Computers & Industrial Engineering 177 (2023) 109023

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

A novel prediction based portfolio optimization model using deep learning

Yilin Ma [a,b,*], Weizhong Wang [c,*], Qianting Ma [d]

[a] School of Management, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, 210003, China
[b] Research Center for Information Industry Integration, Innovation and Emergency Management, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, 210003, China
[c] School of Economics and Management, Anhui Normal University, Wuhu, Anhui, 241002, China
[d] College of Finance, Nanjing Agricultural University, Nanjing, Jiangsu, 210095, China

ARTICLE INFO

Keywords:
Prediction based portfolio
Return forecast
LSTM network
Worst-case omega model

ABSTRACT

Portfolio optimization is an important part of portfolio management. It realizes the trade-off between maximizing expected return and minimizing risk. A better portfolio optimization model helps investors achieve higher expected returns under the same risk level. This paper proposes a novel prediction based portfolio optimization model. This model uses autoencoder (AE) for feature extraction and long short term memory (LSTM) network to predict stock return, then predicted and historical returns are utilized to build a portfolio optimization model by advancing worst-case omega model. In order to show the effect of AE, the LSTM network without any feature extraction methods is used as a benchmark in stock prediction. Also, an equally weighted portfolio is considered as a comparison to reveal the advantage of the worst-case omega model. Empirical results show that the proposed model significantly outperforms the equally weighted portfolio, and a high risk–return preference is more suitable to this model. In addition, even after deducting transaction fees, this model still achieves a satisfying return and performs better than the state-of-art prediction based portfolio optimization models. Thus, this paper recommends applying this model in practical investment.

1. Introduction

Portfolio management is a worthy issue in the process of financial investment, which has been widely concerned by investors and researchers. In recent years, many prediction based portfolio optimization models have been proposed for portfolio management. Under the same expected return, the portfolio optimization model with a better efficient frontier can keep the risk lower. Therefore, it is of great importance to further improve the performance of prediction based portfolio optimization models.

Since machine learning (ML) models, such as support vector regression (SVR), random forest (RF) and artificial neural networks (ANNs), have been used in stock prediction and achieved satisfying results (Ballings, Poel, Hespeels, & Gryp, 2015; Lu, Lee, & Chiu, 2009; Matías & Reboredo, 2012; Pang, Zhou, Wang, Lin, & Chang, 2018; Patel, Shah, & Thakkar, 2015; Sezer & Ozbayoglu, 2018; Tsai & Hsiao, 2010; Zhang, Li, & Guo, 2018) recommend combining ML models with classical portfolio optimization models in forming portfolios. Then, many scholars expand this research direction mainly in three directions. To be specific, first, some researchers only apply ML models in the stock preselection process, and then use selected stocks to build portfolios with classical portfolio optimization models (Paiva, Cardoso, Hanaoka, & Duarte, 2019; Vo, He, Liu, & Xu, 2019; Wang, Li, Zhang, & Liu, 2020). These works show that stocks selected by ML models can further improve the performance of classical portfolio optimization models. Second, some experts use ML models to predict future stock returns, and then apply the predicted information to generate new objective functions. These functions are used to supplement and improve the objective functions of classical portfolio optimization models (Ma, Han, & Wang, 2021; Ustun & Kasimbeyli, 2012; Yu, Chiou, Lee, & Lin, 2020; Yu, Chiou, Lee, & Yu, 2017). These researches present that objective functions generated by predicted information of ML models can further ameliorate classical portfolio optimization models. Third, some scholars utilize ML models for stock return prediction, and use predictive errors instead of historical returns in building classical portfolio optimization models (i.e., prediction based portfolio optimization models) since the normality of historical returns is smaller than predictive errors (Freitas, De Souza, & de Almeida, 2005; Freitas, Souza, & Almeida, 2009; Hao, Wang, & Xu, 2013; Ma, Han, & Wang, 2020). These studies show that predictive errors of ML models are more suitable than historical returns in building portfolio optimization models.

Thus, these three research directions improve the out-of-sample performance of classical portfolio optimization models from different

* Corresponding author.
E-mail addresses: ylmaseu@163.com (Y. Ma), wangweizhongky@163.com (W. Wang), mqt2626@sina.com (Q. Ma).

https://doi.org/10.1016/j.cie.2023.109023
Received 25 October 2020; Received in revised form 20 September 2022; Accepted 17 January 2023
Available online 20 January 2023

perspectives. In order to make full use of the advantages of these important discoveries to further improve the out-of-sample performance of prediction based portfolio optimization models, it is interesting and necessary to combine these research findings. As far as we know, there is no existing research focus on this point. Thus, the main contribution of this paper is to propose a more efficient prediction based portfolio optimization with return forecast. This model can help investors own higher expected returns under the same risk level and improve the out-of-sample portfolio performance in actual investment practices. Therefore, this paper tries to follow this direction by further extending relative researches. On the one hand, since autoencoder (AE) is a classical unsupervised learning model and frequently used for feature extraction (Bao, Yue, & Rao, 2017; Chen, Lin, Zhao, Wang, & Gu, 2017), this paper applies AE for feature extraction before stock return prediction. Moreover, the long short term memory (LSTM) network is often applied in stock prediction and achieves satisfying performance (Nelson, Pereira, & Oliveira, 2017; Qiu, Yang, Lu, & Chen, 2020; Sezer, Gudelek, & Ozbayoglu, 2020). Thus, the LSTM network is used in stock return prediction with extracted features from AE. On the other hand, some researchers apply the worst-case omega model in portfolio optimization, whose objective function is to maximize the worst-case Omega ratio (Kapsos, Christofides, & Rustem, 2014; Yu, Chiou, Lee, & Chuang, 2019). This ratio as a robust alternative avoids the limitations of the Omega ratio (Kapsos et al., 2014). Also, the omega model, which aims to maximize the Omega ratio, achieves satisfying performance in building portfolio optimization models with return forecast (Ma et al., 2021; Yu et al., 2020). Based on the superiority of the worst-case omega ratio, this paper applies the worst-case omega model in portfolio optimization process.

In general, the purpose of this paper is to build a more efficient prediction based portfolio optimization model with return forecast, i.e., AE+LSTM+OMEGA. More concretely, this model first applies AE for feature extraction and then uses the LSTM network with extracted features to predict future stock returns. Next, stocks with higher predicted returns are selected in the stock preselection process. Then, predictive errors are applied to build the worst-case omega model, and four additional objective functions are generated using predicted and historical returns. Last, the linearity weighted method is used to convert the multiple objective portfolio optimization model to a single objective portfolio optimization model. The proposed model applies historical returns as input features instead of financial indexes, technical indicators, and economic information, since this paper mainly focuses on demonstrating a better approach to optimize the weightings of stocks in a portfolio after the stock preselection process rather than choosing the optimal stock based on the input features.

On this point, the contributions of this paper to the existing literature are as follows. First, this paper applies the LSTM network in stock prediction with extracted features from AE, and stocks with higher predicted returns are selected, which guarantees high-quality stocks for portfolio optimization. Second, this model applies predictive errors instead of historical returns to build the worst-case omega model for the first time, which is more suitable for its normality hypothesis. Third, four objective functions are added to the above worst-case omega model based on the predicted returns and short-term performance of individual stocks, which can fully use predicted and historical information. Moreover, the linearity weighted method is used to convert the multiple objective model into a single objective model, which balances the importance of different objective functions. Fourth, this model significantly outperforms the state-of-art prediction based portfolio optimization models. Also, an equally weighted (EW) portfolio is used as a benchmark to reveal the advantage of the worst-case omega model. In addition, China Securities 100 Index (CSI 100) component stocks are used as total experimental data. This data ranges from 2007 to 2015, which contains nine years' data. Moreover, the last four years' data is adopted to show the out-of-sample performance of different models.

Table 1
Parameters of AE.

Parameter	Value
Hidden nodes	10, 20, 30, 40
Activation function	Relu, Sigmoid
Loss function	Mean absolute error
Optimizer	SGD, RMSprop, Adam
Learning rate	0.0001, 0.001, 0.01, 0.1
epoch	100, 500, 1000, 1500
Batch size	50, 100, 200

This paper is organized as follows. Section 2 introduces some utilized models. Section 3 gives the experimental process in detail. Experimental results are discussed in Section 4. Finally, the conclusion is drawn in Section 5.

2. Models

This section firstly presents the topologies of the AE model, the LSTM network, and their applied parameters. Then, prediction based worst-case omega model is clarified. Finally, the proposed model and benchmark model are given.

