ISSN: 2710-0170

Nonlinear Pricing Mechanism of China's Public REITs: A Study on Predicting REITs Returns Based on Explainable Machine Learning

Haibo Liao *

School of Economics and Business Administration, Yibin University, Yibin 644007, Sichuan, China

* Corresponding Author

Abstract

To address the research gap where the nonlinear pricing mechanism of China's public REITs market remains unelucidated, this study constructs an explainable machine learning (XAI) analytical framework, innovatively realizing a full-process research of "prediction-attribution-discovery". Based on the daily trading data (2021-2024) of 40 public REITs, the XGBoost algorithm is adopted to build a return prediction model. The pure long-only portfolio (ML_top20) constructed from this model achieved a cumulative return of 41.29% (Sharpe ratio = 0.12) during the out-of-sample testing period (2022-2024), significantly outperforming the market benchmark portfolio which recorded a cumulative return of -9.51% (Sharpe ratio = -0.03). Comprehensive analysis using machine learning interpretation tools reveals that stock market return, momentum effect, Treasury bond yield, and daily price range constitute the core driving factors, while feature interaction and nonlinear effects dominate the REITs price formation mechanism. The findings of this study suggest that a new generation of pricing models integrating interaction factors and nonlinear responses should be developed, providing support for the establishment of cross-market risk early warning mechanisms.

Keywords

Public REITs, Explainable Machine Learning, Return Prediction, Nonlinear Pricing, Interaction Effect.

1. Introduction

The development of publicly offered infrastructure investment trust funds (referred to as publicly offered infrastructure securities investment funds after listing on stock exchanges, hereinafter "public REITs") is a key measure for the reform of China's capital market, and has been included in the 14th Five-Year Plan and the Long-Range Objectives Through the Year 2035 of the People's Republic of China. Since the pilot launch in 2021, by July 2024, a total of 46 public REITs have been publicly listed on the Shanghai and Shenzhen Stock Exchanges, with the market's trading activity and scale expanding steadily. As an innovative financial instrument with both equity and debt attributes, the pricing mechanism of public REITs exhibits significant particularities. Its return is nonlinearly affected by various factors such as the operational efficiency of infrastructure assets, interest rate policies, and market sentiment. Uncovering its complex driving mechanism is not only a cutting-edge academic issue for improving asset pricing theory, but also a policy demand for optimizing the risk monitoring system (the "New Nine Measures for Capital Market Development" explicitly proposes to "improve the capital market risk monitoring and early warning system").

ISSN: 2710-0170

2. Domestic and Foreign Relevant Studies and Innovations of This Study

Domestic and foreign studies on REITs returns mostly rely on linear regression models in traditional research, which struggle to capture feature interactions and nonlinear effects; while studies applying machine learning are mostly limited to improving prediction accuracy, with the model's prediction mechanism remaining a "black box". To address the research gap where nonlinear interaction effects in the formation mechanism of public REITs returns are unelucidated, this study constructs an explainable machine learning (XAI) framework, innovatively realizing a full-process research of "prediction-attribution-discovery".

2.1. Studies on Influencing Factors of REITs Returns

Scholars at home and abroad have conducted extensive research on the influencing factors of REITs returns, identifying various macro and micro-level factors, but studies on the relative importance of these factors are relatively scarce. Foreign scholars have confirmed that macro-level influencing factors mainly include industrial production index, loan interest rate, stock index, etc. [1]; micro-level factors include REITs' own structure [2], momentum effect [3], and scale [4].

Domestic scholars have also studied the influencing factors of REITs returns. HUANG Shida (2015) theoretically discussed the factors affecting REITs returns, including economic growth rate, inflation rate, interest rate, real estate market prosperity, stock market return, scale factor, investor sentiment, and internal corporate characteristics [5]. DU Yanhui et al. (2022) sorted out the key risk factors affecting industrial park REITs returns [6]. MA Shichang and LI Tingting (2023) found that there is a price correlation effect between the public REITs market and the stock market [7]. YI Danhong (2024) analyzed the impact of three factors on the price changes of expressway REITs: the daily rise and fall of the expressway stock price index, the 10-year Treasury bond interest rate, and the 7-day repo rate (Dr007) [8].

