Journal of Computational Science & Engineering 58 (2022) 9-19

Available at http://www.asocse.org ISSN 1710-4068

Analysis and forecast of the development trend of Building-Integrated Photovoltaic plate index

Li Wu[1], Shupei Du[1], Dong liu[2], Jieyu Ding [1*]

[1]School of Mathematics and Statistics, Qingdao University, Qingdao 266071, China

[2]School of Economics, Qingdao University, Qingdao 266071, China

Abstract: **Building-Integrated Photovoltaic (BIPV) plays an important role in the realization of "carbon peak" and "carbon neutralization" in China. Therefore, it is important to study the BIPV plate index to predict future trends and make relevant policies. Firstly, some valuable data which can effectively reflect the stock trend are screened and preprocessed. Define a calculation formula of a plate index and a moving average line of the plate index, using entropy weight method, we select the index that can depict the market value of each stock to calculate the corresponding weight of each stock, and then get the plate index and moving average price, and draw the correlative image. Then, divide the historical data into a training set and a test set (8:2), process the features through feature engineering, use the Gradient Boosting Decision Tree model (GBDT) to learn the historical sector index data, and predict the trading day index. The predicted data and the calculated real data are subjected to error analysis and parameter adjustment to obtain a fitted model, and draw the correlation coefficient diagram between the BIPV sector index and the Shanghai Stock Exchange index to obtain the correlation coefficient between them. Finally, take the average value of VaR value of each stock within three months as the risk index of the stock, sort the stocks, and establish a portfolio optimization model: mean-variance portfolio model, give the optimal investment plan, adopt PEST The analysis method analyzes the future development trend of China's BIPV industry.**

Key words: Feature engineering; GBDT; Grid search; Quantitative trading; VaR value.

Received 03 September 2022, Revised 10 September 2022, Accepted 17. September 2022

*Corresponding Author: Jieyu Ding, djy@qdu.edu.cn

1. Introduction

Building-Integrated Photovoltaic (BIPV) is a new concept of utilizing solar energy to generate electricity. It makes full use of resources such as industrial buildings and public building roofs to implement distributed PV power generation projects and integrate PV products into buildings. It is both a power generation device and an external structure of the building part, effectively reducing costs without sacrificing aesthetics. In recent years, the BIPV market is in its infancy. The newly added and renovated building area in China can reach 4 billion square meters every year, and the market potential is huge. In recent years, the BIPV market is still in its infancy. The annual increase and renovation of the building area can reach 4 billion square meters, the market potential is huge.

At present, under the environment of global warming and depletion of fossil energy, vigorously developing renewable energy has become the consensus of the world. Solar power generation is the frontier of science and technology in the 21st century, and governments around the world support solar power generation. As a new concept of improving the climate environment and realizing energy regeneration in the future, building integration of PV is undoubtedly at the forefront of the future PV application field, with very broad development prospects and huge market potential. In order to predict the development trend of the emerging sector index of BIPV, and to further understand the future development trend of the corresponding industry in the sector, this paper studies from the following aspects.

Using the known data, a moving average model of the BIPV sector index is established. The established model is analyzed and corrected, and the future development trend of the sector is predicted according to the corrected model. The correlation analysis of the Shanghai Stock Exchange Index and the BIPV sector index is carried out. Assess the investment risk of individual stocks in the BIPV sector, and give the optimal investment plan. Through the above research, the future development trend of China's BIPV industry is analyzed.

2. Journals reviewed

With the continuous introduction of China's policies, 31 places including Beijing and Shanghai have successively issued relevant policies for photovoltaic building integration (BIPV) in the next three to five years, affecting the development of China's BIPV and related industry stock markets. As a long-term credit tool of the capital market, the stock market can reflect the current economic form to a certain extent. How to better build the relevant index of BIPV to reflect and predict the current market trend is a problem that investors are concerned about. Because the stock itself has multiple volatility, predicting the sector index is a good method. There are many regression models based on time series. Adebiyi predicts the stock price through ARIMA model, and the results

2022@ The American Computational Science Society. All rights reserved.

show that this model has short-term prediction potential[4]. Due to the characteristics of financial time series such as heteroscedasticity, autoregressive conditional heteroscedasticity model (arch) and its generalized variant (GARCH) are also widely used in financial time series analysis. Wei Yu applied GARCH to the Shanghai and Shenzhen 300 index to predict its volatility[5] Hassan used hidden Markov (HMM) model to model the stock market and obtained more accurate prediction results[6]. In addition to traditional statistical models, some models in machine learning are also commonly used in time series prediction. Meesad et al. Used SVR combined with different window setting methods to model stock data, and verified the effectiveness of this model[7]. Deep neural network (DNN) also has unique advantages in dealing with big data problems, and its effectiveness in financial time series prediction has also been proved[8]. Yang Qing et al. Used long short-term memory (LSTM) to conduct different term studies on 30 stock indexes around the world, verifying the good performance of LSTM[9].

3. Basic Assumptions

Assumption 1: Assume that the BIPV sector index is only affected by the indicators given in Annex 1;
Assumption 2: Assume that the volume and price information can adequately reflect the trend of a stock;
Assumption 3: Assume that the confidence interval used by the VaR value estimation method is 95% and 99%;
Assumption 4: Efficient market theory is established in this question.