2.1. AE

AE is usually composed of encoder and decoder (Hinton & Salakhutdinov, 2006). The encoder first compresses input features to generate deep features, then the decoder reconstructs output features based on obtained deep features. The stochastic gradient descent method is often used for training, and the training aim of AE is to minimize the error between input features and output features. The main hyperparameters of the AE model contain hidden nodes, activation function, loss function, optimizer, learning rate, epoch, and batch size. The considered values of these hyperparameters are given in Table 1.

The values of hyperparameters in the AE model are fixed by using the grid research method. Through this method, the AE model with 40 hidden nodes, Relu activation function, Adam optimizer, 0.001 learning rate, 1000 epoch, and 100 batch size owns the lowest fitting error during the training process. Thus these values are the best. Since the AE model with these values can generate satisfying results for the remainder of the experiments according to our experiments and the grid research with 1152 potential combinations in Table 1 is time consuming for each training process, these values are fixed for the remainder of the experiments. This method is also used and proved to be efficient in Ma et al. (2020, 2021). Therefore, this model is used for feature extraction before stock prediction.

2.2. LSTM network

As a novel kind of recurrent neural network, LSTM network is introduced to surmount the known limitations of classical recurrent neural network (Graves & Schmidhuber, 2005). LSTM network has shown overwhelming performance than traditional recurrent neural network (Sezer et al., 2020). Usually, LSTM network is composed of one input layer, many hidden layers, and one output layer. The stochastic gradient descent method is used for training the LSTM network, and we apply earlystopping technology to reduce overfitting.

The focused hyperparameters of the LSTM network are hidden layers, hidden nodes, learning rate, recurrent dropout rate, dropout rate, optimizer, batch size, patient, activation function, and loss function. Relu function is utilized as activation function based on the recommendation in Orimoloye, Sung, Ma, and Johnson (2020). Moreover, the grid research method is utilized to discover the optimal hyperparameter. The considered values of other hyperparameters are shown in Table 2.

Similar to the AE, the optimal topology of the LSTM network is also specified. The final topology of the LSTM network is composed of

Table 2
Parameters of LSTM network.

Parameter	Value
Hidden layers	$1, 2, 3, \dots, 10$
Hidden nodes	$5, 10, 15, 20, 25, 30$
Learning rate	$0.0001, 0.001, 0.01, 0.1$
Patient	$0, 5, 10$
Batch size	$50, 100, 200$
Dropout rate	$0.1,\ 0.2,\ \dots, 0.5$
Recurrent dropout rate	$0.1,\ 0.2,\ \dots, 0.5$
Loss function	Mean absolute error
Optimizer	SGD, RMSprop, Adam

one hidden layer with ten hidden nodes. And, recurrent dropout rate, dropout rate, optimizer, learning rate, batch size, and patient are set as 0.2, 0.1, SGD, 0.001, 100, and 0, respectively.

Based on the obtained features from AE, this paper applies this LSTM network for stock prediction. For simplicity, AE+LSTM is used to denote this stock return prediction model.

2.3. Prediction based worst-case omega model

Kapsos et al. (2014) first proposed the worst-case Omega ratio, which avoids the limitation of the Omega ratio. As calculating the Omega ratio is based on stock returns' probability distribution, the final value will be biased if the probability distribution is inaccurate (Kapsos et al., 2014). The worst-case omega model is presented as follows.

$$\max \psi \tag{1}$$

Subject to

$$\delta\Big(\sum_{i=1}^{n} x_i \bar{r}_i^j - \tau\Big) - \frac{1-\delta}{T^j}\sum_{t=1}^{T^j} \eta_t^j \geq \psi \tag{2}$$

$$\eta_t^j \geq -\sum_{i=1}^{n} x_i r_{it}^j + \tau \tag{3}$$

$$\eta_t^j \geq 0 \tag{4}$$

$$\sum_{i=1}^{n} x_i = 1 \tag{5}$$

$$0 \leq x_i \leq 1 \tag{6}$$

$$t = 1, 2, \dots, T^j \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, l \tag{7}$$

where δ is risk–return preference, x_i represents the weight of stock i in portfolio, T^j indicates the jth normal distribution's sample period, $\bar{r}_i^j$ denotes stock i's the average return of sample period of the jth normal distribution, r_{it}^j is the return of stock i at time t under the jth normal distribution, η_t^j means auxiliary variable which is adopted to linearize the original portfolio model, l represents the number of normal distributions and n means the number of stocks in the portfolio. The efficient frontier of this model can be derived by solving this model with different δ's. Since δ is determined by investors, this paper investigates the model performance under different risk–return preferences in the following.

This paper sets T^j, τ, l as 10, 0 and 2 respectively on the basis of Yu et al. (2019, 2020). According to the conclusion of Freitas et al. (2005, 2009), the normality of ANN's predictive errors is better than the historical return series. Thus, it is more suitable to replace historical returns with predictive errors in building the worst-case omega model. In addition, we also add four additional objective functions according to Ustun and Kasimbeyli (2012) and Yu et al. (2020). Hence, the advanced worst-case omega model, i.e., prediction based worst-case omega model, is presented as follows.

$$\max \psi \tag{8}$$

$$\max \sum_{i=1}^{n} x_i \hat{r}_i \tag{9}$$

$$\max \sum_{i=1}^{n} x_i \bar{\varepsilon}_i \tag{10}$$

$$\max \sum_{i=1}^{n} x_i R_i^{20} \tag{11}$$

$$\max \sum_{i=1}^{n} x_i R_i^{5} \tag{12}$$

Subject to

$$\delta\Big(\sum_{i=1}^{n} x_i \bar{\varepsilon}_i^j\Big) - \frac{1-\delta}{10}\sum_{t=1}^{10} \eta_t^j \geq \psi \tag{13}$$

$$\eta_t^j \geq -\sum_{i=1}^{n} x_i \varepsilon_{it}^j \tag{14}$$

$$\eta_t^j \geq 0$$

$$\sum_{i=1}^{n} x_i = 1$$

$$0 \leq x_i \leq 1$$

$$t = 1, 2, \dots, 10 \quad i = 1, 2, \dots, n \quad j = 1, 2 \tag{15}$$

where $\bar{\varepsilon}_i^j$ represents stock i's average predictive error of sample period under the jth normal distribution, $\varepsilon_{it}^j = r_{it}^j - \hat{r}_{it}^j$, where $\varepsilon_{it}^j, r_{it}^j$, and $\hat{r}_{it}^j$ are predictive error, return, and predicted return of stock i at time t under the jth normal distribution respectively, $\hat{r}_i$ is the predicted return of stock i, R_i^{20} and R_i^{5} denote the average returns of stock i during the past 20 and 5 trading days respectively. Eq. (9)–(10) are to maximize the expected portfolio return and abnormal return respectively. Eq. (11)–(12) denote maximization of portfolio's short-term performance. As the first part of Eq. (13), $\sum_{i=1}^{n} x_i \bar{\varepsilon}_i^j$, means portfolio abnormal return and the second part, $\frac{1}{10}\sum_{t=1}^{10} \eta_t^j$, means portfolio loss. The abnormal return $\bar{\varepsilon}_i^j$ of stock i represents the mean value of returns minus predicted returns during the sample period of the jth normal distribution. The high risk–return preference (i.e., $\delta > 0.5$) means investors pay more attention to the portfolio abnormal return than the portfolio loss. The low risk–return preference (i.e., $\delta < 0.5$) indicates investors focus more on the portfolio loss than the portfolio abnormal return. The medium risk–return preference (i.e., $\delta = 0.5$) represents investors pay equal attention to both of them.