2.2. Studies on Predicting REITs Returns Using Machine Learning

HABBAB & KAMPOURIDIS (2024) studied the application of machine learning algorithms in REITs return prediction. They compared the performance of five machine learning algorithms and other financial benchmark models (e.g., Holt Exponential Smoothing, TBATS, and ARIMA) in predicting the returns of 30 REITs in the United States, the United Kingdom, and Australia. The results showed that machine learning algorithms performed excellently in return prediction after risk adjustment [9].

Relevant domestic studies include GU Zongyuan (2024), who constructed a multi-agent REITs return prediction model, used machine learning methods to simulate the price formation process, and calibrated the returns of 20 REITs [10].

2.3. Main Innovations of This Study

First, methodological innovation. Aiming at the insufficient attention to the interpretability of machine learning models in REITs prediction research, this study systematically integrates cutting-edge explainable technologies, breaks through the "black box" limitation of traditional machine learning, realizes full-process interpretable analysis of "prediction-attribution-discovery", and shifts the research method of asset pricing from "statistical inference" to "mechanism analysis".

Second, theoretical expansion through model mechanism analysis. This study uses the machine learning interpretation tool SHAP to rank the importance of factors influencing REITs returns; identifies the nonlinear response mechanism of REITs returns to key features through the Accumulated Local Effect (ALE) method; and quantitatively reveals the prediction mechanism dominated by interaction effects by calculating the H-statistic.

ISSN: 2710-0170

3. Sample Data and Variable Description

The samples selected in this study are all 40 public REITs with more than 30 trading days of transaction records in the Oriental Fortune Choice Financial Database as of September 19, 2024. The sample period is from June 21, 2021 to September 19, 2024, during which no public REITs were delisted, and the data frequency is daily trading data. The data are sourced from the Oriental Fortune Choice Financial Database.

The target variable (Target) — the variable to be predicted — is the daily return rate of public REITs on each trading day. Based on the findings of existing studies and considering research objectives such as examining the relative importance of influencing factors, the following features are selected:

(1) R_1: Lag-1 return rate of REITs, as REITs returns may exhibit a momentum effect.

(2) PriceRange: Daily price range, reflecting investor sentiment.

(3) IPOM: IPO amount of public REITs (100 million yuan), which directly affects the asset expansion capacity and operational efficiency of REITs, thereby influencing returns.

(4) Divden: Dividend per unit (fen, 1 yuan = 100 fen), i.e., the recent dividend amount per fund unit.

(5) Tuofee: Custodian fee rate (%).

(6) NumofPro: Number of projects managed by the public REITs fund, a feature of the internal structure of REITs.

(7) Yuans: Whether the project is managed by the original equity holder (a feature of the internal structure of REITs); assigned a value of 1 if managed by the original equity holder, otherwise 0; determined based on the actual controller information.

(8) ThirtyY: Daily data of 30-year Treasury bond yield (%), aligned with R_1.

(9) Hs300: Daily return rate (%) of the CSI 300 Index, representing the stock market return rate.

The descriptive statistics of sample variables are shown in Table 1.

Table 1: Descriptive Statistics of Sample Variables

Variables	Obs	Mean	Std. Dev.	Min	Max
R_1（%）	17531	.019	1.237	-9.916	30
PriceRange（%）	17531	1.436	1.229	0.000	19.759
IPOM（100 million yuan）	17531	32.971	26.381	8.240	108.800
Divden（fen）	17168	6.920	3.673	1.150	15.530
Tuofee（%）	17531	.015	.010	.010	.050
NumofPro（unit）	17531	2.222	1.316	1.000	7.000
Yuans	17531	.747	.435	0.000	1.000
ThirtyY（%）	17531	2.925	.356	2.170	3.670
Hs300（%）	17531	-.001	.010	-.049	.043

To avoid data bias, except for the binary variable "Yuans" (whether managed by the original equity holder), other data were standardized to have a mean of 0 and a standard deviation of 1 before training.

