	0.00***	0.00***	0.71

* means significant on the level of 10%
** means significant on the level of 5%
*** means significant on the level of 1%.

D. Portfolio performance and analysis

The researches results of Bo et al [18] show that the investor opinions generating from machine learning algorithms are timeliness. In this paper, we set up the monthly warehouse transfer rule following the convention in the financial industry. The rolling window scheme (see Fig. 2) is used [48]. We set the training period to 150 days and the testing period to 30 days. BL series models consider the most up-to-date training samples at each period.

Fig. 3 reveals the cumulative returns of BL series models and benchmarks. It can be found that during the whole period of backtesting, BL-RF obtains the highest cumulative return, 230.9%. The profitability and risk-adjusted-performance indicators of portfolio models and benchmarks are shown in TABLE IV. Consistent with our theoretical analysis, all optimized strategies are higher than S&P 500 index in terms of cumulative return. As far as portfolio β, the mean-variance model and worst-case robust model prefer choosing a conservative and defensive investment style, which could also be derived from the objective functions. However, BL series models prefer choosing an active and aggressive investment style, that is, sacrificing certain stability for higher returns. From the view of macroeconomics, in 2010-2019 the U.S.

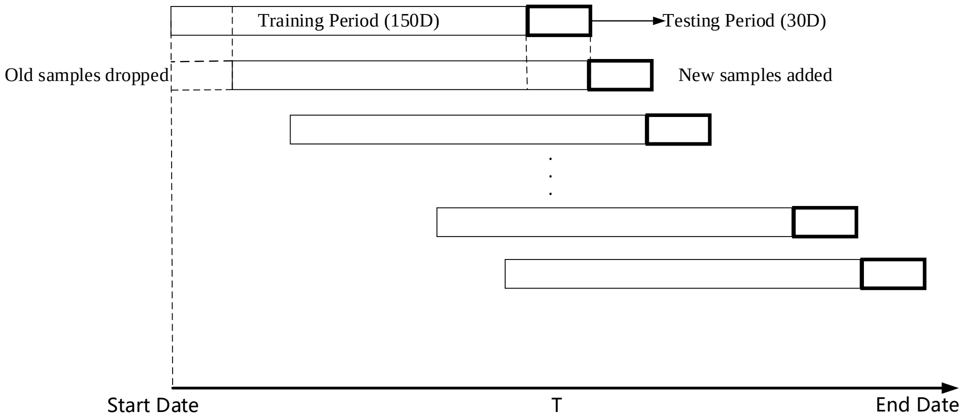


Fig. 2. Rolling window scheme for evaluating portfolios. The most recent obtainable data are added for updating investor opinions.

economy has been gradually out of the woods of the subprime mortgage crisis and began to recover slowly, which means the total tendency of economics is in a narrow range of shocks. Some BL series models such as BL-RF, BL-MLP, and BL-Xg still obtain significantly higher cumulative return and information ratio than other portfolios or benchmarks. In terms of risk, BL-RF, BL-MLP, and BL-Xg models get similar maximum drawdown (MDD) to S&P 500 index. Overall, the empirical results show that machine algorithms that are adapted in modeling non-linear relationships including tree models and shallow neural networks are more suitable for low SNR data like financial series. The conclusion is similar to the results obtained by Gu et al [19]. It is noteworthy that the worst-case robust portfolio earns a higher cumulative return than the mean-variance model, BL-LR model, and BL-SVM model in backtesting. Meanwhile, the robust portfolio has a lower annualized standard deviation than the market index. From the perspective of model structure analysis, the robust model [5] considers the decomposition of variance-covariance $\Sigma = VFV^T + \Omega$, and calibrates the uncertainty of the factor loading matrix V is a space of elliptical distribution, namely, $V = V_0 + W, \|W_i\|_g \leq \rho_i$. Nevertheless, in the BL series models, the posterior of the variance-covariance matrix is given by combining opinions and moments of samples. Generally, the model for the uncertainty of Σ is more passive in the robust portfolio model because of certain human-defined static parameters. Moreover, due to the dynamic nature of the financial market, the solution of a static robust model that pays more attention to risk-averse may result in an extreme case instead. Empirical results show that the robust model is at a disadvantage to other models in terms of MDD. To overcome the shortcoming of conservatism to some extent, investor opinions generating from machine learning algorithms provide the classical Black-Litterman model with predictive information. The validity of this approach is demonstrated by BL-RF, BL-Xg, and BL-MLP portfolio models in our numerical tests.