Table 1: Symbol description

Symbol	Significance	Symbol	Significance
v_i	Stock value	d	Information effect value
N	Total days	Cov	Covariance
L	Sector Index	ρ	Pearson correlation coefficient
X_i	Indicator	W_E	Window period
X	Sample matrix	$\hat{\sigma}_t^2$	Variance
Z	Normalized matrix of X	r_p	Investment hinders return
P	Probability matrix	ω_i	Proportion of each asset
e	Information entropy	r_i	Rate of return on each asset
w_i	Index weight	σ_p	Risk
E	Expect	μ_p	Expected rate of return

4. Model establishment and solution

Through data preprocessing, the number of valid trading days is selected to calculate the sector index, so as to facilitate the calculation of the moving average and the subsequent learning and prediction of the data.

4.1 Sector Index Model

(1) Index selection

For the weight, the market value and total share capital of individual stocks are used as the characterization model to determine. The weight is determined by using the entropy weight method through the four indicators of "days on the market", "historical average trading volume", "historical average trading volume" and "days when the effective trading day rises by more than 5%".

(2) Calculate the index weight

①Positivization of index

Step 1: Forward the four indicators, that is, convert all extremely small indicators into very large indicators, and obtain a positive matrix：

$$X=\begin{pmatrix} x_{11} & \cdots & x_{1j} \\ \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} \end{pmatrix} \quad (4.1)$$

Step 2: Normalize the normalization matrix X to get the normalized matrix Z, where:

$$z_{ij}=\frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}} \quad (4.2)$$

2022@ The American Computational Science Society. All rights reserved.

If there are negative numbers in the matrix Z, another normalization method is applied to X to obtain a normalized matrix $\tilde{z}$, where：

$$\tilde{z}_{ij} = \frac{x_{ij} - \min\{x_{ij}, x_{2j}, ..., x_{nj}\}}{\max\{x_{ij}, x_{2j}, ..., x_{nj}\} - \min\{x_{ij}, x_{2j}, ..., x_{nj}\}} \quad (4.3)$$

The following normalization matrices are unified as Z.

②The Entropy Weight Method (EWM)：

EWM is an objective weighting method, based on the principle that the smaller the degree of data variation, the less the amount of information reflected, and the lower the corresponding weight.

Step 1: Compute the probability matrix p, where $p_{ij} = \frac{z_{ij}}{\sum_{i=1}^{n} z_{ij}}$, and $\sum_{i=1}^{n} p_{ij} = 1, j = 1,2,...m$, that is, the sum of the probabilities corresponding to each indicator is 1.

Step 2: For the jth indicator, calculate its information entropy:

$$e_j = -\frac{1}{\ln n}\sum_{i=1}^{n} p_{ij} \ln(p_{ij}) \quad (4.4)$$

Then its information effect value is $d_j = 1 - e_j$, the larger the d_j ,the more information.

Step 3: Normalize d_j to obtain the entropy weight of the four indicators:

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j} \quad (4.5)$$

(3) Sector index model

For the calculation of sector index and moving average price, the following formula is defined:

$$L = \frac{\sum_{i=1}^{37} v_i \cdot w_i}{N} \times 1000 \quad (4.6)$$

Here, $\sum_{i=1}^{37} v_i \cdot w_i$ is the sum of N-day closing prices.

(4) Model solution

Import module: Use the Pandas library to read data, use the sheet to which each stock belongs as a key, and store the two-dimensional data matrix in each sheet as a value in a dictionary, which is convenient for subsequent loop calls.

Carry out data preprocessing: screen out all the trading day data that can reflect the valid trading days of the PV construction industry sector, and fill in the blanks for the actual trading days of the suspension to facilitate the calculation of the unified index dimension in the later stage.

Calculate the characteristics of indicators: Use the relevant statistical functions and if functions in the Pandas library to solve the indicators of "days on the market", "historical average trading volume", "historical average trading volume", and "the number of days that the effective trading day rose by more than 5%". Use the entropy weight method to find the weight of each stock, see Table 2 (retain four decimal places):

4.2 Prediction model

Use the GBDT model to learn the plate data on valid trading days, predict the plate data and then compare it with the real data for error analysis. A model of the trend of sector indices.

(1) GBDT algorithm

GBDT is an iterative decision tree algorithm that constructs trees in a sequential manner, with each tree trying to correct the mistakes of the previous tree. By default, gradient-lifted regression trees do not randomize, but strong pre-pruning is used. Gradient boosted trees usually use trees with very small depth, so that the model takes up less memory and makes predictions faster.

The iterative model is $f_m(x) = f_{m-1}(x) + T_m(x)$ (4.7)

Where $f_{m-1}(x)$ is the m-1 th iteration model, that is, the last iteration model, $T_m(x)$ is the decision tree to be built this time, and is also the decision tree that fits the residual value of the previous model, and $f_m(x)$ is the new model generated after this iteration.

Algorithm steps:

①Initialize $f_0(x) = 0$;

2022@ The American Computational Science Society. All rights reserved.