ISSN: 2710-0170

4. Research Methods

4.1. Algorithm Introduction

XGBoost (eXtreme Gradient Boosting, hereinafter "XGB") is an ensemble machine learning algorithm based on decision trees. It optimizes the loss function through second-order Taylor expansion and introduces a regularization term to control model complexity, thereby achieving high efficiency and accuracy, and has achieved state-of-the-art performance in many fields (LUNDBERG, ERION, CHEN et al., 2020) [11]. For tabular datasets, tree-based models are usually superior to standard deep learning models (CHEN & GUESTRIN, 2016) [12].

The hyperparameter combination used in the XGB model of this study is as follows: learning rate (eta) = 0.13, maximum tree depth (max_depth) = 5, minimum splitting loss (gamma) = 0. The process of determining this hyperparameter combination is as follows: the XGB model was trained using data from 365 trading days from June 21, 2021 to December 16, 2022, and cross-validation and parameter grid search were applied to determine the optimal hyperparameter combination for the model. The parameter grid included learning rate (eta ∈ [0.01, 0.3], step size = 0.03), maximum tree depth (max_depth ∈ {4, 5}), and minimum splitting loss (gamma ∈ {0, 0.1}).

In each model training process, the early stopping method was adopted: if the performance of the model on the validation set did not improve for 50 consecutive iterations, the training was terminated early to prevent overfitting.

4.2. Method for Constructing Fund Portfolios

First, the XGB model was trained using data from 365 trading days from June 21, 2021 to December 16, 2022. Then, the lag-1 data was used to predict the REITs return rate of the next trading day (December 19, 2022) to avoid future information leakage. Based on the prediction results, the REITs with the top 20% performance were selected to form an equal-weighted portfolio named ML_top20; the top 30% REITs formed ML_top30; the bottom 30% REITs formed ML_bottom30; and the bottom 20% REITs formed ML_bottom20. On the next trading day, the training sample was expanded by one day, the previous training process was repeated, then the return rate of the sample REITs for the next trading day (December 20, 2022) was predicted, and the machine learning portfolios were constructed using the method of the previous day. Through the rolling training window, the daily return rate series of machine learning portfolios from December 19, 2022 to September 18, 2024 (a total of 427 trading days) was finally generated. The number of REITs in the machine learning portfolios ranged from a minimum of 23 to a maximum of 40.

In addition, an equal-weighted full-sample market portfolio (Benchmark) was constructed as a benchmark portfolio to compare and analyze the performance of the machine learning portfolios.

5. Performance of Machine Learning Portfolios

From December 19, 2022 to September 18, 2024, the return rate of the market portfolio (Benchmark) used as the benchmark was -9.51% with a Sharpe ratio of -0.03; the return rate of the ML_bottom20 portfolio was -48.76% with a Sharpe ratio of -0.19; the return rate of the ML_bottom30 portfolio was -43.21% with a Sharpe ratio of -0.17; the return rate of the ML_top30 portfolio was 35.64% with a Sharpe ratio of 0.11; and the return rate of the ML_top20 portfolio was 41.29% with a Sharpe ratio of 0.12 (see Table2).

Considering transaction costs, assuming full position adjustment every trading day with a cost of 0.005% of the net value, the return rate of the ML_top30 portfolio was 9.57% with a Sharpe ratio of 0.04; the return rate of the ML_top20 portfolio was 14.14% with a Sharpe ratio of 0.05.

ISSN: 2710-0170

The performance of the two outperforming portfolios was still far better than that of the market portfolio (assuming no position adjustment for the market portfolio, so no transaction costs, with a return rate of -9.51% and a Sharpe ratio of -0.03).