E. Weights Analysis

According to the present empirical analysis results, the portfolio weights of the Markowitz mean-variance model may be concentrated on a few assets, which would result in insufficient dispersion of non-systemic risk. One of the intuitive solutions to this problem is to add cardinality constraints then reformulate it to mixed-integer programming (MILP), but in the case of the large-scale portfolio, the efficiency of MILP is unsatisfactory. We would inspect the dispersion of weights in BL series models from an empirical perspective in this section.

The following two indicators are defined to evaluate the dispersion of portfolio weights:

Indicator 1: The number of assets invested in a portfolio, N. It is a direct indicator reflecting the degree of portfolio dispersion. When more assets are invested, a higher probability of reducing non-systemic risk and obtaining stable performance in out-of-sample [20].

Indicator 2: The maximum weight in the portfolio, W_{max}. If W_{max} is too high, it means that there exists excessively concentrated weight in a portfolio and high non-systemic risk. This indicator is inversely proportional to the degree of portfolio dispersion.

Fig. 4 to Fig. 6 reveal that the risk indicators defined above of BL-RF, BL-Xg, and BL-MLP. The descriptive statistics of indicators are shown in TABLE V and TABLE VI. It could be found that the number of allocated assets is about 10 to 20 in the three proposed models, that is, about 30% to 60% of all assets are invested. Regarding W_{max}, BL-RF obviously has lower W_{max} than BL-Xg and BL-MLP. It can be inferred from portfolio β that more aggressive assets are allocated in the BL-RF model. Accordingly, the BL-RF model obtains higher annual returns and information ratio at expense of bearing more MDD and annual standard deviation.

F. Transaction cost analysis

In this paper, we simulate a trading environment without friction for backtesting. However, there are lots of transaction costs in the real conditions, some explicit costs including stamp tax, commissions, and transfer fees can not be ignored. Moreover, there exist several indirect influences in the process of warehouse transfer such as slippage and price impact, etc. In this section, we inspect the sensitivity of proposed models to transaction costs. For simplicity, we set bi-directional transaction costs which is