②When $m = 1,2,\cdots,M$, calculate the residual $r_{mi} = y_i - f_{m-1}(x)$, fit the residual, get the decision tree $T_m(x)$, and update $f_m(x) = f_{m-1}(x) + T_m(x)$;

③When the error or the number of iterations meet the specified requirements, the lifting tree of the regression problem is obtained as: $f_M(x) = \sum_{m=1}^{M} T_m(x)$.

Table 2: The weight of each stock obtained by the entropy weight method

Securities name	Weights	Securities name	Weights	Securities name	Weights
CSG A	0.0188	SHENZHEN ENERGY	0.0303	TUNGHSU AZURE	0.0287
Fangda Group	0.0476	SHENZHEN SEG	0.0233	BAUING DECORATION	0.0267
Southeast Space Frame	0.0076	YANHUA SMARTECH	0.0045	TORRAY Solar	0.0389
Zhongli Group	0.0537	YASHA	0.0225	GRANDLAND GROUP	0.0158
Ruihe Decoration	0.0286	Almaden	0.0447	ERA	0.0140
ZHONGZHAUNG CONSTRUCTION	0.0184	TGOOD	0.0467	JAYU	0.0168
Risen	0.0461	XiuQiang	0.0405	Haida	0.0258
WATERTEK	0.0293	JOLYWOOD	0.0319	HNAC	0.0290
Tus-Design	0.0267	HANJA DESIGN	0.0089	JINGGONG STEEL BUILDING	0.0273
SUMEC	0.0228	LONGi	0.0266	LINYANG	0.0067
MINGYANG SMART ENERGY	0.0470	JANGHO	0.0339	ARTS Group	0.0083
CENTER INT	0.0605	SUNOREN	0.0194	CLENERGY	0.0043
CHINA SOUTHERN POWER GRID	0.0172				

Substitute into formula (3.6) to calculate the sector index, and the solution result is shown in Figure 1.

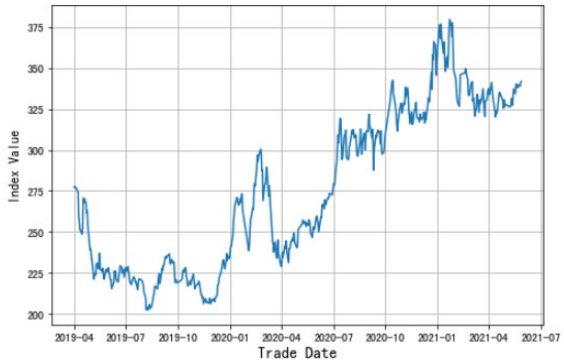

Fig 1. Photovoltaic building integration sector index

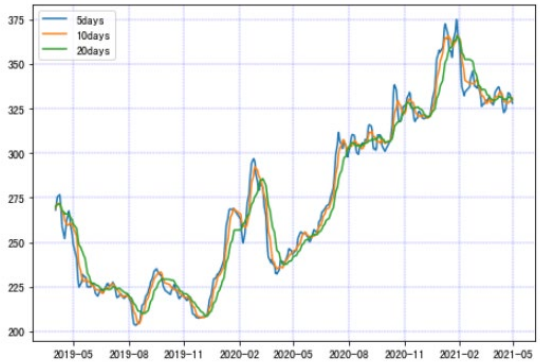

Fig 2. Moving Average of the Sector Index

(2) Feature engineering

Feature engineering plays a very important role in machine learning, and its quality determines whether machine learning can learn a satisfactory result. Generally speaking, feature engineering is divided into three parts: feature construction, feature extraction, and feature selection. In this paper, the feature construction method is adopted, that is, some features with physical meaning are manually found from the original data, and their potential data value is mined. It is necessary to use the features of the given data to obtain features that may have practical significance for the model to learn through transformation, which can improve the accuracy of the model.

Definition:

HL_PCT is the percentage change between the highest price and lowest price of the stock;

PCT_change is the percentage change between the closing price and opening price of the stock;

df['HL_PCT'] = (df['Sector Highest Price'] - df['Sector Lowest Price']) / df['Sector closing price'] * 100.0 df['PCT_change'] = (df['Sector closing price'] - df['Sector opening price']) / df['Sector opening price'] * 100.0

The defined two features of HL_PCT and PCT_change plus four features of daily turnover and daily turnover constitute the X vector dimension of GBDT model learning.

2022@ The American Computational Science Society. All rights reserved.

(3) Grid search model with cross-validation

Cross-validation is a statistical method for evaluating generalization performance, which is more stable and comprehensive than the method of dividing the training set and the test set once. In cross-validation, the data is divided multiple times and multiple models need to be trained. Grid search is a way to improve the generalization ability of the model by adjusting parameters. It can exhaustively traverse the combination possibilities of all parameters to find the best parameters.

The grid search model with cross-validation is a combination of the above two models. An ensemble function is provided in Sklearn, GridSearchCV, which implements this approach in the form of an estimator. Before use, you need to construct a combination of parameters in the form of a dictionary, and then call the function.

(4) Model solution

Through the GBDT model, the predicted data results can be obtained, as shown in Table 3. The comparison with the real data is shown in Figure 2.