Since the linearity weighted method is frequently used to transform multiple objective models into single objective models and proved to be effective in existing literature (Ma et al., 2021; Yu et al., 2020, 2017), this paper applies this method to covert the prediction based worst-case omega model, which is shown as follows.

$$\max k_1 \psi + k_2 \sum_{i=1}^{n} x_i \hat{r}_i + k_3 \sum_{i=1}^{n} x_i \bar{\varepsilon}_i + k_4 \sum_{i=1}^{n} x_i R_i^{20} + k_5 \sum_{i=1}^{n} x_i R_i^{5} \tag{16}$$

Subject to

$$\delta\Big(\sum_{i=1}^{n} x_i \bar{\varepsilon}_i^j\Big) - \frac{1-\delta}{10}\sum_{t=1}^{10} \eta_t^j \geq \psi$$

$$\eta_t^j \geq -\sum_{i=1}^{n} x_i \varepsilon_{it}^j$$

$$\eta_t^j \geq 0$$

$$\sum_{i=1}^{n} x_i = 1$$

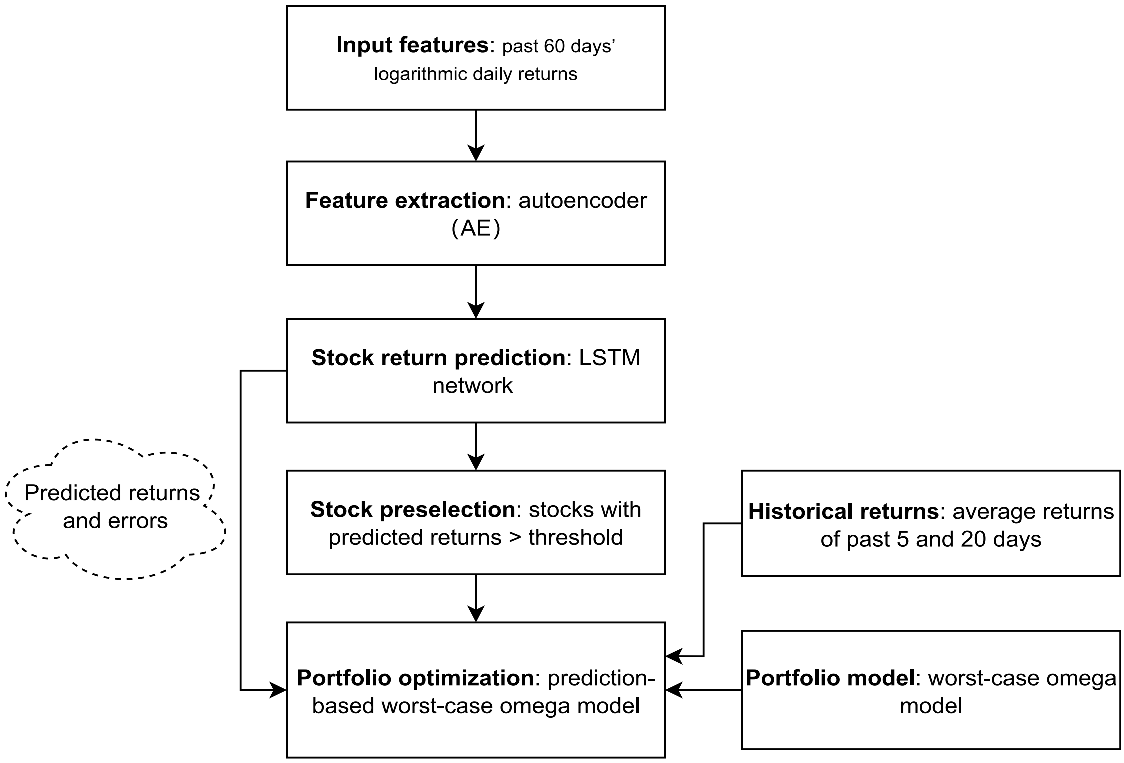

Fig. 1. The flow chart of the AE+LSTM+OMEGA model.

$0 \le x_i \le 1$

$t = 1, 2, \ldots, 10 \quad i = 1, 2, \ldots, n \quad j = 1, 2$

where k_1, k_2, k_3, k_4, k_5 are weighted coefficients. This paper sets k_1, k_2, k_3, k_4, k_5 as $2/8, 2/8, 2/8, 1/8, 1/8$ respectively since worst-case Omega ratio, expected portfolio return and abnormal return are more important than short-term performance of portfolio. Actually, many experiments are tested to verify the efficiency of these values.

2.4. Proposed model: AE+LSTM+OMEGA

This paper proposes a prediction based worst-case omega model with the AE+LSTM model's return forecast, i.e., AE+LSTM+OMEGA. This model first uses the AE+LSTM model to predict individual stock returns, then stocks with higher predicted returns are selected for portfolio formation. Last, the prediction based worst-case omega model is built for portfolio optimization. The flow chart of the AE+LSTM+OMEGA model is given in Fig. 1.

2.5. Benchmark models

In order to show the superiority of our proposed model, this paper uses an equally weighted portfolio with the AE+LSTM model, i.e., AE+LSTM+EW, as a comparison. This model adopts the AE+LSTM model in stock return prediction, then applies the equally weighted method to build a portfolio with selected stocks.

3. Experiments

In this paper, historical data of the CSI 100 component stocks is used as experimental data. The CSI 100 index consists of the largest 100 constituent stocks of the Shanghai–Shenzhen 300 index, which comprehensively reflects the overall performance of a group of companies with the largest market value and market effect in the Chinese stock market. Historical data ranges from January 4, 2007 to December 31, 2015. When neglecting the halted or unlisted stocks during this period, the remainder of CSI 100 index component stocks contains 49 stocks, which are displayed in Table 3.

Table 3
Selected stocks' tickers.

000001	000002	000063	000069	000538	000625
000651	000725	000858	000895	002024	300059
600000	600010	600011	600015	600016	600018
600019	600028	600030	600031	600036	600048
600050	600104	600111	600115	600150	600276
600340	600372	600398	600485	600518	600519
600585	600637	600690	600795	600837	600886
600887	600893	600900	601006	601111	601398
601988					

In this paper, the past 60 days' logarithmic daily returns are utilized as input features of the AE+LSTM model to forecast daily logarithmic stock returns. Since the fluctuation range of each input feature is different, it is necessary to preprocess them before training. To be specific, for input feature series $\{{}_q s_t\}_{t=1}^T$, q=1,2, …,60, ${}_q s_t$ is handled as follows

$$ {}_q s_t = \begin{cases} {}_q s_m + 5\,{}_q s_{mm} & if \quad {}_q s_t \ge \; {}_q s_m + 5\,{}_q s_{mm}, \\ {}_q s_m - 5\,{}_q s_{mm} & if \quad {}_q s_t \le \; {}_q s_m - 5\,{}_q s_{mm}. \end{cases} \tag{17} $$

where ${}_q s_m$ and ${}_q s_{mm}$ are the median of $\{{}_q s_t\}_{t=1}^T$ and $\{|{}_q s_t -{}_q\, s_m|\}_{t=1}^T$ respectively. Then, these features are standardized to unify the fluctuation range as follows,

$$ {}_q \hat{s}_t = \frac{{}_q s_t - \mu}{\sigma} \tag{18} $$

where μ and σ are $\{{}_q s_t\}_{t=1}^T$'s mean and standard deviation respectively. Since the daily logarithmic return series $\{r_t\}_{t=1}^T$ is mainly range from −0.1 to 0.1, it is processed as follows in order for better model training.

$$ r'_t = \begin{cases} min(10r_t, 1) & if \quad 10r_t \ge 1, \\ max(10r_t, -1) & if \quad 10r_t \le -1. \end{cases} \tag{19} $$