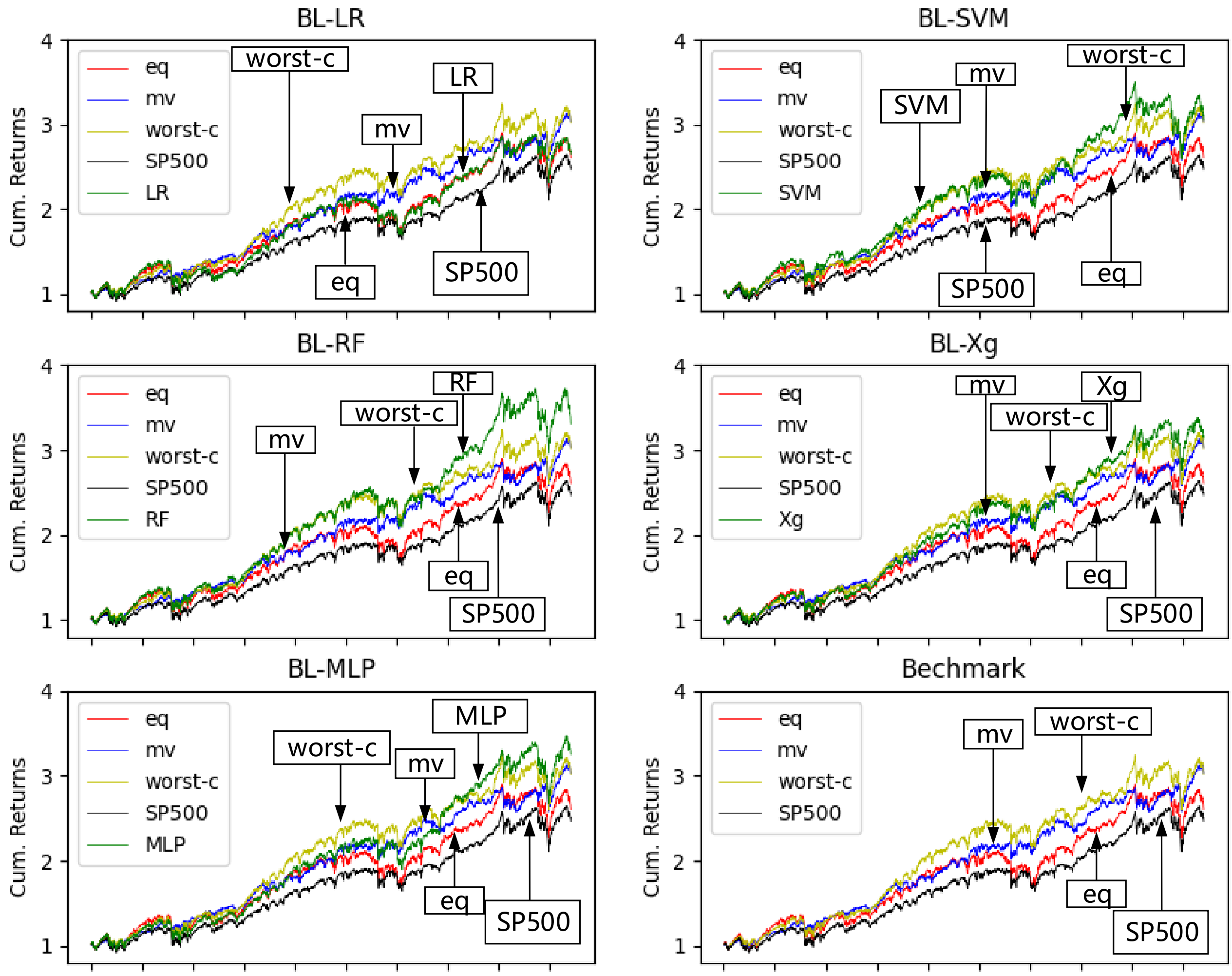


Fig. 3. Cumulative returns of portfolios

TABLE IV
INDICATORS OF PORTFOLIOS

Indicators	Benchmarks				BL series models				
	SP500	1/N	MV	Worst case	BL-LR	BL-SVM	BL-RF	BL-Xg	BL-MLP
Cum. Ret.	146.80%	160.78%	202.82%	211.68%	164.86%	205.24%	230.9%	214.1%	225.4%
Annual Ret.	14.94%	15.92%	18.62%	19.15%	16.20%	18.77%	20.26%	19.29%	19.94%
Daily Ret.	4.29%%	4.57%%	4.94%%	5.18%%	4.67%%	5.23%%	5.58%%	5.31%%	5.49%%
Annual Std.	14.91%	16.15%	11.56%	13.78%	16.67%	16.19%	16.15%	15.49%	15.90%
MDD	19.78%	22.99%	11.34%	25.70%	27.72%	25.77%	21.96%	19.84%	21.26%
Port. β	1.0	1.05	0.66	0.83	1.06	1.04	1.04	1.01	1.03
Sharpe Ratio*	1.00	0.99	1.61	1.39	0.98	1.16	1.25	1.25	1.25
Calmar Ratio	0.75	0.69	1.64	0.75	0.58	0.73	0.92	0.97	0.94
Info. Ratio	0.0	0.046	0.195	0.207	0.056	0.174	0.242	0.202	0.229
Jensen's α	0.0	0.002	0.087	0.067	0.0023	0.032	0.046	0.042	0.045
Treynor Ratio	0.149	0.151	0.280	0.23	0.152	0.180	0.194	0.191	0.193