Table 3: Prediction results

Predictive value	Transaction date	The real value
331.03	May 6	326.87
329.30	May 7	326.37
326.78	May 10	326.72
325.26	May 11	331.27
322.52	May 12	331.40
328.31	May 13	327.44
329.67	May 14	333.65
331.75	May 17	337.32
329.29	May 18	334.42
334.90	May 19	340.24
340.43	May 20	340.19
344.12	May 21	340.04
337.87	May 24	338.02
329.50	May25	339.73
343.65	May 26	338.94
330.21	May 27	340.60
330.08	May 28	321.15

‘learning_rate’：0.1	‘loss’：1s	‘max_depth’：3	‘n_estimators’：None

Model accuracy: 0.8636

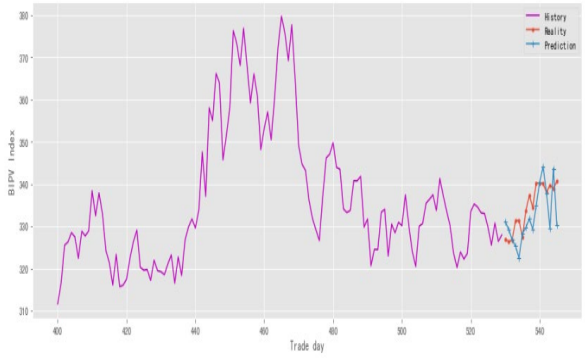

Fig3. Comparison of predicted results and real data

‘learning_rate’：0.1	‘max_depth’：4	‘n_estimators’：150

Model accuracy: 0.9737

2022@ The American Computational Science Society. All rights reserved.

The optimal parameters after grid search with cross-validation are:

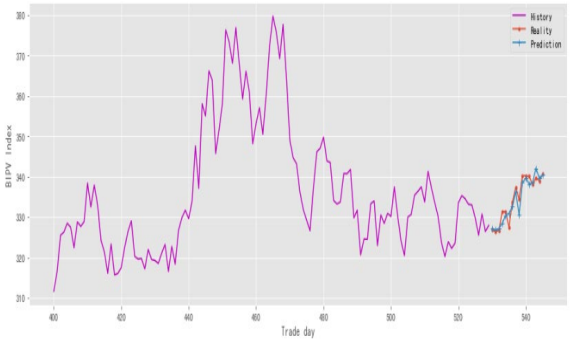

Fig 4. Results after the modified model

4.3 Correlation analysis

The correlation analysis is carried out using the obtained sector index and the Shanghai Stock Exchange index. If the two sets of data $X:\{X_1,X_2,\cdots,X_n\}$ and $Y:\{Y_1,Y_2,\cdots,Y_n\}$ are population data, then the population mean:

$$E(X)=\frac{\sum_{i=1}^{n}X_i}{n},\quad E(Y)=\frac{\sum_{i=1}^{n}Y_i}{n} \tag{4.8}$$

Population covariance:

$$Cov(X,Y)=\frac{\sum_{i=1}^{n}(X_i-E(X))(Y_i-E(Y))}{n} \tag{4.9}$$

Overall Pearson correlation coefficient:

$$\rho_{XY}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}=\frac{\sum_{i=1}^{n}\frac{(X_i-E(X))}{\sigma_X}\frac{(Y_i-E(Y))}{\sigma_Y}}{n} \tag{4.10}$$

Table 4: Correlations

Correlation	Negative	Positive
No correlation	-0.09~0.0	0.0~0.09
Weak correlation	-0.3~-0.1	0.1~0.3
Medium correlation	-0.5~-0.3	0.3~0.5
Strong correlation	-1.0~-0.5	0.5~1.0

The solution results are shown in Figure 5-6.

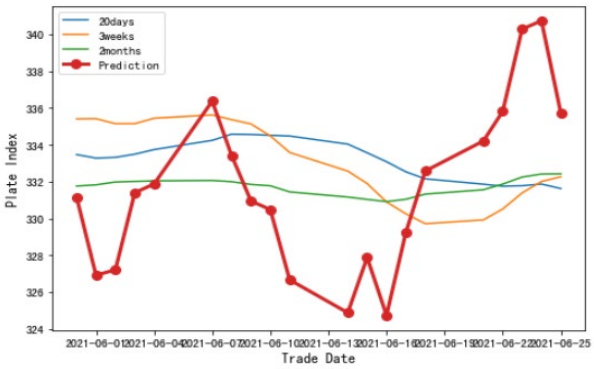

Fig5. Forecast Moving Average

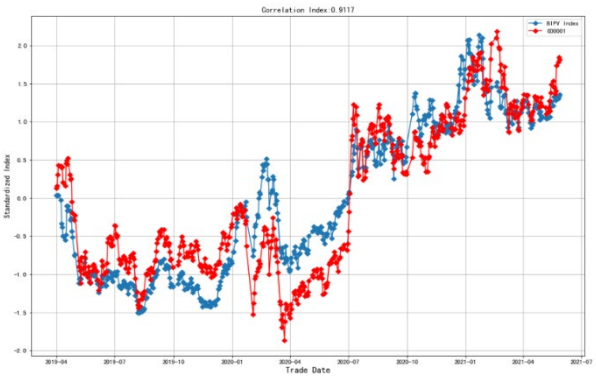

Fig6. Overall correlation analysis

It can be seen from Figure 5 and Figure 6 that the BIPV sector has been trending upward since April 1, 2019, and in recent years, due to the support of relevant policies, the development speed is relatively fast. From Figure 3-1, it can be seen that the PV building integration sector index defined in this paper basically matches the trend of the Shanghai Stock Exchange Index, and the correlation coefficient reaches 0.9177, which is a relatively satisfactory result.