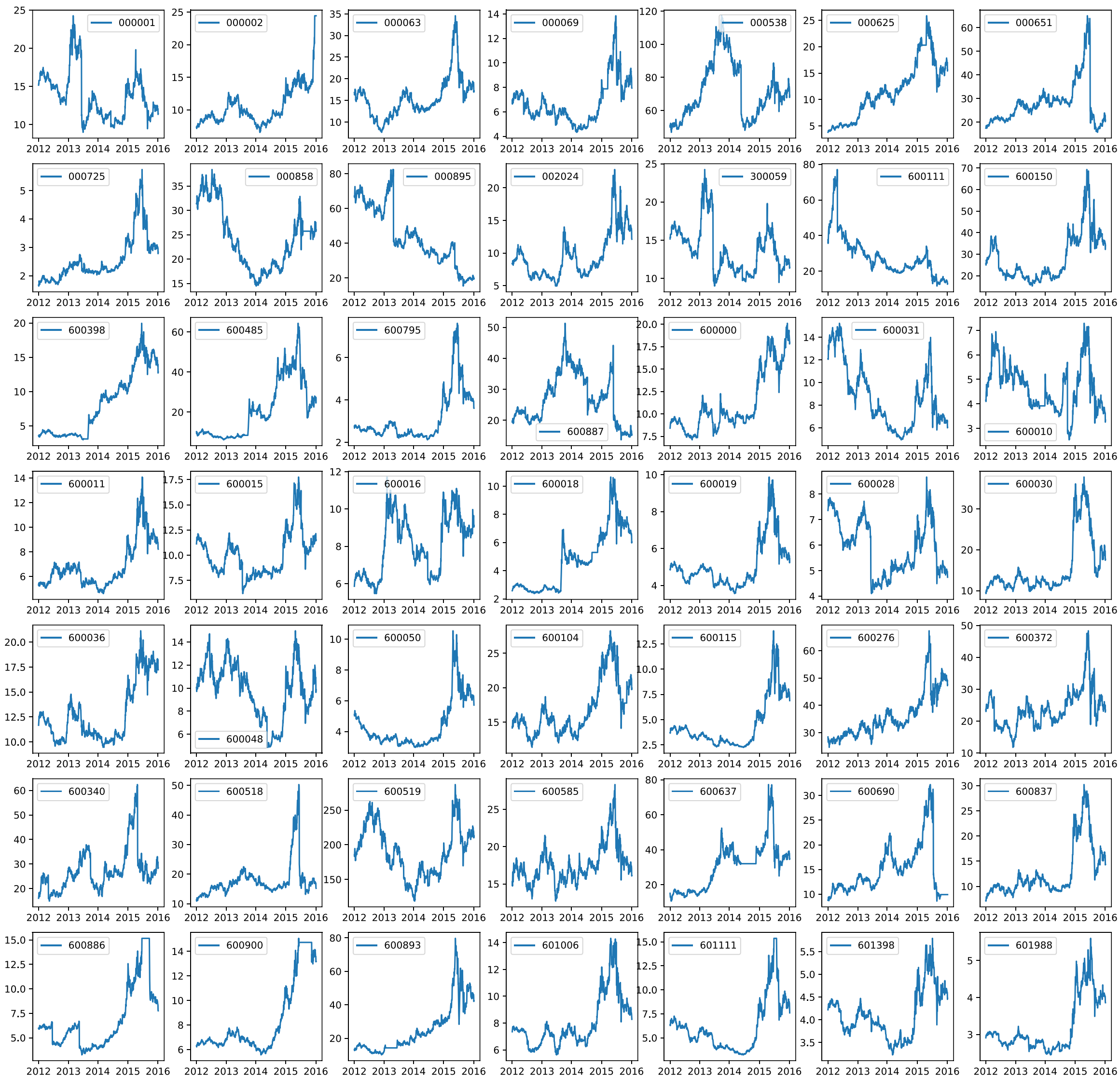

Fig. 2. Daily close price of different stocks.

Thus, the processed input features ${}_1\hat{s}_t,\ {}_2\hat{s}_t,\ \cdots,\ {}_{60}\hat{s}_t$ at time t and their corresponding label r_t' compose one sample, i.e., ${}_1\hat{s}_t,\ {}_2\hat{s}_t,\ \cdots,\ {}_{60}\hat{s}_t$ are used as an input for the AE+LSTM model, and label r_t' is the target output that we train the AE+LSTM model to predict with this input, and daily samples of nine years form the entire samples. Actually, input feature series $\{{}_q s_t\}_{t=1}^{T}$ is daily logarithmic return series $\{r_t\}_{t=1}^{T}$ with a lag of q days. Sliding window method is adopted to slice the entire samples in order to train the AE+LSTM model for stock return prediction in our experiment. To be specific, a sliding window contains six years' data, training set consists of the first four years' data, the next year's data is used as validation set, then the following year's data is utilized as test data. For example, when a sliding window contains daily data from January 4, 2007 to December 31, 2012, training data ranges from January 4, 2007 to December 31, 2010 is used to train the AE+LSTM model, validation data ranges from January 4, 2011 to December 30, 2011 is used to reduce overfitting problem, test data ranges from January 4, 2012 to December 31, 2012 is used measure the out-of-sample performance of this model. The rolling window is rolled yearly, and generates a continuous test data of non-overlapping test sets, i.e.,2012–2016, and the whole testing period consists of 941 trading days. Also, the stock prices of the selected stocks from January 4, 2012 to December 31, 2015 are presented in Fig. 2.

Based on the predicted returns of individual stocks, stocks with higher predicted returns are selected. And the portfolio weights of these stocks are obtained from the proposed model with predicted and historical returns of these stocks. To be specific, when using the AE+LSTM+OMEGA model to decide which stocks to trade and their proportions for day t+1, investors can first select stocks with the predicted returns of individual stocks obtained from day t, and then implement the AE+LSTM+OMEGA model to obtain their proportions with past 20 trading days' average return, past 5 trading days' average return, and predicted errors of selected stocks during the past 20 trading days. The portfolio is rebalanced on each trading day of the four year's testing period. Therefore, the last four years' data is applied to examine the out-of-sample performance of different models. This experiment implements AE and LSTM network with Keras deep learning package.

4. Experimental results

This section first discusses the predictive performance of the AE+LSTM model and stock preselection process. Then, when neglecting transaction fees, a trading simulation is organized to measure the profitability of its corresponding prediction based worst-case omega model. Last, in order to test the practical performance of this model, this

Table 4
The predictive performance of different models.

Year	Model		MAE	MSE	H_R	H_{R+}	H_{R-}
2012	AE+LSTM	mean	0.1397	0.0389	46.97%	47.03%	48.35%
		standard deviation	0.0421	0.0221	0.0423	0.0464	0.2132
2012	LSTM	mean	0.1398	0.0389	47.91%	47.67%	47.78%
		standard deviation	0.0422	0.0221	0.0421	0.0436	0.1682
2013	AE+LSTM	mean	0.1718	0.0594	47.02%	46.62%	53.61%
		standard deviation	0.0458	0.0300	0.0457	0.0454	0.2366
2013	LSTM	mean	0.1720	0.0594	47.52%	46.02%	50.22%
		standard deviation	0.0455	0.0298	0.0486	0.0827	0.1581
2014	AE+LSTM	mean	0.1523	0.0499	49.29%	49.49%	45.33%
		standard deviation	0.0373	0.0238	0.0325	0.0358	0.1086
2014	LSTM	mean	0.1525	0.0500	48.81%	49.88%	50.01%
		standard deviation	0.0373	0.0237	0.0389	0.0501	0.1721
2015	AE+LSTM	mean	0.2597	0.1318	50.28%	50.54%	44.83%
		standard deviation	0.0529	0.0476	0.0320	0.0311	0.0766
2015	LSTM	mean	0.2600	0.1322	50.42%	51.02%	46.47%
		standard deviation	0.0532	0.0478	0.0323	0.0390	0.1255

section also investigates the performance of this model after deducting its transaction fees.

4.1. Stock return prediction

This paper applies five metrics, i.e., mean squared error (MSE), mean absolute error (MAE), H_R, H_{R-} and H_{R+}, to comprehensively show different models' predictive performance in the stock return prediction process. These metrics are often adopted as evaluating indicators since they comprehensively describe the predictive ability (Freitas et al., 2009; Gandhmal & Kumar, 2019; Ma et al., 2020, 2021; Wang et al., 2020). The definitions of these metrics are as follows.

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(r_t - \hat{r}_t)^2 \tag{20}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|r_t - \hat{r}_t| \tag{21}$$

$$H_R = \frac{Count_{t=1}^{n}(r_t\hat{r}_t > 0)}{Count_{t=1}^{n}(r_t\hat{r}_t \neq 0)} \tag{22}$$

$$H_{R-} = \frac{Count_{t=1}^{n}(r_t < 0 \quad AND \quad \hat{r}_t < 0)}{Count_{t=1}^{n}(\hat{r}_t < 0)} \tag{23}$$

$$H_{R+} = \frac{Count_{t=1}^{n}(r_t > 0 \quad AND \quad \hat{r}_t > 0)}{Count_{t=1}^{n}(\hat{r}_t > 0)} \tag{24}$$

where $\hat{r}_t$ and r_t denote predicted return and actual return at time t, respectively. And, H_R, H_{R-}, and H_{R+} mean total hit rate, the accuracy of negative prediction, and positive prediction, respectively. In addition, since the predictive errors are used in forming prediction based worst-case omega models, this paper uses MAE and MSE metrics as key evaluation indicators in stock return prediction.

As we can see from Table 4, the AE+LSTM model's MAE and MSE are lower than the LSTM network. And, its standard deviation is similar to the LSTM network. This means that the LSTM network with AE can further improve its predictive performance. In addition, the predictive errors of the AE+LSTM model, i.e., the values of MAE and MSE, are also smaller than some existing literature (Ma et al., 2020; Weng, Lu, Wang, Megahed, & Martinez, 2018).