* Assuming that the risk-free rate in US is 0 from 2010 to 2019

a fixed ratio of volume. Fig. 7(a) and Fig. 7(b) show the cumulative returns and the Sharpe Ratio of BL-RF, BL-Xg, and BL-MLP under different transaction costs respectively. We find that BL-RF is more sensitive to the change of transaction cost. When the transaction cost is about 5%%, BL series models still obtain cumulative returns higher than the market index. And when the transaction cost is slightly higher than 30%%, the performances of the BL series models and the market index are approximately flat. In terms of the Sharpe Ratio, BL-MLP reveals better performance than BL-Xg and BL-RF, hence BL-MLP is less sensitive to transaction costs than BL-Xg and BL-

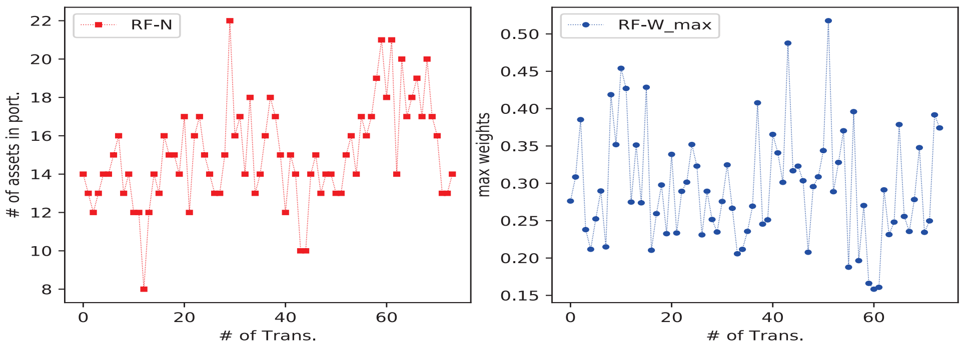


Fig. 4. Dispersion Indicators of BL-RF

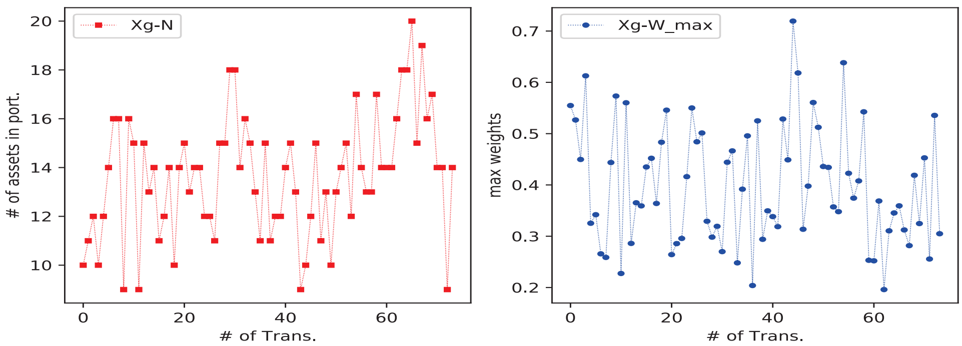


Fig. 5. Dispersion Indicators of BL-Xg

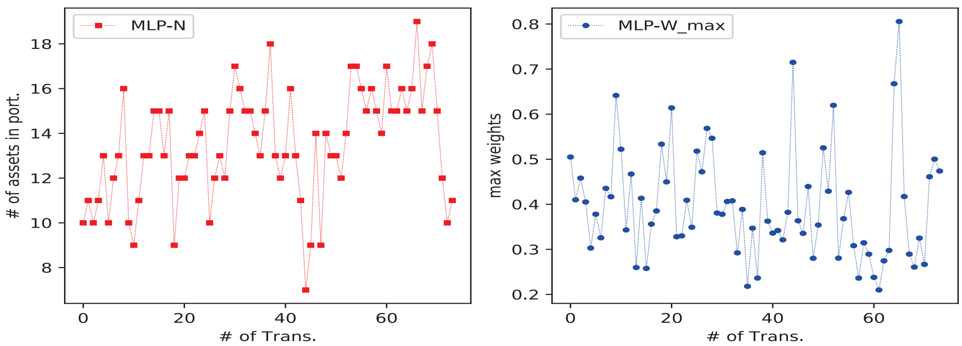


Fig. 6. Dispersion Indicators of BL-MLP

RF. Overall, with the development of artificial intelligence and the popularization of automatic trading, the transaction costs would gradually decrease.

