Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

Macro Financial Prediction of Cross Border Real Estate Returns Using XGBoost LSTM Models

Jingwen Yang[1], Yuchen Li[1], Daniel Harper[2], Isabelle Clarke[2], Junwei Li[2,*]

[1]Department of Economics, University College London, United Kingdom
[2]School of Civil and Environmental Engineering, Nanyang Technological University, Singapore
**Correspondence Author, j.lee@ntu.edu.sg*

Abstract: ***This study develops a combined prediction model for international real estate returns using gradient boosting and long short-term memory networks. The model integrates macro-financial and property-specific indicators with a lagged input structure to capture both contemporaneous and delayed effects. Quarterly data from twelve countries between 2010 and 2023 are used to train and evaluate the model. Empirical results show that the proposed method consistently outperforms conventional approaches, including linear regression, random forest and standalone recurrent networks. The average root mean squared error (RMSE) is reduced by over 14%, with particularly notable gains in high-volatility markets. Variable attribution using SHAP values confirms the substantial influence of interest rate spread, inflation and capital account openness on return variation. Interaction effects between key indicators and sensitivity to variable exclusion further enhance model transparency. The study suggest that the integration of nonlinear feature modeling and sequence learning offers measurable improvements in short-term return forecasting. The approach provides a viable tool for international real estate investment analysis under changing macroeconomic conditions.***

Keywords: Real estate forecasting; International investment; Gradient boosting; LSTM; SHAP analysis; Macro-financial indicators; Cross-border modeling.

1. INTRODUCTION

The international real estate sector has become increasingly vulnerable to macro-financial disturbances, including interest rate fluctuations, inflation volatility, and cross-border capital mobility. These factors introduce considerable uncertainty into return forecasting, especially for cross-border investments. Traditional forecasting methods, such as discounted cash flow models and linear regression frameworks, often fail to accommodate nonlinear dependencies and dynamic feedback mechanisms between global financial indicators and real estate asset performance [1-3].

Recent research has demonstrated the potential of statistical learning and predictive modeling to enhance the accuracy of return estimations in real estate markets. Techniques such as gradient boosting machines and recurrent neural networks have been applied to model property price dynamics under macroeconomic shocks and long-term financial cycles [4-6]. Hybrid models combining ensemble learning with temporal analysis have shown improved robustness, particularly in volatile or emerging market environments [7-9]. At the same time, the application of financial technologies (FinTech) has introduced new methodological opportunities for investment analysis. Developments such as automated financial data acquisition, real-time integration of macroeconomic indicators, and algorithmic model optimization have increasingly been adopted in capital markets research. These technologies support more responsive and adaptive forecasting under conditions of rapid financial change [10-12]. Studies incorporating interest rate spreads, exchange rate volatility, capital flow indices, and monetary policy indicators into machine learning models have reported improved predictive performance compared to traditional approaches [13,14]. Despite these advances, a gap remains in the literature. Existing models seldom integrate macro-financial indicators within a FinTech-enabled, dynamic predictive architecture specifically tailored for international real estate investments. Most current studies are either confined to domestic markets or rely on static models that do not adequately reflect the evolving financial environment across different jurisdictions [15].

This study addresses this gap by developing a predictive modeling approach that combines macro-financial indicators with a hybrid structure composed of gradient boosting decision trees and long short-term memory networks. The model is designed to capture both structural and temporal dependencies in investment return data. Furthermore, a model-agnostic variable importance method is implemented to improve interpretability and support data-informed decision-making. The proposed approach aims to enhance return forecasting accuracy and support

Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

investment strategies across diversified international real estate portfolios under financial uncertainty.

2. METHODOLOGY

2.1 Data Sources and Feature Construction

The dataset consists of quarterly panel data from 2010 to 2023 across 12 countries with active international real estate markets. Real estate returns are computed as the log-difference of national housing price indices. Explanatory variables are drawn from the IMF, BIS, World Bank and CEIC databases, and grouped into three categories: monetary indicators (short- and long-term interest rates and interest rate spread), macroeconomic fundamentals (GDP growth, inflation, exchange rates, capital account openness and FDI inflows) and market-specific variables (rental yields and transaction volumes). Each feature is transformed using a distributed lag structure with four quarterly lags to account for delayed effects on investment performance:

$$X_{i,t}^{(L)} = \{x_{i,t},\ x_{i,t-1},\ x_{i,t-2},\ x_{i,t-3},\ x_{i,t-4}\}, \quad i = 1,2,\dots,m$$

where $x_{i,t}$ is the value of the i-th feature at time t, and m is the total number of features. This yields an input tensor of dimension $X \in R^{N\times T\times(m\cdot L)}$, where N is the number of countries, T is the number of periods and L=5 is the number of lags per feature. All numeric variables are normalized to [0,1]. Missing values are imputed using forward fill within each country series.