4.2. Stock preselection

Stock preselection is an important process before portfolio optimization, and better stock preselection can further improve the performance of portfolio management (Wang et al., 2020). Thus, it is necessary to select stocks with better predicted returns before portfolio optimization. This paper uses thresholds to select stocks since this approach is natural for their predicted returns, i.e., stocks with predicted returns larger than each threshold are selected to build portfolios, and the others are removed. The process can be described in the following equation.

$$Portfolio = \begin{cases} stock\ i\ is\ included\quad if\quad \hat{r}_i > threshold, \\ stock\ i\ is\ removed\quad if\quad \hat{r}_i \leq threshold \end{cases} \tag{25}$$

Generally, with the increase of the threshold, the size of the equally weighted portfolio is decreased, and portfolio risk is increased since the number of stocks with accurate predicted returns becomes less. For example, when an investor uses one threshold to select stocks in building a portfolio, this portfolio contains 20 stocks. There may be eight stocks that do not meet this standard, i.e., stocks whose predicted returns for day t+1 are larger than the threshold but whose actual returns on day t+1 are not satisfactory, which influences the average return of this portfolio on day t+1. When the investor increases the threshold to select stocks in building a portfolio, the portfolio contains fewer stocks, such as 10. Five stocks may not meet this standard, which increases the risk of portfolio failure since half of the stocks in this portfolio have unsatisfactory returns on day t+1. When the investor further increases the threshold to select stocks in building a portfolio, the portfolio contains fewer stocks, such as 3. Two stocks may not meet this standard, which further increases the risk of portfolio failure. Also, the portfolio sizes of different portfolios are not equal for the same period.

Thus, the performance of the equally weighted portfolio gets better first then becomes worse with the decrease of portfolio size. Therefore, it is necessary to balance between them. As the labels of the AE+LSTM model range from −1 to 1, the expected range of predicted returns is $[-1, 1]$. Actually, the predicted returns range from −0.035 to 0.15. This paper considers four thresholds, i.e., 0, 0.01, 0.02, and 0.03, to select stocks since the predicted returns of different stocks are mostly smaller than 0.04, and 0 is often used to select stock in relative literature.

Fig. 3 shows the size of different equally weighted portfolios. This figure indicates that the size of each equally weighted portfolio varies over time, and the differences between different portfolios are transparent. As we can see from this figure, the number of selected stocks with a threshold of 0.03 is extremely small, which means there is no need to investigate the performance of a threshold of 0.04. In order to measure the performance of different thresholds, equally weighted portfolios are built with selected stocks. Net value is adopted as evaluate metric since it begins with 1 yuan and describes accumulated value of different portfolios over time, and the performance of these equally weighted portfolios is presented in Fig. 4. As we can see from this figure, 0 is not the best threshold for AE+LSTM, and 0.02 is the best threshold since its corresponding equally weighted portfolio significantly outperforms the others. The poor performance of the threshold of 0.03 is due to the

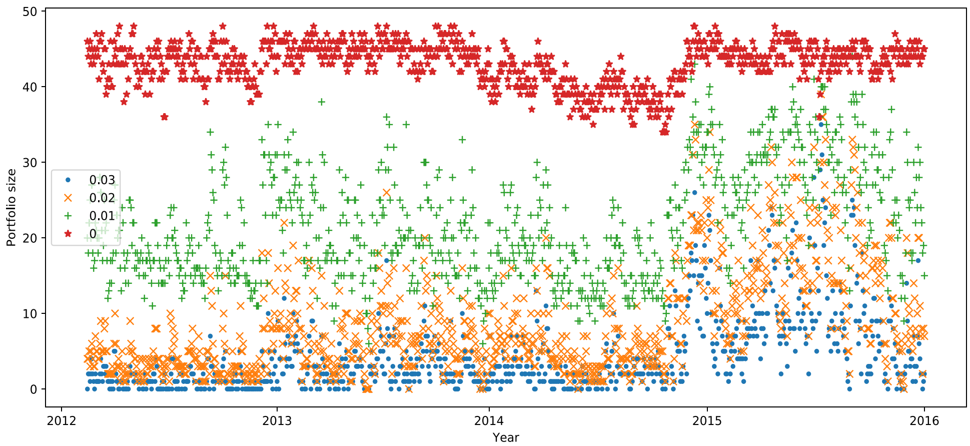

Fig. 3. Portfolio size of different equally weighted portfolios.

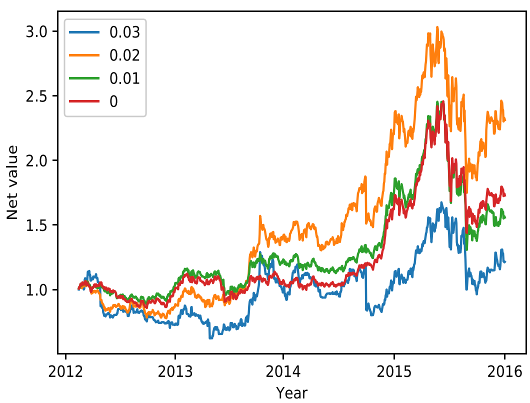

Fig. 4. Net value of different equally weighted portfolios.

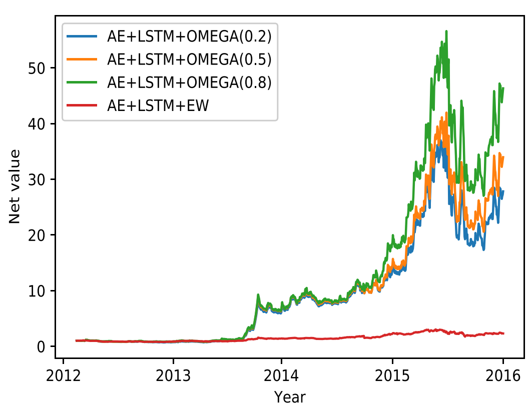

Fig. 5. Net value of AE+LSTM+OMEGA model with different δ's.

extremely small portfolio size, which means the number of stocks with accurate predicted returns is rare. Therefore, this paper applies 0.02 as the threshold to select stocks for portfolio optimization since this value can better balance satisfying return and low risk.

4.3. Trading simulation with different risk–return preferences

Based on the analysis of the AE+LSTM model in stock return prediction, we conduct a trading simulation to further investigate the performance of its corresponding prediction based worst-case omega model. We simulate trading like an investor. Typically, an investor needs to decide which stocks to buy and their proportions. Thus, the proposed model can generate such information for the investor before each trading day. Then the investor chooses to trade certain proportions of different stocks after obtaining their calculated proportions. This paper overlooks dividends, relative fees, leveraging, and short selling during investing. And, this trading simulation is conducted throughout the whole testing period.

In this paper, six metrics, i.e., excess return, information ratio, standard deviation, total return, turnover rate, and maximum drawdown are used to evaluate the performance of different models comprehensively. Actually, the excess return is the mean value of monthly excess returns, and excess return equals return minus the average return of the whole 49 stocks in this paper. The standard deviation denotes the volatility of monthly excess returns. The information ratio indicates risk adjusted return. The total return is the total cumulative return of daily portfolio returns in the whole testing period, where daily portfolio return is calculated based on the daily returns of different stocks in the portfolio. The maximum drawdown denotes the maximum holding risk on account of historical net values. The turnover rate indicates the turnover frequency of each transaction. The detailed definitions of maximum drawdown, information ratio, and turnover rate are as follows (Ma et al., 2020, 2021).

$$Maximum \quad drawdown = \max_{p<q} \frac{Nev_p - Nev_q}{Nev_p} \tag{26}$$

$$Information \quad ratio = \frac{excess \quad return}{stanard \quad deviation} \tag{27}$$

$$Turnover \quad rate = \sum_{i=1}^{n} |x_{i,t} - x_{i,t-1}| \tag{28}$$

where Nev_p is net value of time p, $x_{i,t}$ means the weight of stock i at time t and n denotes the number of stocks. In addition, in order to thoroughly compare the profitability of different models, excess return, total return, and information ratio are set as core metrics. Moreover, this paper uses AE+LSTM+OMEGA and AE+LSTM+EW to represent prediction based worst-case omega model and equally weighted portfolio model with AE+LSTM's return forecast respectively for simplicity.

In the following, this paper discusses the performance of the AE+LSTM+OMEGA model when δ equals 0.2, 0.5, and 0.8, which typically represent low, medium, and high risk–return preferences respectively. Also, other values such as 0.1, 0.5, and 0.9 can be used, and their performance is similar according to our experiments.