V. Conclusions & Discussions

The issues of conservatism and stability in uncertainty are part of major research areas being under extensive study. Several attempts have been made to solve these problems in recent years. For example, some psychological theories such as regret theory and prospect theory are integrated into classical portfolio selection framework [49], [50]; skewness and kurtosis are taken into account breaking the limitations of normality and symmetry in asset distribution [50], [51], [52], [53], [54]; fuzzy programming and robust programming provide the fundamental theoretical framework for dealing with uncertainty, which has been widely used in both industries and academics [50], [53], [55], [56]; statistical forecasting methodology can also be incorporated into the Markowitz framework through the Bayesian theory [15], etc. On the one hand, the methodology of robust modeling exhibits excellent stability in out-of-sample numerical tests, but relative high conservatism. On the other hand, the Black-Litterman model paves the path for combining investor opinions with the classical Markowitz mean-variance portfolio model. This work aims at providing relatively objective investor opinions generating from multiple machine learning algorithms for the Black-Litterman portfolio model. Moreover, to further overcome the potential conservatism in portfolio models, a dynamic scheme is designed for regularly updating investor opinions and rolling portfolio models.

In this paper, we implement the Black-Litterman se-

TABLE V
DESCRIPTIVE STATISTICS OF N

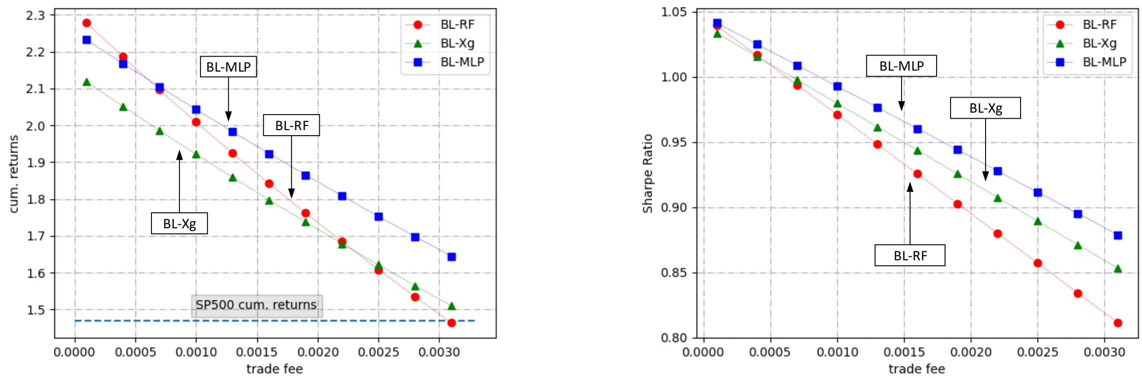
Model	obs*	Mean	Std	Min	25% quan.	Median	75% quan.	Max
BL-RF	74	15.00	2.62	8.0	13.0	14.5	17.0	22.0
BL-Xg	74	13.66	2.50	9.0	12.0	14.0	15.0	22.0
BL-MLP	74	13.47	2.51	7.0	12.0	13.0	15.0	19.0

* obs is the number of warehouse transfer during back test.

TABLE VI
DESCRIPTIVE STATISTICS OF $W_{\max}$

Model	obs*	Mean	Std	Min	25% quan.	Median	75% quan.	Max
BL-RF	74	29.52%	7.59%	15.83%	23.63%	28.91%	34.31%	51.77%
BL-Xg	74	39.94%	11.63%	19.60%	31.10%	37.15%	48.39%	71.94%
BL-MLP	74	39.85%	12.07%	20.99%	31.62%	37.93%	46.04%	80.56%

* obs is the number of warehouse transfer during back test.