Table 2: Performance of Market Portfolio and Machine Learning Portfolios During the Sample Period

Performance Indicators	Benchmark	ML_top20	ML_top30	ML_bottom30	ML_bottom20
Return Rate	-9.51%	41.29%	35.64%	-43.21%	-48.76%
Sharpe ratio	-0.03	0.12	0.11	-0.17	-0.19
Sortino Ratio	-0.06	0.21	0.21	-0.25	-0.27

Figure 1 shows the cumulative net value changes of the market portfolio and the 4 machine learning portfolios with an initial value of 1 yuan from December 19, 2022 to September 18, 2024. It can be seen from Figure 1 that as time progresses, the cumulative net value curves of different portfolios show a significant divergence trend. The net values of the predicted outperforming portfolios (ML_top20 and ML_top30) were gradually higher than that of the market portfolio, and the net value of ML_top20 was gradually higher than that of ML_top30 among the outperforming portfolios. On September 18, 2024, the net value of the ML_top20 portfolio reached 1.413 yuan, and the net value of the ML_top30 portfolio reached 1.356 yuan. The net values of the predicted underperforming portfolios (ML_bottom30 and ML_bottom20) were gradually lower than that of the market portfolio, and the net value of ML_bottom20 was gradually lower than that of ML_bottom30 among the underperforming portfolios. On September 18, 2024, the net value of the ML_bottom30 portfolio reached 0.568 yuan, and the net value of the ML_bottom20 portfolio reached 0.516 yuan.

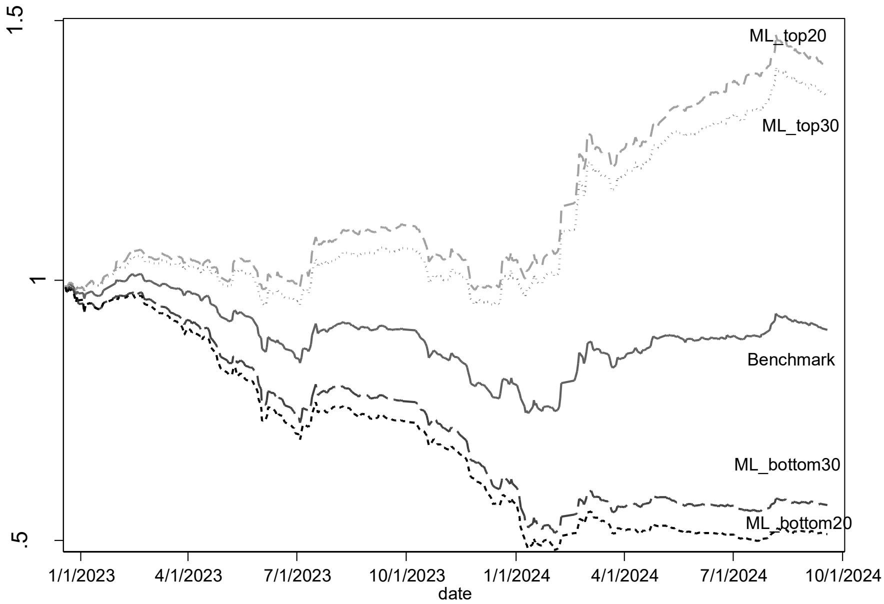

Figure 1 Cumulative Net Values of Market Portfolio and Machine Learning Portfolio

The aforementioned results indicate that the machine learning model established in this study has high prediction accuracy and can effectively distinguish between outperforming and underperforming REITs.

ISSN: 2710-0170

6. Interpretation of the Constructed Machine Learning Model

This section interprets the prediction mechanism of the previously established XGB model from three dimensions: feature importance, interaction strength, and feature marginal effects, to explore the formation law of public REITs returns.