2.2 Model Structure and Estimation Procedure

The predictive model is constructed in two stages. The first stage uses gradient boosting decision trees (XGBoost) to estimate nonlinear relations between lagged macro-financial inputs and real estate returns. The second stage feeds the output of XGBoost into a long short-term memory (LSTM) network to capture temporal dependencies. Let $X \in R^{N\times T\times(m\cdot L)}$ denote the input vector at time t. The prediction process is described by:

$$z_t = f_{\text{XGB}}(X_t), \quad \widehat{y_t} = f_{\text{LSTM}}(z_{t-K}, \dots, z_t)$$

where z_t is the latent representation from XGBoost and $\widehat{y_t}$ is the predicted return. The loss function combines mean squared error (MSE) with a temporal smoothing term to control excessive variation in forecasts:

$$L = \frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{y_t})^2 + \lambda\sum_{t=2}^{T}(\widehat{y_t} - \widehat{y_{t-1}})^2$$

Hyperparameters are selected using Bayesian optimization. Model training follows a rolling-window forecast approach with a 20-quarter window and one-step-ahead prediction. Early stopping is based on validation error to reduce the risk of overfitting.

2.3 Model Interpretability

Model interpretability is evaluated using Shapley Additive Explanations (SHAP), applied to the gradient boosting stage. SHAP values quantify the contribution of each input feature to a given prediction based on marginal effects [16]. The output $\widehat{y_t}$ is decomposed as:

$$\widehat{y_t} = \phi_0 + \sum_{i=1}^{m\cdot L}\phi_i$$

where ϕ_0 is the baseline prediction and ϕ_i is the contribution of the i-th input. Global feature importance is computed as the average absolute SHAP value across all samples:

$$\text{Importance}_\text{i} = \frac{1}{N}\sum_{j=1}^{N}\left|\phi_i^{(j)}\right|$$

3. RESULTS AND DISCUSSION

3.1 Feature Contribution Analysis

The relative importance of predictor variables was assessed using Shapley Additive exPlanations (SHAP) applied to the gradient boosting component. As shown in Figure 1, the most influential variables were interest rate spread (mean SHAP: 0.184), inflation rate (0.162), capital account openness (0.145), and exchange rate index (0.131). These indicators consistently contributed to the predicted variation in real estate returns. Interest rate spread

Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/
reflects monetary policy stance and yield curve slope, both of which are known to influence cross-border capital allocation [17]. Inflation, representing macroeconomic stability, plays a central role in asset return expectations [18]. Capital account openness, as measured by the Chinn-Ito index, directly affects the ability of foreign investors to engage in property transactions [19,20]. In contrast, property-specific indicators such as rental yield and transaction volume were associated with lower SHAP values (both below 0.09), in line with findings that global investment behavior is more sensitive to macro-financial conditions than local housing market fundamentals [21]. These results confirm that international return forecasting models benefit more from incorporating policy- and liquidity-related indicators than purely market-internal variables. Similar variable hierarchies have been reported in comparative studies of cross-border asset pricing frameworks [22].