(a) Cum. Returns

(b) Sharpe Ratio

Fig. 7. Performance of proposed portfolio models under different transaction costs

ries portfolio models incorporated with investor opinions generating from several machine learning algorithms, including Logistic Regression, SVM, Random Forest, XGBoost, and Multi-layer perceptron neural network. In the empirical research stage, the effectiveness and robustness of our proposed models are demonstrated using U.S.A 30 industry portfolio data set. For comparing and analyzing the performance of BL series models, we select S&P 500 market index, 1/N portfolio, classical Markowitz mean-variance portfolio, and worst-case robust portfolio as references. It can be observed that:

1) All of the BL series portfolio models obtain higher cumulative returns than S&P 500 and 1/N even if transaction costs are taken into account. Except for BL-LR, our proposed models outperform the mean-variance portfolio in terms of returns. BL-RF, BL-Xg, and BL-MLP obtain higher returns than all benchmarks.
2) There exists relatively large volatility in our proposed models than S&P 500 index, mean-variance model, and worst-case robust portfolio model. However, BL-RF, BL-Xg, and BL-MLP reveal smaller MDD than the worst-case robust portfolio model. Regarding the risk-adjusted indicators, all BL series models except BL-LR obtain higher Sharpe Ratios than S&P 500 and the 1/N portfolio. BL-RF obtains the highest Information Ratio, and BL-Xg obtains the second-highest Calmar Ratio (lower than the mean-variance model). Overall, BL series models reflect a relatively active and aggressive style which is opposite to these chosen benchmark portfolios.
3) From the weight analysis, we can find that, about $30\% \sim 60\%$ of all assets are selected, which is acceptable according to general risk management regulations. Additionally, Indicator 1 and Indicator 2 we defined in section 4.E reflect that the proposed models exhibit rather diversification and robustness to some extent.

Finally, possible further research direction may involve different constraints such as cardinality and higher moments. Also, in this work, we only consider the overall risk, several tail risk measures including VaR, CVaR would be considered in our future work.

ACKNOWLEDGMENT

Special thanks for the guidance from J.J. Gao, the associate professor in School of Information Management & Engineering, Shanghai University of Finance & Economics.

REFERENCES

[1] H. Markowitz, "Portfolio selection," *The journal of finance*, vol. 7, no. 1, pp. 77–91, 1952.

[2] M. Broadie, "Computing efficient frontiers using estimated parameters," *Annals of Operations Research*, vol. 45, no. 1, pp. 21–58, 1993.

[3] M. J. Best and R. R. Grauer, "On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results," *The review of financial studies*, vol. 4, no. 2, pp. 315–342, 1991.

[4] J. Gao and W. Wu, "Sparse and multiple risk measures approach for data driven mean-cvar portfolio optimization model," in *Optimization and Control for Systems in the Big-Data Era*. Springer, 2017, pp. 167–183.

[5] D. Goldfarb and G. Iyengar, "Robust portfolio selection problems," *Mathematics of operations research*, vol. 28, no. 1, pp. 1–38, 2003.

[6] S. Zhu and M. Fukushima, "Worst-case conditional value-at-risk with application to robust portfolio management," *Operations research*, vol. 57, no. 5, pp. 1155–1168, 2009.

[7] F. Black and R. Litterman, "Global portfolio optimization," *Financial analysts journal*, vol. 48, no. 5, pp. 28–43, 1992.

[8] ——, "Asset allocation: combining investor views with market equilibrium," *Goldman Sachs Fixed Income Research*, vol. 115, 1990.

[9] G. He and R. Litterman, "The intuition behind black-litterman model portfolios," *Available at SSRN 334304*, 2002.

[10] T. Idzorek, "A step-by-step guide to the black-litterman model: Incorporating user-specified confidence levels," in *Forecasting expected returns in the financial markets*. Elsevier, 2007, pp. 17–38.