6.1. Feature Importance: Identifying Core Driving Factors

To determine the importance of each feature, this study uses the SHAP (Shapley Additive Explanations) method proposed by LUNDBERG & LEE (2017) [13]. Figure 2 shows the absolute values of the SHAP values of each feature (completed using the R package "shapviz"). SHAP is a method derived from game theory, whose core idea is to calculate the contribution value of each feature to different prediction results, thereby assigning an importance score (i.e., SHAP value) to each feature. The SHAP method is model-agnostic and applicable to any type of data. In recent years, it has become one of the important tools for interpreting machine learning models and has been widely used (BALI et al., 2023) [14].

As can be seen from Figure 2, the key features ranked by importance are: Hs300 (CSI 300 Index return rate), R_1 (lag-1 return rate), ThirtyY (30-year Treasury bond yield), and PriceRange (daily price range).

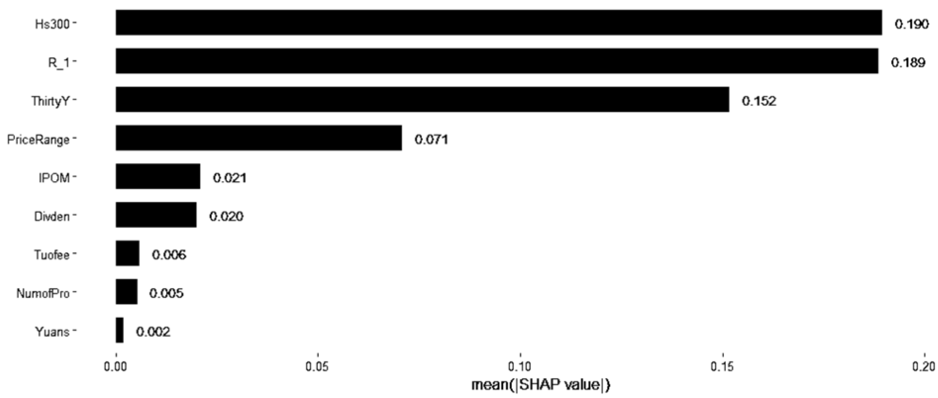

Figure 2 SHAP values of features

6.2. Interaction Effects

Studies by GU et al. (2020) [16] and DEMIGUEL et al. (2023) [17] show that the advantage of machine learning over traditional linear models in prediction accuracy lies in its ability to capture the interaction effects between features.

This section uses the H-statistic proposed by FRIEDMAN & POPESCU (2008), which is widely used, to measure the strength of interaction effects between features [17]. The total interaction strength (Total interaction strength of all variables together) index H^2 of the XGB prediction model established in this study is 0.492 (standardized value), indicating that 49.2% of the prediction variance is driven by the interaction effects between features, and interaction effects dominate the formation of REITs returns. This finding highlights the importance of considering interaction effects in modeling REITs returns.

6.3. Marginal Effects of Key Features

To accurately characterize the marginal effects of key features, this study adopts the Accumulated Local Effects (ALE) method, and the results are shown in Figure 4. Compared with the Partial Dependence Plot (PDP), ALE avoids false extrapolation caused by feature correlation by accumulating local effects under conditional distribution (APLEY & ZHU, 2020) [18], which is particularly suitable for the scenario of this study where there are significant interaction effects ($H^2 = 0.492$).

ISSN: 2710-0170

The Accumulated Local Effects (ALE) analysis shows that the response of public REITs returns to key driving factors exhibits significant nonlinear characteristics. Specifically:

(1) Asymmetry of stock market correlation: The sensitivity of REITs to stock market fluctuations increases significantly under extreme market conditions, suggesting that cross-market risk contagion may have an unbalanced mechanism.

(2) State dependence of momentum effect: The marginal effect of R_1 on the current period's return changes with its value level, increasing significantly in the medium range but decreasing in the high range, which contradicts the constant strength assumption of traditional linear models.

(3) Direction reversal of interest rate elasticity: The impact of long-term interest rates on REITs valuation has a positive-negative elasticity conversion interval.