2022@ The American Computational Science Society. All rights reserved.

At the beginning of 2020, there were two sharp declines in the plate. Combining with the relevant facts, it can be concluded that it should be affected by the new crown pneumonia epidemic and the Sino-US trade war, so it went out of the downward trend. The validity of the model is verified again.

4.4 Investment risk assessment

The risk of each stock can be determined by VaR value. The average value of VaR value of each stock within three months is taken as the risk index of this stock for evaluation, and the portfolio optimization model is further established.

(1) Value-at-risk assessment model

First, use three methods to calculate the VaR value of the first stock, and perform a VaR value back-test to screen out the most reasonable method to calculate the VaR value of all stocks.

①Normal distribution method

Assume that asset gains and losses follow a normal distribution. Within each confidence interval, the VaR value is calculated by multiplying one standard deviation of the return by the corresponding z-value. Because the backtest of the VaR value evaluates the value based on historical data, the current VaR value is based on the data of the past 60 trading days, excluding the data of the current day.

Advantages and disadvantages: The advantage of using the normal distribution method is that it is simple and easy to calculate, but the premise of using this method is to assume that the rate of return obeys a normal distribution.

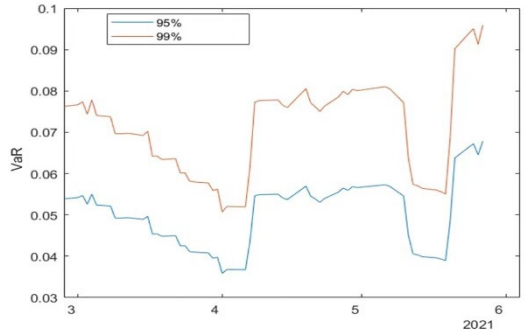

Fig 7. VaR values calculated by the normal distribution method

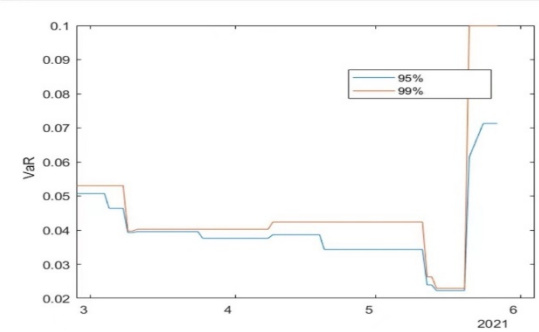

Fig 8. VaR value calculated by historical simulation method

②Historical simulation method

The historical simulation method assumes that past profits and losses can continue to be used to reflect the distribution of profits and losses in the next period. The VaR value of the current day corresponds to the value of the pth quantile of the past N historical returns.

Advantages and disadvantages: The VaR value calculated by the historical simulation method can show a relatively stable situation. This is because, over time, only when new extreme events enter the window period in which the samples are obtained can the VaR values of the quantiles change. Therefore, when the market fluctuates, the VaR value calculated by the historical simulation method cannot quickly reflect this change.

③Exponentially Weighted Moving Average (EWMA)

The first two methods assume that all past returns are equally weighted. Exponentially Weighted Moving Average (EWMA) assigns different weights to returns in an exponentially decreasing fashion. Recent returns are given a higher weight because they have a greater impact on today's returns than data from earlier periods. Corresponding to a certain window period W_E the formula for calculating the variance using the exponentially weighted average method is:

$$\hat{\sigma}_t^2 = \frac{1}{c}\sum_{i=1}^{W_E}\lambda^{i-1}y_{t-i}^2 \quad (4.11)$$

Where c is a normalized constant term:

$$c = \sum_{i=1}^{W_E}\lambda^{i-1} = \frac{1-\lambda^{W_E}}{1-\lambda} \to \frac{1}{1-\lambda} \text{当} W_E \to \infty \quad (4.12)$$

For simplicity, assuming that the window period W_E tends to infinity, the variance is approximately equal to:

$$\hat{\sigma}_t^2 \approx (1-\lambda)(y_{t-1}^2+\sum_{i=2}^{\infty}\lambda^{i-1}y_{t-1}^2)=(1-\lambda)y_{t-1}^2+\lambda\hat{\sigma}_{t-1}^2 \quad (4.13)$$

Among them, the value of the attenuation factor λ usually takes 0.

2022@ The American Computational Science Society. All rights reserved.