[11] D. Alexander and D. Svetlana, "Application of ensemble learning for views generation in meucci portfolio optimization framework," *Review of Business and Economics Studies*, no. 1, 2013.

[12] M. Asad, "Optimized stock market prediction using ensemble learning," in *2015 9th International Conference on Application of Information and Communication Technologies (AICT)*. IEEE, 2015, pp. 263–268.

[13] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models," *Expert Systems with Applications*, vol. 103, pp. 25–37, 2018.

[14] W. Kristjanpoller and M. C. Minutolo, "A hybrid volatility forecasting framework integrating garch, artificial neural network, technical analysis and principal components analysis," *Expert Systems with Applications*, vol. 109, pp. 1–11, 2018.

[15] M. Kara, A. Ulucan, and K. B. Atici, "A hybrid approach for generating investor views in black–litterman model," *Expert Systems with Applications*, vol. 128, pp. 256–270, 2019.

[16] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, vol. 33, no. 1, pp. 3–56, 1993.

[17] ——, "Size, value, and momentum in international stock returns," *Journal of financial economics*, vol. 105, no. 3, pp. 457–472, 2012.

[18] B. L, Z. D. P, and e. a. Fang Y, "Multi-period portfolio selection model based on investor's views," *Systems Engineering — Theory & Practice*, vol. 37, no. 8, pp. 2024–2032, 2017.

[19] S. Gu, B. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," National Bureau of Economic Research, Tech. Rep., 2018.

[20] V. DeMiguel, L. Garlappi, and R. Uppal, "Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?" *The review of Financial studies*, vol. 22, no. 5, pp. 1915–1953, 2007.

[21] M. Bildirici and Ö. Ö. Ersin, "Improving forecasts of garch family models with the artificial neural networks: An application to the daily returns in istanbul stock exchange," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7355–7362, 2009.

[22] S. P. Das and S. Padhy, "Support vector machines for prediction of futures prices in indian stock market," *International Journal of Computer Applications*, vol. 41, no. 3, 2012.

[23] A. Hossain and M. Nasser, "Comparison of the finite mixture of arma-garch, back propagation neural networks and support-vector machines in forecasting financial returns," *Journal of Applied Statistics*, vol. 38, no. 3, pp. 533–551, 2011.

[24] O. Ican, T. B. Celik *et al.*, "Stock market prediction performance of neural networks: A literature review," *International Journal of Economics and Finance*, vol. 9, no. 11, pp. 100–108, 2017.

[25] N. Masoud, "Predicting direction of stock prices index movement using artificial neural networks: The case of libyan financial market," *Journal of Economics, Management and Trade*, pp. 597–619, 2014.

[26] Y. Kara, M. A. Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange," *Expert systems with Applications*, vol. 38, no. 5, pp. 5311–5319, 2011.

[27] F. Pérez-Cruz, J. A. Afonso-Rodriguez, J. Giner *et al.*, "Estimating garch models using support vector machines*," *Quantitative Finance*, vol. 3, no. 3, pp. 163–172, 2003.

[28] W. Shen, X. Guo, C. Wu, and D. Wu, "Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm," *Knowledge-Based Systems*, vol. 24, no. 3, pp. 378–385, 2011.

[29] M. Rout, B. Majhi, U. M. Mohapatra, and R. Mahapatra, "Stock indices prediction using radial basis function neural network," in *International Conference on Swarm, Evolutionary, and Memetic Computing*. Springer, 2012, pp. 285–293.

[30] A. Patra, S. Das, S. Mishra, and M. R. Senapati, "An adaptive local linear optimized radial basis functional neural network model for financial time series prediction," *Neural Computing and Applications*, vol. 28, no. 1, pp. 101–110, 2017.

[31] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Mathematics of operations research*, vol. 23, no. 4, pp. 769–805, 1998.

[32] L. El Ghaoui, F. Oustry, and H. Lebret, "Robust solutions to uncertain semidefinite programs," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 33–52, 1998.