(4) Asymmetric three-regime pattern of sentiment response: The impact of daily price range (PriceRange) on REITs returns shows a significant state dependence. In the low-volatility regime, returns are positively correlated with daily price range; in the medium-volatility regime, the impact direction reverses but is not significant; in the high-volatility regime, the negative effect increases significantly. This asymmetric pattern challenges the traditional assumption of a linear correlation between volatility and returns, suggesting that sentiment factors should be incorporated into the pricing framework in the form of regime switching.

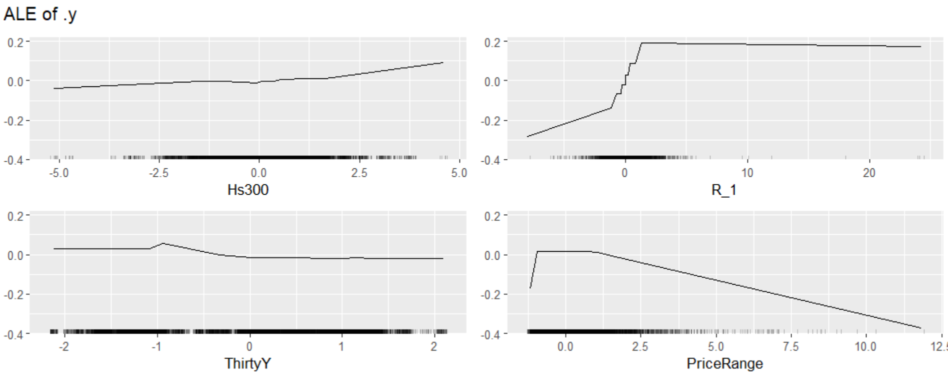

Figure 3 Accumulated Local Effects (ALE) Plot of Key Features

7. Conclusions and Discussions

The XGB model established in this study has high prediction accuracy and economic effectiveness. Even after considering transaction costs, the performance of the pure long-only portfolio ML_top20 constructed based on model predictions is still significantly better than that of the market benchmark portfolio.

7.1. Key Findings

By constructing an explainable machine learning framework, this study reveals the driving mechanism and nonlinear pricing characteristics of China's public REITs returns. The key findings are:

(1) The core driving factors of REITs returns are, in order, stock market return rate (Hs300), momentum effect (R_1), Treasury bond interest rate (ThirtyY), and daily price range (PriceRange).

(2) Interaction effects between features and nonlinear effects dominate the pricing mechanism of the model.

ISSN: 2710-0170

7.2. Theoretical Implications

The findings of this study provide the following insights for the theoretical research on asset pricing:

(1) The dominance of interaction effects and nonlinear responses indicates that REITs pricing needs to incorporate interaction factor terms and state switching mechanisms, which challenges the applicability of the linear pricing paradigm.

(2) The asymmetry of stock market linkage indicates that REITs are not only derivatives of underlying assets but also have the volatility contagion attribute of equity assets. This requires theoretical models to integrate cross-market volatility.

(3) The regime-dependent characteristics of sentiment response suggest that investor sentiment should be incorporated into the pricing framework in the form of dynamic thresholds, rather than as simple linear proxy variables.

7.3. Policy Implications

Regulatory authorities should establish a dynamic elasticity monitoring system for the REITs market, and incorporate interaction effects and nonlinear response analysis into stress tests to avoid underestimating systemic risks based on linear extrapolation.

Different from some studies that believe public REITs have low correlation with the stock market and bond market [19], this study finds that public REITs returns are strongly correlated with stock market returns and Treasury bond yields. It is suggested to establish a sound "stock-REITs-bond" cross-market volatility early warning system to implement the requirement of the "New Nine Measures for Capital Market Development" on "strengthening cross-market, cross-industry, and cross-border risk monitoring and early warning".

Through the explainable machine learning framework, this study reveals the existence evidence of the nonlinear pricing mechanism in the public REITs market, breaks through the interpretation dimension of traditional linear models, provides a new perspective for the innovation of asset pricing theory, and suggests a potential path for the development of regulatory technology. Although the research data of this study covers the period from 2021 to 2024, the long-term robustness of the conclusions remains to be observed. However, the discovery of interaction effects and asymmetric effects provides a new research idea for deepening the study of the REITs price formation mechanism.