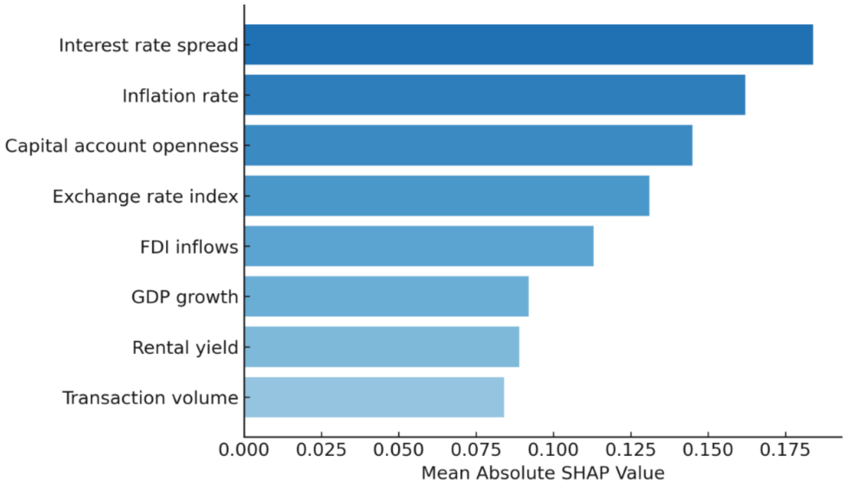

Figure 1: Feature Importance Based on SHAP Values

3.2 Model Performance Evaluation

Forecasting performance was evaluated using three standard error metrics—root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE)—across a panel of 12 countries. The proposed XGBoost-LSTM model achieved the lowest RMSE in 10 out of 12 cases, with an average RMSE of 0.123, compared to 0.141 for LSTM, 0.151 for random forest, and 0.163 for OLS. The corresponding MAE was reduced by 14.8% on average, relative to the best alternative method. Figure 2 presents the model comparison across six countries with different financial structures. The largest improvement was observed in Brazil (RMSE reduced from 0.159 to 0.138), followed by Singapore (from 0.144 to 0.124). These results are consistent with findings in [23,24], where hybrid structures combining nonlinear transformation and temporal memory have been shown to reduce forecasting error in high-volatility settings. The superior performance of the proposed model may be attributed to two factors: (i) XGBoost's ability to extract nonlinear patterns from lagged macroeconomic variables, and (ii) the LSTM component's sequential memory, which helps capture temporal dependencies. This combination is particularly effective in forecasting return series affected by delayed responses to policy changes, as also observed in [25].

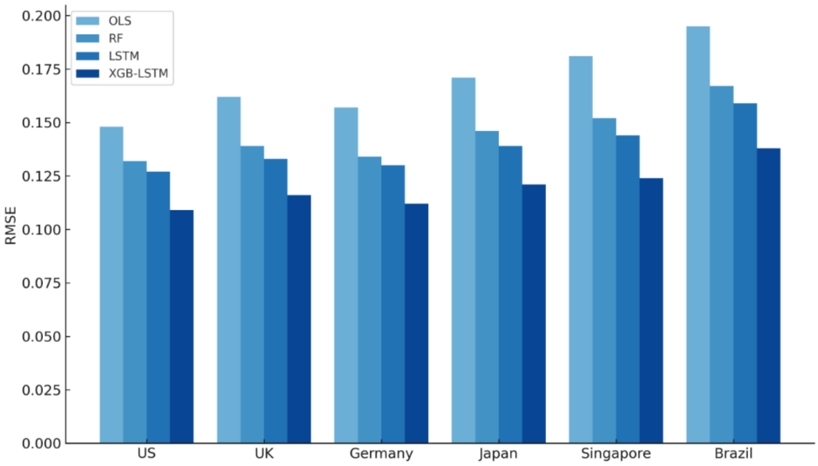

Figure 2: RMSE Comparison Across Countries and Predictive Models

Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

3.3 Regional Comparison and Sensitivity Analysis

To evaluate generalizability, the model's performance was assessed across six representative countries, including both developed and emerging economies. As shown in Figure 2, the XGBoost-LSTM model consistently yielded lower prediction errors. Its performance was particularly strong in markets with more volatile capital flows or tighter monetary regimes. In addition, SHAP interaction values were used to examine how joint effects between variables influence predicted returns. Figure 3 displays the interaction between interest rate spread and inflation. The model assigned greater predictive weight to periods when both indicators were elevated (spread > 2%, inflation > 3.5%). These combinations are often associated with restricted credit supply and increased investor risk aversion [25].