[33] J.-S. Pang, "A new and efficient algorithm for a class of portfolio selection problems," *Operations Research*, vol. 28, no. 3-part-ii, pp. 754–767, 1980.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[35] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[36] X. Zhang, "Introduction to statistical leanring theory and support vector machines," *Acta automatica sinica*, vol. 26, no. 1, pp. 32–42, 2000.

[37] R. E. Caraka, R. C. Chen, S. A. Bakar, M. Tahmid, T. Toharudin, B. Pardamean, and S.-W. Huang, "Employing best input svr robust lost function with nature-inspired metaheuristics in wind speed energy forecasting," *IAENG Int. J. Comput. Sci*, vol. 47, no. 3, pp. 572–584, 2020.

[38] W. L. Al-Yaseen, "Improving intrusion detection system by developing feature selection model based on firefly algorithm and support vector machine," *IAENG Int. J. Comput. Sci*, vol. 46, no. 4, pp. 534–540, 2019.

[39] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[40] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.

[41] M. Arafat, S. Hoque, S. Xu, and D. M. Farid, "Machine learning for mining imbalanced data," *IAENG International Journal of Computer Science*, vol. 46, no. 2, pp. 332–348, 2019.

[42] W.-Y. Loh, "Classification and regression trees," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 14–23, 2011.

[43] L. Wu, J. Zhou, and Z. Li, "Applying of ga-bp neural network in the land ecological security evaluation," *IAENG International Journal of Computer Science*, vol. 47, no. 1, pp. 11–18, 2020.

[44] L. Zaman, S. Sumpeno, M. Hariadi, Y. Kristian, E. Setyati, and K. Kondo, "Modeling basic movements of indonesian traditional dance using generative long short-term memory network." *IAENG International Journal of Computer Science*, vol. 47, no. 2, pp. 262–270, 2020.

[45] G. Chen, J. Chen, Z. Zhang, and Z. Sun, "Short-term wind speed forecasting based on fuzzy c-means clustering and improved mea-bp." *IAENG International Journal of Computer Science*, vol. 46, no. 4, pp. 768–776, 2019.

[46] V. N. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," in *Measures of complexity*. Springer, 2015, pp. 11–30.

[47] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The journal of finance*, vol. 19, no. 3, pp. 425–442, 1964.

[48] H. Yun, M. Lee, Y. S. Kang, and J. Seok, "Portfolio management via two-stage deep learning with a joint cost," *Expert Systems with Applications*, vol. 143, p. 113041, 2020.

[49] R. Baule, O. Korn, and L.-C. Kuntz, "Markowitz with regret," *Journal of Economic Dynamics and Control*, vol. 103, pp. 1–24, 2019.

[50] X. Gong, C. Yu, L. Min, and Z. Ge, "Regret theory-based fuzzy multi-objective portfolio selection model involving dea cross-efficiency and higher moments," *Applied Soft Computing*, vol. 100, p. 106958, 2021.

[51] J. Wilcox, "Better portfolios with higher moments," *Journal of Asset Management*, vol. 21, no. 7, pp. 569–580, 2020.

[52] B. Chen, J. Zhong, and Y. Chen, "A hybrid approach for portfolio selection with higher-order moments: Empirical evidence from shanghai stock exchange," *Expert Systems with Applications*, vol. 145, p. 113104, 2020.

[53] X. Gong, C. Yu, and L. Min, "A cloud theory-based multi-objective portfolio selection model with variable risk appetite," *Expert Systems with Applications*, p. 114911, 2021.

[54] M. Aksaraylı and O. Pala, "A polynomial goal programming model for portfolio optimization based on entropy and higher moments," *Expert Systems with Applications*, vol. 94, pp. 185–192, 2018.

[55] G. Pandolfo, C. Iorio, R. Siciliano, and A. D'Ambrosio, "Robust mean-variance portfolio through the weighted l^p depth function," *Annals of Operations Research*, vol. 292, no. 1, pp. 519–531, 2020.

[56] D. Chakrabarti, "Parameter-free robust optimization for the maximum-sharpe portfolio problem," *European Journal of Operational Research*, 2021.