Acknowledgements

The research was supported by the Open Fund of High-Quality Development Research Center for Modern Industries of Jialing River Basin (No. JLJLY2025-01) and Yibin Social Science Planning Research Project(No. 2024YBSKL63).

References

[1] CHEN J H, CHANG T T, HO C R, et al. Grey Relational Analysis and Neural Network Forecasting of REIT Returns, Quantitative Finance, Vol. 14 (2014) No.14, p.2033-2044.

[2] CHAN S H, CHEN J, WANG K. Are REIT IPOs Unique? The Global Evidence, The Journal of Real Estate Finance and Economics, Vol. 47 (2013) No.4, p.719-759.

[3] DERWALL J, MARQUERING W. REIT Momentum and the Performance of Real Estate Mutual Funds, Financial Analysts Journal, Vol. 65 (2009) No.5, p.24-34.

[4] LIU C H, HARTZELL D J, HOESLI M E. International Evidence on Real Estate Securities as an Inflation Hedge, Real Estate Economics, Vol. 25 (1997) No.2, p.193-221.

[5] HUANG S D. Analysis of Influencing Factors of REITs Investment Return Rate, Journal of Xinjiang University of Finance and Economics, (2015) No.4, p.5-14. (In Chinese)

ISSN: 2710-0170

[6] DU Y H, WU D, WANG S Q, et al. Analysis of Key Risks in Industrial Park REITs Returns, Construction Economy, Vol. 43 (2022) No.1, p.32-38. (In Chinese)
[7] MA S C, LI T T. Study on the Correlation Between Infrastructure Public REITs Price Index and Stock Market Price, Price: Theory & Practice, (2023) No.11, p.197-201, 216. (In Chinese)
[8] YI D H. Empirical Analysis of Influencing Factors on the Price of China's Expressway Public REITs, Investment Research, Vol. 43 (2024) No.4, p.25-36. (In Chinese)
[9] HABBAB F Z, KAMPOURIDIS M. An In-depth Investigation of Five Machine Learning Algorithms for Optimizing Mixed-Asset REITs, Expert Systems with Applications, Vol. 235 (2024), p. (to be supplemented if available).
[10] GU Z Y. Study on Market Price and Investor Structure of China's Public REITs, Management Modernization, Vol. 44 (2024) No.3, p.57-65. (In Chinese)
[11] LUNDBERG S M, ERION G, CHEN H, et al. From Local Explanations to Global Understanding with Explainable AI for Trees, Nature Machine Intelligence, Vol. 2 (2020), p.56-67.
[12] CHEN T, GUESTRIN C. XGBoost: A Scalable Tree Boosting System, Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (New York, USA, 2016), p.785-794.
[13] LUNDBERG S, LEE S. A Unified Approach to Interpreting Model Predictions, Proc. 31st International Conference on Neural Information Processing Systems (Long Beach, USA, 2017), p. (to be supplemented if available).
[14] BALI T G, BECKMEYER H, MOERKE M, et al. Option Return Predictability with Machine Learning and Big Data, Review of Financial Studies, Vol. 36 (2023) No.9, p.3548-3602.
[16] GU S, KELLY B, XIU D. Empirical Asset Pricing via Machine Learning, Review of Financial Studies, Vol. 33 (2020) No.5, p.2223-2273.
[17] DEMIGUEL V, GIL-BAZO J, NOGALES F J, et al. Machine Learning and Fund Characteristics Help to Select Mutual Funds with Positive Alpha, Journal of Financial Economics, Vol. 150 (2023) No.3, p. (to be supplemented if available).
[18] APLEY D W, ZHU J. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models, Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. (to be supplemented if available) (2020), p. (to be supplemented if available).
[19] LI Z Z, SHENG L. Study on the Return and Risk Characteristics of China's Infrastructure REITs, Macroeconomic Research, (2022) No.9, p.97-105. (In Chinese)