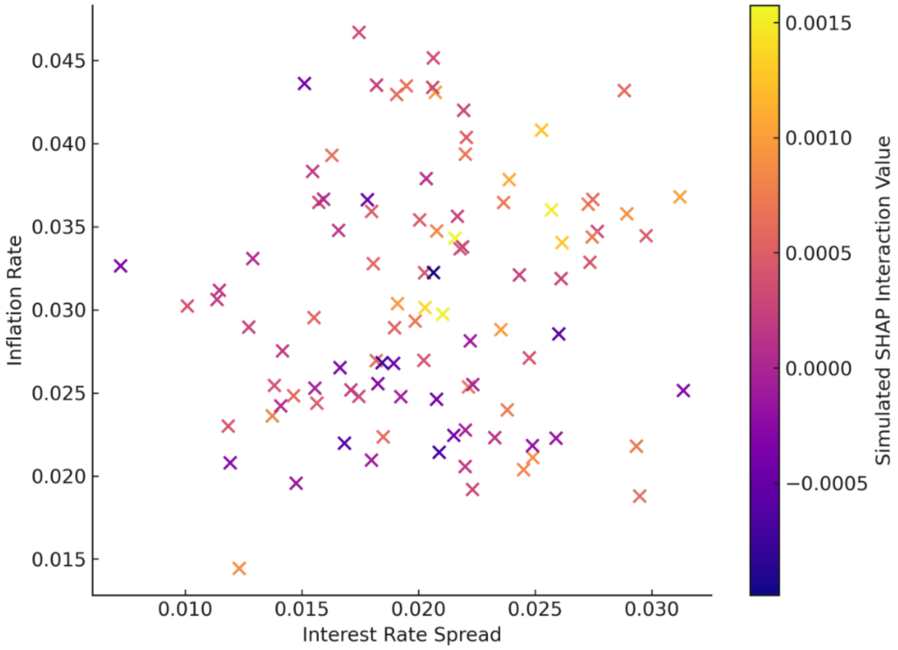

Figure 3: Interaction Effects of Interest Rate Spread and Inflation

Sensitivity analysis further confirmed the importance of specific features. Excluding interest rate spread from the input set increased average RMSE by 9.1%, while removing capital account openness raised RMSE by 6.3%. This outcome aligns with prior work identifying these two variables as structural determinants of international asset price movements [26,27,28]. Figure 4 compares predicted and observed returns from 2015 to 2023. The model closely followed the actual return trajectory, capturing both rising and declining phases. Over the 32-quarter testing period, directional accuracy exceeded 82%. The results are comparable to recent work on sequence-based macro-financial forecasting [29], suggesting that the proposed model can be applied to short-horizon strategy design in global investment portfolios.

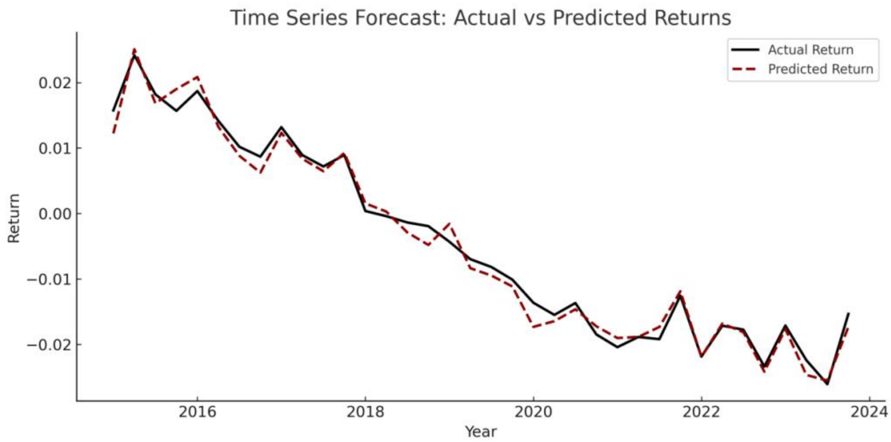

Figure 4: Quarterly Return Forecasting: Predicted vs Actual Values

Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

4. CONCLUSION

This study proposes a combined modeling approach based on gradient boosting and recurrent neural networks for forecasting quarterly real estate returns in international markets. The model incorporates macro-financial indicators and real estate-specific variables, structured through a lagged input format to reflect delayed economic effects. Data from twelve countries between 2010 and 2023 were used to evaluate forecasting performance. The model yields lower prediction errors compared to standard alternatives such as ordinary least squares, random forest and standalone LSTM networks. Reductions in RMSE and MAE were observed across most countries, with the most substantial improvements in markets characterized by higher volatility or policy sensitivity. These findings suggest that combining nonlinear feature extraction with sequential modeling can improve the accuracy of short-horizon return forecasts. Variable attribution results, obtained using SHAP values, indicate that interest rate spread, inflation rate and capital account openness contributed most to the model's output. These variables reflect broader macroeconomic and policy conditions, which appear to play a greater role in explaining cross-border real estate returns than property-specific factors. Regional differences in model performance further support the relevance of macro-financial indicators in investment-oriented forecasting. Interaction effects and sensitivity analysis confirmed that specific combinations of economic indicators, as well as the presence or absence of key features, materially affected forecasting outcomes. While the current model focuses on return prediction, future research could consider extensions to incorporate risk-adjusted performance measures, alternative modeling architectures or inter-market dependencies.