First, the AE+LSTM+OMEGA models with different risk–return preferences are compared. As we can see from Table 5, the excess return, information ratio and total return of the AE+LSTM+OMEGA model with the high risk–return preference are the highest. And the excess return, information ratio and total return of the AE+LSTM+OMEGA

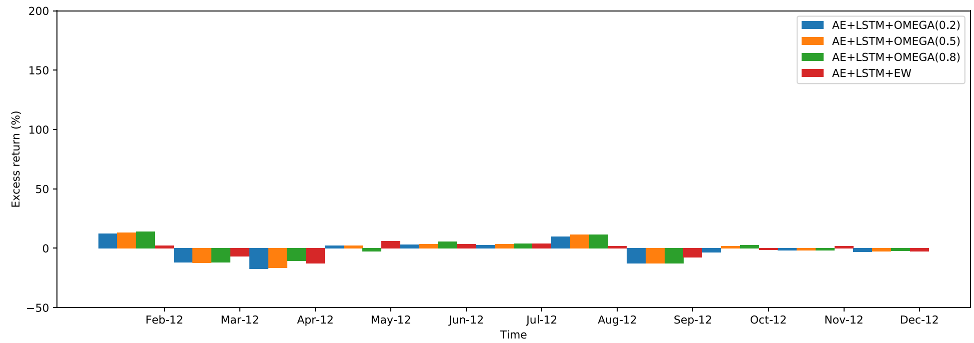

Fig. 6. Excess return of AE+LSTM+OMEGA model with different δ's in 2012.

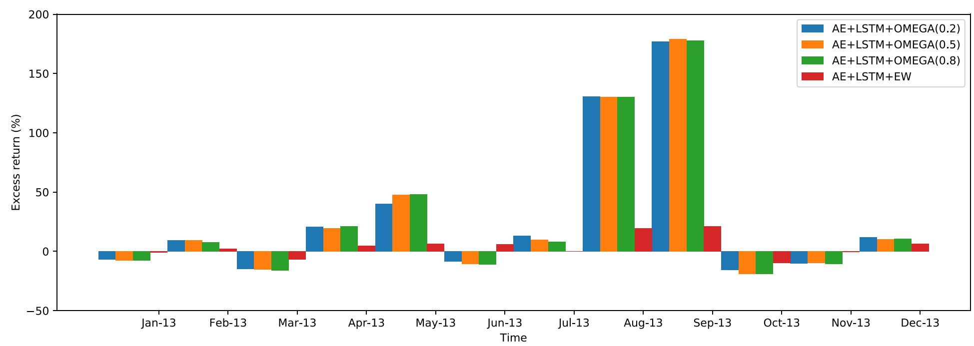

Fig. 7. Excess return of AE+LSTM+OMEGA model with different δ's in 2013.

Table 5
The performance of AE+LSTM+OMEGA model with different δ's.

Model	δ	ER	SD	IR	TOR	MD	TUR
AE+LSTM+OMEGA	0.2	9.68%	0.3426	0.2826	2683.06%	53.25%	145.51%
AE+LSTM+OMEGA	0.5	10.18%	0.3462	0.2939	3296.35%	50.98%	144.85%
AE+LSTM+OMEGA	0.8	10.93%	0.3453	0.3166	4532.75%	52.25%	143.99%
AE+LSTM+EW		0.91%	0.0653	0.1396	131.16%	42.33%	111.00%

ER, SD, IR, TOR, MD, and TUR denote excess return, standard deviation, information ratio, total return, maximum drawdown, and turnover rate respectively.

Table 6
Statistical information on the numbers of invested stocks of different models.

Model	δ	Mean	Standard deviation	Minimum	Maximum	Median
AE+LSTM+OMEGA	0.2	1.45	0.62	0	4	1
AE+LSTM+OMEGA	0.5	1.36	0.57	0	4	1
AE+LSTM+OMEGA	0.8	1.27	0.48	0	3	1
AE+LSTM+EW		8.73	7.03	0	39	7

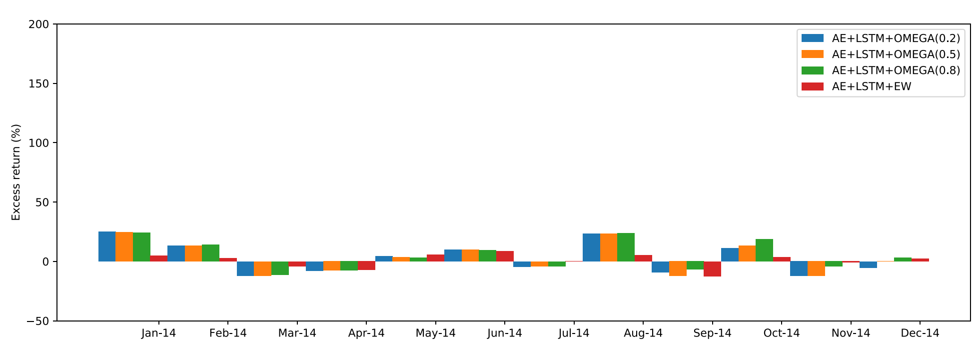

Fig. 8. Excess return of AE+LSTM+OMEGA model with different δ's in 2014.

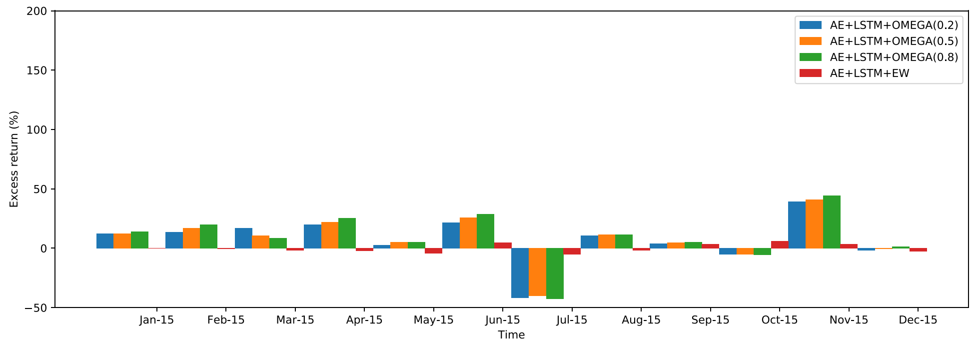

Fig. 9. Excess return of AE+LSTM+OMEGA model with different δ's in 2015.

Table 7
Statistical information on the daily returns of different models.

Model	δ	Mean	Standard deviation	Minimum	Maximum	Median
AE+LSTM+OMEGA	0.2	0.44%	0.0363	−10%	10%	0
AE+LSTM+OMEGA	0.5	0.46%	0.0369	−10%	10%	0
AE+LSTM+OMEGA	0.8	0.49%	0.0374	−10%	10%	0
AE+LSTM+EW		0.12%	0.0209	−10%	10%	0.03%

model with low risk–return preference are the lowest. Also, Figs. 5 to 9 show that the net value and excess returns of the AE+LSTM+OMEGA model with the high risk–return preference outperforms the others. These findings indicate that higher risk–return preference is more suitable to the AE+LSTM+OMEGA model and is recommended for the investor in our trading simulation. In the following, this paper only considers the high risk–return preference for the AE+LSTM+OMEGA model for simplicity.

Second, the AE+LSTM+OMEGA model is compared with the AE+LSTM+EW model. From Table 5, we can see that the excess return, information ratio and total return of the AE+LSTM+OMEGA model are much higher than the AE+LSTM+EW model. Also, Figs. 5 to 9 present that the superiority of the AE+LSTM+OMEGA model is not transparent over the AE+LSTM+EW model from the beginning of 2012 to Mar-2013, since then its net value and excess return increase sharply and are much higher than the AE+LSTM+EW model in general. Actually, the different excess returns between these two models are due to their different portfolio weights. As we can see from Table 6, the AE+LSTM+OMEGA model assigns more weights to fewer stocks than the AE+LSTM+EW model.

Moreover, the daily returns of different models are given in Fig. 10 and related statistical information is presented in Table 7. From this table, we can see that the mean value of daily returns of the AE+LSTM+OMEGA model with the high risk–return preference is more than four times that of the AE+LSTM+EW model. And the satisfying performance of the AE+LSTM+OMEGA model is mainly because its portfolio contains fewer stocks and the Chinese stock market experiences a bull market from Mid-2014 to Mid-2015. This finding means that the AE+LSTM+OMEGA model performs better than the AE+LSTM+EW model, which indicates that the worst-case omega model is preferable to equally weighted method for portfolio optimization.