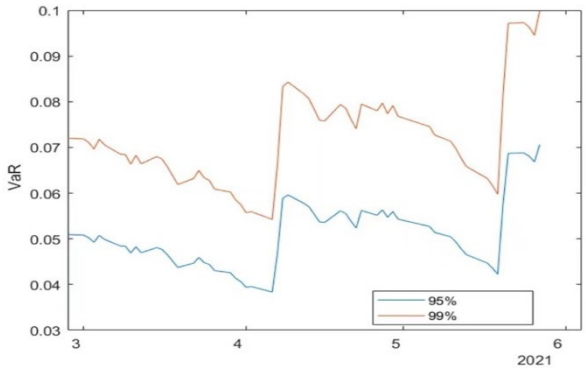

Fig9. VaR value calculated by the normal distribution method

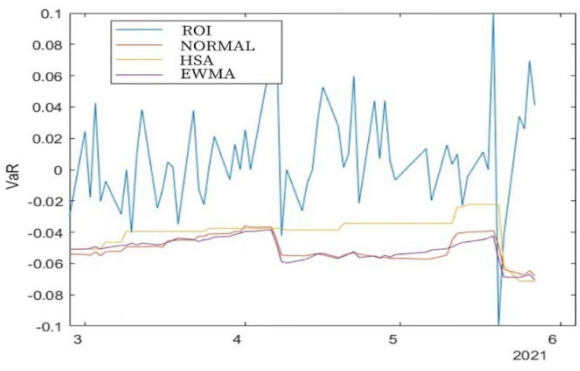

Fig10. Comparison of VaR values and rates of return for different methods

(2) VaR value backtest

The purpose of the VaR value backtest is to evaluate and compare different VaR value calculation methods. If the confidence interval is 95%, then theoretically only 5% of the day's actual losses should exceed the VaR value, and these 5% should not be clustered at a particular time. If the loss exceeds the VaR value concentrated in a certain period of time, it means that these 5% cases are not independent of each other in time, and the VaR value obtained by the commonly used calculation method should reflect the market changes slowly.

Plot the VaR value calculated by the three methods with 95% confidence interval and the rate of return over the corresponding time on the same graph for comparison:

In addition to the above visual tools, mathematical statistics are also used to backtest VaR values. Using the varbacktest provided in the MATLAB risk management toolbox, a variety of statistical tests are used for backtesting, as shown in Table 5.

Table 5: Normal distribution method backtest results

1	2	3	4	5	6	7	8	9
PortfolioID	VaRID	VaRLevel	ObservedLevel	Observations	Failures	Expected	Ratio	FirstFailure
“S&P”	“Norma195”	0.9500	0.9836	61	1	3.0500	0.3279	56
“S&P”	“Norma199”	0.9900	0.9836	61	1	0.6100	1.6393	56

As can be seen from Table 5, the actual observed interval "ObservedLevel" is very close to the confidence interval of the pre-defined VaR value. If the observed sample size is N, then at most (1-VaR_level) x N samples can be greater than the VaR value of the corresponding day at the corresponding 95% and 99% confidence intervals. From the "error rate" (Ratio) in the table, it can be seen that under the 95% confidence interval, the error rate is within the acceptable range, while under the 99% confidence interval, the VaR value calculated by the normal distribution method is not enough accurate (it underestimates the risk).

Similarly, using the historical simulation method and EWMA method to calculate the VaR value back test results are shown in Table 6 and Table 7.

Table 6: Historical simulation method backtest results

1	2	3	4	5	6	7	8	9
PortfolioID	VaRID	VaRLevel	ObservedLevel	Observations	Failures	Expected	Ratio	FirstFailure
“S&P”	“Historical95”	0.9500	0.9508	61	3	3.0500	0.9836	9
“S&P”	“Historical99”	0.9900	0.9508	61	3	0.6100	4.9180	9

Table 7: EWMA Backtest Results

1	2	3	4	5	6	7	8	9	10
PortfolioID	VaRID	VaRLevel	ObservedLevel	Observations	Failures	Expected	Ratio	FirstFailure	missing
“S&P”	“EWMA95”	0.9500	0.9836	61	1	3.0500	0.3279	56	0
“S&P”	“EWMA99”	0.9900	0.9836	61	1	0.6100	1.6393	56	0

It can be seen that the results obtained by the normal distribution method and EWMA method are better within the 95% confidence interval, because the ideal situation of normal distribution is unlikely to occur due to market reasons, the EWMA method will be used. The arithmetic mean of the VaR value calculated by the method (multiplied by 1000) is used as the standard to measure the size of each stock risk, that is, the risk value.

(3) Ranking of value at risk

2022@ The American Computational Science Society. All rights reserved.