4.4. Model comparison with transaction fee

According to the analysis of the above section, it is apparent that the AE+LSTM+OMEGA model significantly outperforms the AE+LSTM+EW model. However, the turnover rate of the AE+LSTM+OMEGA model is also much higher than the AE+LSTM+EW model. Since high turnover causes high transaction fees in the practical trading investment process, which largely erodes the final profit. Thus, it is necessary to investigate the practical performance of the proposed model after deducting its transaction fee caused by turnover. In the following, this paper applies a turnover of 0.05% per unit to approximate the complete transaction fee in practical trading investment for simplicity. Similarly, three evaluation metrics, i.e., excess return, information ratio, and total return, are set as key metrics to measure their performance.

First, the AE+LSTM+OMEGA model is compared with the stat-of-art models, i.e., DMLP+MSAD in Ma et al. (2020) and RF+MVF in Ma et al. (2021). From Table 8, we discover that the AE+LSTM+OMEGA model possesses the highest excess return, information ratio and total return. Thus, the AE+LSTM+OMEGA model performs better than the stat-of-art models. Second, the SVR+MV model in Hao et al. (2013) is also used as a benchmark. Table 8 shows that SVR+MV's excess return, information ratio, and total return are the lowest among these models. Also, Fig. 15 shows that the net value of the AE+LSTM+OMEGA model is higher than the other models. In addition, Figs. 11 to 14 present excess returns of different models after deducting their transaction fees. As we can see from Fig. 11, RF+MVF outperforms the other models in 2012. Figs. 12 to 14 show that the advantage of AE+LSTM+OMEGA is gradually revealed since March 2013, and its excess return is much higher than the other models in general.

Therefore, the performance of AE+LSTM+OMEGA is superior to DMLP+MSAD, RF+MVF, and SVR+MV, which further improves the out-of-sample performance of prediction based portfolio optimization models.

5. Conclusion

This paper proposes a prediction based worst-case omega model with AE+LSTM's return forecast, i.e., AE+LSTM+OMEGA. Under different risk–return preferences, the performance of this model is discussed. Also, this model is compared with the state-of-art models in existing works. The key findings are as follows.

First, the AE+LSTM model performs better than the LSTM network in stock return prediction. The MAE and MSE of the AE+LSTM model are lower than the LSTM network. And, stocks with higher predicted returns are more suitable for portfolio optimization. Second, when transaction fee is neglected, the proposed model significantly outperforms its corresponding equally weighted model, and the high risk–return preference is more suitable to this model. Finally, after deducting the transaction fee, the proposed model is superior to the state-of-art prediction based portfolio optimization models.

This paper extends the existing literature on portfolio formation with return forecasts. This study proposes a novel prediction based portfolio optimization model, which not only applies autoencoder and LSTM network in stock preselection, but also advances the worst-case

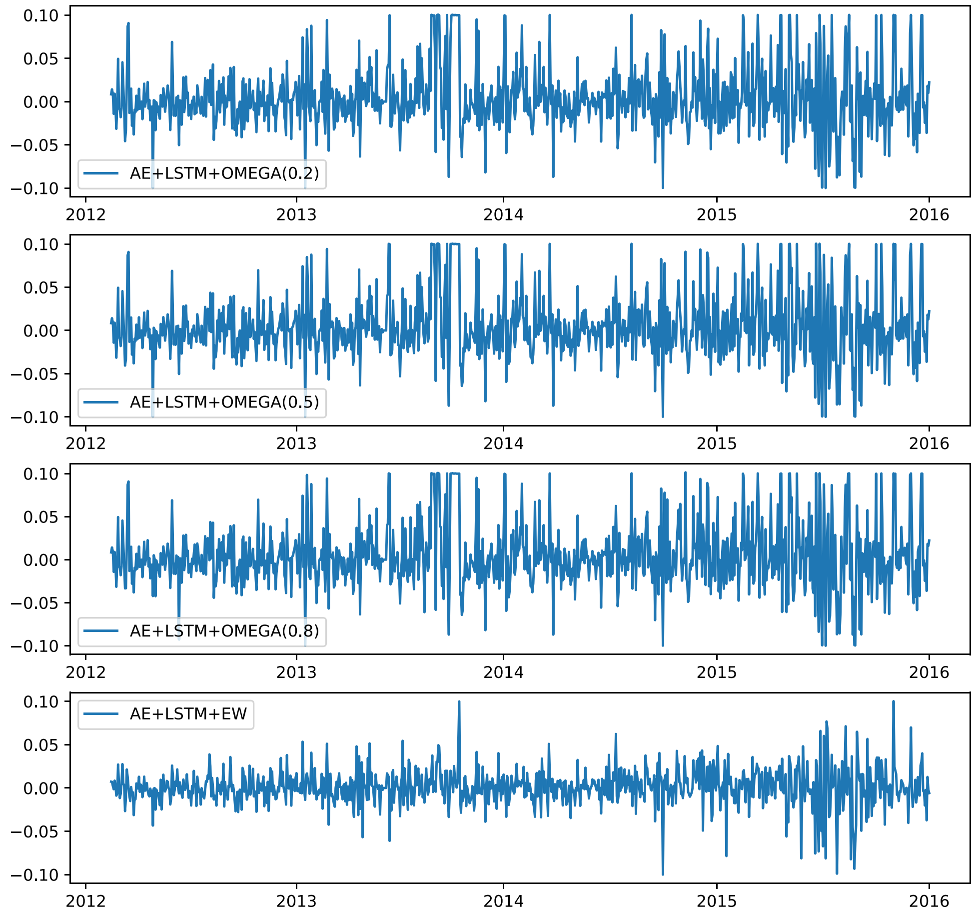

Fig. 10. Daily returns of AE+LSTM+OMEGA model with different δ's.

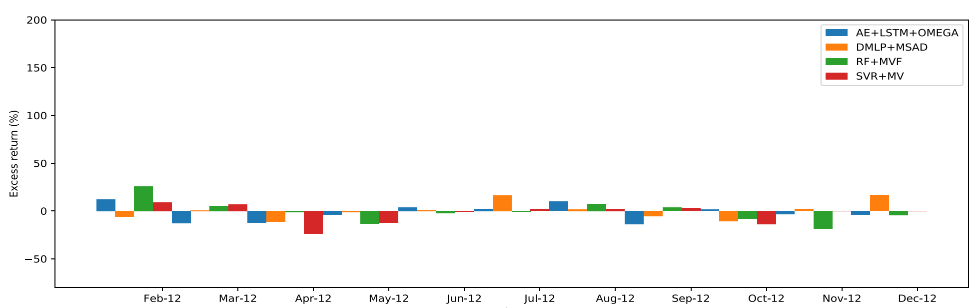

Fig. 11. Excess return of different models with transaction fee in 2012.

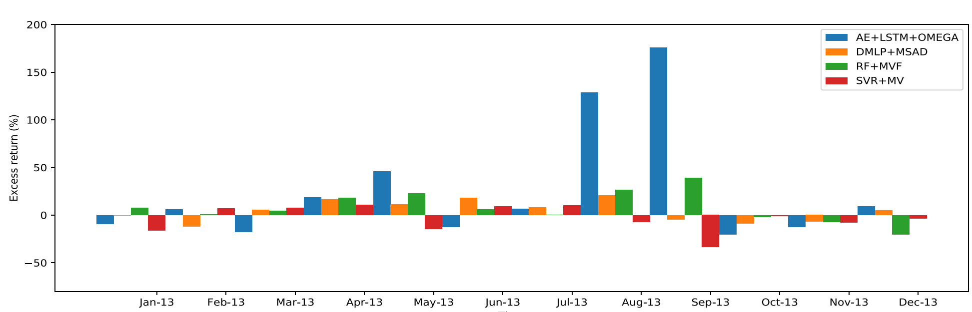

Fig. 12. Excess return of different models with transaction fee in 2013.

Table 8
The performance of different models with transaction fee.

Model	ER	SD	IR	TOR	MD
AE+LSTM+OMEGA	9.38%	0.3437	0.2728	2252.86%	55.27%
RF+MVF (Ma et al., 2021)	5.17%	0.1929	0.2679	806.52%	63.10%
DMLP+MSAD (Ma et al., 2020)	2.33%	0.1464	0.1590	195.85%	61.17%
SVR+MV (Hao et al., 2013)	−1.61%	0.1876	−0.0859	−12.33%	54.70%

ER, SD, IR, TOR, and MD denote excess return, standard deviation, information ratio, total return, and maximum drawdown respectively.

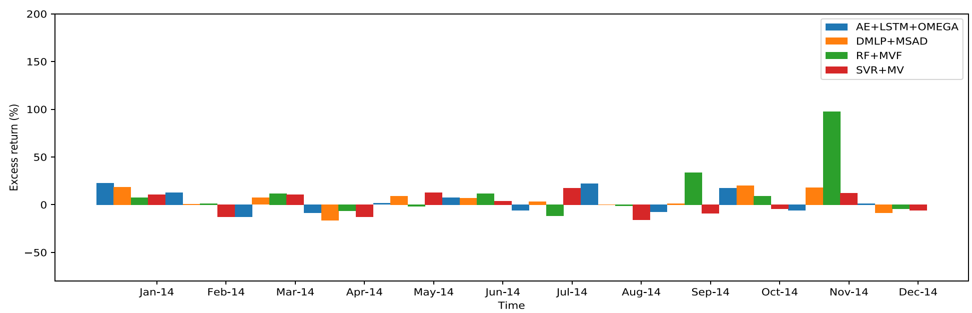

Fig. 13. Excess return of different models with transaction fee in 2014.

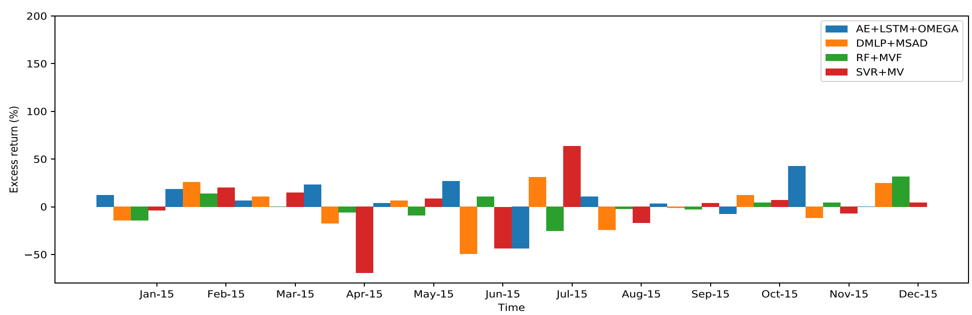

Fig. 14. Excess return of different models with transaction fee in 2015.

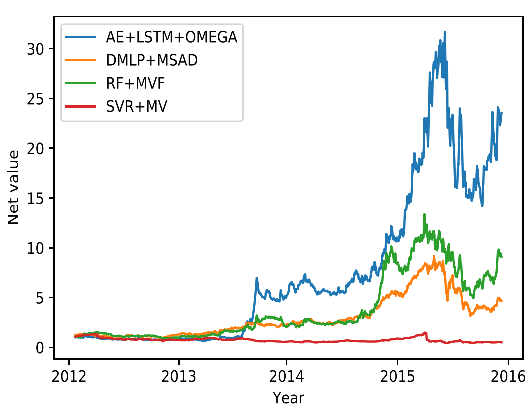

Fig. 15. Net value of different models with transaction fee.

omega model by using predicted returns and short-term performance of individual stocks. This model significantly outperforms the benchmark equally weighted portfolio model. Also, it is superior to the state-of-art models, which further improves the out-of-sample performance of existing prediction based portfolio optimization models.

Although this study achieves some useful conclusions, there are also some limitations existing in this work. First, we only utilize historical lagged returns as input features for stock return prediction since this paper only pays attention to portfolio formation. According to some existing literature, news, investor sentiment, economic situation, and financial information can influence short-term stock price. Future researchers can try these features in stock return prediction. Second, there may exist some more efficient portfolio optimization models than the worst-case omega model in building prediction based portfolio optimization models, which can further improve its out-of-sample performance.

CRediT authorship contribution statement

Yilin Ma: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Weizhong Wang:** Writing – review & editing. **Qianting Ma:** Writing – review & editing.

Data availability

Data will be made available on request.

Acknowledgments

The research is supported by National Natural Science Foundation of China, China (72101004), Jiangsu Provincial Double-Innovation Doctor Program, China (JSSCBS20210524), and Research Fund for Humanities and Social Sciences of Nanjing University of Posts and Telecommunications, China (NYY221009).

References

Ballings, M., Poel, D. V. D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, *42*, 7046–7056.

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS One*, *12*(7), Article e0180944.

Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2017). Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(6), 2094–2107.

Freitas, F. D., De Souza, A. F., & de Almeida, A. R. (2005). Autoregressive neural network predictors in the Brazilian stock markets. In *IEEE Latin American robotics symposium*. Brasil.

Freitas, F. D., Souza, A. F. D., & Almeida, A. R. D. (2009). Prediction-based portfolio optimization model using neural networks. *Neurocomputing*, *72*, 2155–2170.

Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. *Computer Science Review*, *34*, Article 100190.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, *18*, 602–610.

Hao, C. Y., Wang, J. Q., & Xu, W. (2013). Prediction-based portfolio selection model using support vector machines. In *Sixth international conference on business intelligence and financial engineering*. Hangzhou.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.

Kapsos, M., Christofides, N., & Rustem, B. (2014). Worst-case robust Omega ratio. *European Journal of Operational Research*, *234*, 499–507.

Lu, C. J., Lee, T. S., & Chiu, C. C. (2009). Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, *47*, 115–125.

Ma, Y. L., Han, R. Z., & Wang, W. Z. (2020). Prediction-based portfolio optimization models using deep neural networks. *IEEE Access*, *8*, Article 115393-115405.

Ma, Y. L., Han, R. Z., & Wang, W. Z. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, *165*, Article 113973.

Matías, J. M., & Reboredo, J. C. (2012). Forecasting performance of nonlinear models for intraday stock returns. *Journal of Forecasting*, *31*, 172–188.

Nelson, D. M. Q., Pereira, A. C. M., & Oliveira, R. A. D. (2017). Stock market's price movement prediction with LSTM neural networks. In *International joint conference on neural networks*. Anchorage.

Orimoloye, L. O., Sung, M. C., Ma, T. J., & Johnson, J. E. V. (2020). Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices. *Expert Systems with Applications*, *139*, Article 112828.

Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems with Applications*, *115*, 635–655.

Pang, X., Zhou, Y., Wang, P., Lin, W. W., & Chang, V. (2018). An innovative neural network approach for stock market prediction. *The Journal of Supercomputing*, http://dx.doi.org/10.1007/s11227-017-2228-y.

Patel, J., Shah, S., & Thakkar, P. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*, 259–268.

Qiu, Y., Yang, H. Y., Lu, S., & Chen, W. (2020). A novel hybrid model based on recurrent neural networks for stock market timing. *Soft Computing*, http://dx.doi.org/10.1007/s00500-020-04862-3.

Sezer, O. B., Gudelek, U., & Ozbayoglu, M. (2020). Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing*, *90*, Article 106181.

Sezer, O. B., & Ozbayoglu, A. M. (2018). Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Applied Soft Computing*, *70*, 525–538.

Tsai, C. F., & Hsiao, Y. C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems*, *50*(1), 258–269.

Ustun, O., & Kasimbeyli, R. (2012). Combined forecasts in portfolio optimization: A generalized approach. *Computers & Operations Research*, *39*, 805–819.

Vo, N. N., He, X., Liu, S., & Xu, G. (2019). Deep learning for decision making and the optimization of socially responsible investments and portfolio. *Decision Support Systems*, *124*, Article 113097.

Wang, W., Li, W., Zhang, N., & Liu, K. C. (2020). Portfolio formation with preselection using deep learning from long-term financial data. *Expert Systems with Applications*, *143*, Article 113042.

Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications*, *112*, 258–273.

Yu, J. R., Chiou, W. J. P., Lee, W. Y., & Chuang, T. Y. (2019). Realized performance of robust portfolios: Worst-case omega vs. CVaR-related models. *Computers & Operations Research*, *104*, 239–255.

Yu, J. R., Chiou, W. J. P., Lee, W. Y., & Lin, S. J. (2020). Portfolio models with return forecasting and transaction costs. *International Review of Economics & Finance*, *66*, 118–130.

Yu, J. R., Chiou, W. J. P., Lee, W. Y., & Yu, K. C. (2017). Does entropy model with return forecasting enhance portfolio performance? *Computers & Industrial Engineering*, *114*, 175–182.

Zhang, Y., Li, X., & Guo, S. (2018). Portfolio selection problems with Markowitz's mean–variance framework: A review of literature. *Fuzzy Optimization and Decision Making*, *17*, 125–158.