Through the above process, the VaR ranking of 37 stocks is obtained, as shown in the table:

Table 8: Stock Value-at-Risk Ranking

Number	Securities name	Value at risk	Number	Securities name	Value at risk
34	ARTS Group	22.488986	11	YASHA	51.26015
4	Fangda Group	29.073065	3	TUNGHSU AZURE	53.39101
12	GRANDLAND GROUP	30.328149	10	Zhongli Group	53.53292
6	BAUING DECORATION	34.532633	31	LINYANG	54.94751
22	Haida	34.637858	24	JOLYWOOD	57.04679
15	ERA	35.679148	18	TGOOD	57.35585
16	ZHONGZHAUNG CONSTRUCTION	37.656043	32	MINGYANG SMART ENERGY	57.45095
5	SHENZHEN SEG	38.105868	26	Tus-Design	57.62248
13	Ruihe Decoration	38.292959	19	JAYU	57.9026
27	HANJA	39.223038	29	SUMEC	58.65028
7	Southeast Space Frame	39.747106	23	WATERTEK	60.69652
36	SUNOREN	39.819329	30	LONGi	63.6429
21	XiuQiang	40.789873	8	YANHUA SMARTECH	64.99118
33	JANGHO	41.738966	25	HNAC	66.56034
14	Almaden	42.629204	20	Risen	66.64723
28	JINGGONG STEEL BUILDING	44.500243	35	CENTER INT	81.82947
37	CLENERGY	45.477602	2	SHENZHEN ENERGY	85.75417
9	TORRAY Solar	46.816745	17	CHINA SOUTHERN POWER GRID	1.13E+02
1	CSG A	50.9178			

4.5 Portfolio Optimization

(1) Mean-variance combination model

① Portfolio expected rate of return

$$r_p = \sum_{i=1}^{n} \omega_i r_i$$
$$\mu_p = E(r_p) = E(\sum_{i=1}^{n} \omega_i r_i) \quad (4.14)$$

Where r_p: investment resistance combined yield, ω_i: the proportion of each asset, r_i: the rate of return of each asset.

② Portfolio risk

$$\sigma_p^2 = Var(E(r_p)) = Var(E(\sum_{i=1}^{n} \omega_i r_i))^2$$
$$\sigma_p^2 = \pi r^2 \sum_i \sum_j \omega_i \omega_j Cov(r_i r_j) \quad (4.15)$$

where $Cov(r_i r_j)$: covariance between assets.

③ Optimization objective

$$\min_{\omega} : \sigma_p^2 = \pi r^2 \sum_i \sum_j \omega_i \omega_j Cov(r_i r_j)$$
$$s.t. \begin{cases} E(r_p) = E(\sum_{i=1}^{n} \omega_i r_i) \\ \sum_{i=1}^{n} \omega_i = 1 \end{cases} \quad (4.16)$$

(2) Effective boundary curve

Taking the risk σ_p as the abscissa and the expected return μ_p as the ordinate, the mean-variance portfolio model can obtain a most efficient portfolio boundary curve. All points on the curve are valid portfolios.

(3) Optimization scheme

The client buys stocks with all the funds from June 1, adjusts the stock holding weight on June 10, adjusts the stock holding weight again on June 20, and sells all the stocks at the end of the month.

2022@ The American Computational Science Society. All rights reserved.

1st June		June 10		June 20	
CENTER INT	34.3%	CENTER INT	36.4%	CENTER INT	36.4%
CHINA SOUTHERN POWER GRID	45.7%	SOUTHERN POWER GRID	51.8%	SOUTHERN POWER GRID	52%
SHENZHEN ENERGY	13.6%	SHENZHEN ENERGY	11.8%	SHENZHEN ENERGY	11.6%
CSG A	6.4%				

This scheme corresponds to scheme 28, scheme 29 and scheme 30 in the following figure respectively.

However, according to the above table and the previous risk analysis of individual stocks, China South Power Grid is a new stock with great risk. It is inappropriate for this stock to occupy such a large share in the results of our model, so this stock is not considered. select the optimal investment plan among the remaining 36 stocks, and the results are as follows:

1st June		June 10		June 20	
CENTER INT	47.39%	CENTER INT	54.24%	CENTER INT	65.33%
JANGHO	8.28%	JANGHO	4.02%	SHENZHEN ENERGY	24.79%
SUMEC	7.23%	SUMEC	4.25%	CSG A	9.88%
SHENZHEN ENERGY	21.71%	SHENZHEN ENERGY	23.29%		
CSG A	15.39%	CSG A	14.2%		

The three schemes correspond to scheme 27, scheme 29 and scheme 32 in the following figure respectively.

5. Development Trend Analysis

On the basis of the environmental analysis and industry competition analysis of China's BIPV industry, this paper uses the analysis method of PEST to analyze the future development trend of China 's BIPV industry.

(1) Analysis of the political environment

In recent years, major countries in the world have formulated relevant development plans and policies to promote the development of "zero carbon buildings": EU member states require in the "Building Energy Efficiency 2010 Directive" that from December 31, 2020, all new buildings in member states will be Nearly zero energy consumption; China 's "14th Five-Year Plan" emphasizes the establishment of carbon neutrality and green ecology, and advocates more reasonable energy allocation. In the future, it will further promote the development of green energy and energy conservation and emission reduction. The realization of "carbon peak" in 2030 and "carbon neutrality" in 2060 is an important long-term goal of China 's response to global warming.

At present, the domestic PV industry has also entered the era of large-scale application of both centralized and distributed. With the advancement of PV technology, the rapid decline of cost, and the improvement of relevant policies and norms such as technology and standards, the interactive and embedded development between PV and terminal energy consumption will generate more diversified demands.

(2) Economic environment analysis

According to the National Bureau of Statistics, the total urban and rural construction area in China exceeds 60 billion square meters, of which the urban construction area exceeds 30 billion square meters, the PV installable area exceeds 3 billion square meters, and the annual newly completed area exceeds 4 billion square meters.

The Solar Energy Application Research Center of China Academy of Building Research gave a more specific calculation. Up to now, China 's existing building area can install 400 GW of PV, and the annual completed building area can install 40 GW, the potential market space amounts to billions of yuan.

According to statistics from the China PV Industry Association, “PV + building” projects account for about 80% of domestic distributed PV projects, with a total installed capacity of more than 13 GW. According to the statistics of the Prospective Industry Research Institute, as of the end of 2018, the cumulative installed capacity of the domestic BIPV market was only 1.1 GW, and the market investment scale was less than 5 billion yuan, and the industry was still in the gestation stage.

These kinds of data show that the BIPV industry has an unprecedented market effect and scale, and the total market value of related listed companies is still increasing year by year. In the next five years, China 's BIPV market will enter a period of rapid growth.

(3) Analysis of social environment

Our greenhouse gas emission is large, therefore in the energy saving reduction project development and investment in great development potential country. In addition, compared with other countries at the same latitude and period, China has inherent geographical advantages, rich solar resources to drive the development of the industry. Take Tibet as an example. The daily radiation in Tibet is as high as 7 kilowatt hours per square meter,

2022@ The American Computational Science Society. All rights reserved.

about 3 kilowatt hours higher than other regions on average, so it has huge development potential.

(4) Analysis of technical environment

BIPV is to take PV modules as building materials, integrated installation on the building, so that it becomes an integral part of the building. It can not only meet the function of PV power generation, but also has the function of wind block, rain block, heat insulation and other building peripheral protection. The technical features are as follows:

1) Diversified components and their installation methods can be customized to actively save energy to provide electricity. At the same time, it also provides functional requirements such as energy saving and heat preservation, sealing and waterproofing, isolation and noise reduction of the building envelope.

2) Thin-film PV modules have strong power generation in weak light, and the color can be adjusted through PVB film, and the light transmission is uniform, so as to meet the design requirements of architectural functions such as building lighting, sound insulation, and heat insulation while saving energy and generating electricity, and increase the aesthetics of the building.

3) The material and electrical properties of thin-film BIPV modules, different installation inclinations, and different azimuths installed at the same location all affect the power generation of the system to varying degrees. Taking the PV curtain wall with a light transmittance of 20% and vertical installation in Beijing as an example, according to the calculation, the PV curtain wall per square meter can provide users with 860kWh of electricity during the whole life cycle (25 years). In terms of building energy consumption, compared with traditional energy-saving The use of PV curtain wall can reduce the use and maintenance of thermal insulation electric sunshade shutters, aluminum alloy sunshade louvers, etc. At the same time, PV curtain wall inherits the sound insulation and heat insulation performance of LOW-E glass, saving a lot of air conditioning and refrigeration electricity costs, it is estimated that it will save about 2660kwh per square meter of building energy for customers in 25 years.

To sum up, with the country's vigorous promotion of the BIPV industry and the support of corresponding policies, the BIPV industry will usher in rapid development. Nowadays, the technological progress in various aspects is constantly advancing, and the energy conversion efficiency is gradually improved. The implementation of technology first on the basis of the introduction of various policies and the support of financial subsidies will further promote the development of the industry and continue to provide the development of China 's renewable energy industry power.

As China's BIPV industry is booming, with the guidance of policies, the rational layout of enterprises, financial support, and the continuous advancement of technology, the future of China's BIPV industry will be bright, so that the PV industry will become a top industry with international competitiveness and world influence, and provide green power for the development of China and even the world.

References

[1] S K Si, X J Sun. Python Mathematical Experiment and Modeling. Beijing: Science Press, (2020) 274.
[2] Sarah Guido, et al., Python Machine Learning Basic Tutorial. Beijing: People's Posts & Telecommunications Press, (2018)68, 205.
[3] Y T Wang, et al. Big Data Analysis and Machine Learning. Beijing: China Machine Press, (2021)208.
[4] Ariyo A A, Adewumi A O, Ayo C K. Stock price prediction using the ARIMA model[C]. 2014 UKSim-AMSS 16th international conference on computer modelling and simulation, (2014)106-112.
[5] Y Wei. Volatility forecasting models for CSI300 index futures[J]. Journal of management sciences in China, 13(2010)66-76.
[6] Hassan M R, Nath B. Stock market forecasting using hidden Markov model: a new approach[C]. The 5th International Conference on Intelligent Systems Design and Applications. Piscataway, (2005)192-196.
[7] Meesad P, Rasel R I. Predicting stock market price using support vector regression[C].2013 International Conference on Informatics, Electronics and Vision, (2013)1-6.
[8] Heaton J B, Polson N G, Witte J H. Deep learning in finance[J]. arXiv preprint arXiv:1602.0656, (2016).
[9] Q Yang, C Wang. A Study on Forecast of Global Stock Indices Based on Deep LSTM Neural Network[J]. Statistical research, 36(2029) 65-77.

2022@ The American Computational Science Society. All rights reserved.