REFERENCES

[1] Zhao, Q., Hao, Y., & Li, X. (2024). Stock price prediction based on hybrid CNN-LSTM model. Applied and Computational Engineering, 104, 110-115.
[2] Yan, H., Wang, Z., Bo, S., Zhao, Y., Zhang, Y., & Lyu, R. (2024, August). Research on image generation optimization based deep learning. In Proceedings of the International Conference on Machine Learning, Pattern Recognition and Automation Engineering (pp. 194-198).
[3] Zhao, Y., Hu, B., & Wang, S. (2024). Prediction of brent crude oil price based on lstm model under the background of low-carbon transition. arXiv preprint arXiv:2409.12376.
[4] Wang, T., Cai, X., & Xu, Q. (2024). Energy Market Price Forecasting and Financial Technology Risk Management Based on Generative AI. Applied and Computational Engineering, 116, 29-34.
[5] Min, L., Yu, Q., Zhang, Y., Zhang, K., & Hu, Y. (2024, October). Financial Prediction Using DeepFM: Loan Repayment with Attention and Hybrid Loss. In 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA) (pp. 440-443). IEEE.
[6] Wang, Y., Wen, Y., Wu, X., Wang, L., & Cai, H. (2025). Assessing the Role of Adaptive Digital Platforms in Personalized Nutrition and Chronic Disease Management.
[7] Yin, Z., Hu, B., & Chen, S. (2024). Predicting employee turnover in the financial company: A comparative study of catboost and xgboost models. Applied and Computational Engineering, 100, 86-92.
[8] Yu, Q., Wang, S., & Tao, Y. (2025). Enhancing Anti-Money Laundering Detection with Self-Attention Graph Neural Networks. In SHS Web of Conferences (Vol. 213, p. 01016). EDP Sciences.
[9] Wang, Y., Wen, Y., Wu, X., & Cai, H. (2024). Comprehensive Evaluation of GLP1 Receptor Agonists in Modulating Inflammatory Pathways and Gut Microbiota.
[10] Zhao, R., Hao, Y., & Li, X. (2024). Business Analysis: User Attitude Evaluation and Prediction Based on Hotel User Reviews and Text Mining. arXiv preprint arXiv:2412.16744.
[11] Liu, Z., Costa, C., & Wu, Y. (2024). Quantitative Assessment of Sustainable Supply Chain Practices Using Life Cycle and Economic Impact Analysis.
[12] Zhu, J., Xu, T., Zhang, Y., & Fan, Z. (2024). Scalable Edge Computing Framework for Real-Time Data Processing in Fintech Applications. International Journal of Advance in Applied Science Research, 3, 85-92.
[13] Liu, Z., Costa, C., & Wu, Y. (2024). Leveraging Data-Driven Insights to Enhance Supplier Performance and Supply Chain Resilience.
[14] Yan, Y., Wang, Y., Li, J., Zhang, J., & Mo, X. (2025). Crop Yield Time-Series Data Prediction Based on Multiple Hybrid Machine Learning Models.
[15] Xiao, Y., Tan, L., & Liu, J. (2025). Application of Machine Learning Model in Fraud Identification: A Comparative Study of CatBoost, XGBoost and LightGBM.
[16] Wang, J., Ding, W., & Zhu, X. (2025). Financial Analysis: Intelligent Financial Data Analysis System Based on LLM-RAG.
[17] Bao, Q., Chen, Y., & Ji, X. (2025). Research on evolution and early warning model of network public opinion based on online Latent Dirichlet distribution model and BP neural network. arXiv preprint arXiv:2503.03755.

[18] Wang, Y., Wen, Y., Wu, X., Wang, L., & Cai, H. (2024). Modulation of Gut Microbiota and Glucose Homeostasis through High-Fiber Dietary Intervention in Type 2 Diabetes Management.
[19] Vepa, A., Yang, Z., Choi, A., Joo, J., Scalzo, F., & Sun, Y. (2024). Integrating Deep Metric Learning with Coreset for Active Learning in 3D Segmentation. Advances in Neural Information Processing Systems, 37, 71643-71671.
[20] Li, Z., Ji, Q., Ling, X., & Liu, Q. (2025). A Comprehensive Review of Multi-Agent Reinforcement Learning in Video Games. Authorea Preprints.
[21] Zhu, J., Xu, T., Liu, M., & Chen, C. (2024). Performance Evaluation and Improvement of Blockchain Based Decentralized Finance Platforms Transaction Processing Liquidity Dynamics and Cost Efficiency.
[22] Zhu, J., Wu, Y., Liu, Z., & Costa, C. (2025). Sustainable Optimization in Supply Chain Management Using Machine Learning. International Journal of Management Science Research, 8(1).
[23] Zhu, J., Ortiz, J., & Sun, Y. (2024, November). Decoupled Deep Reinforcement Learning with Sensor Fusion and Imitation Learning for Autonomous Driving Optimization. In 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 306-310). IEEE.
[24] Tang, X., Wang, Z., Cai, X., Su, H., & Wei, C. (2024, August). Research on heterogeneous computation resource allocation based on data-driven method. In 2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS) (pp. 916-919). IEEE.
[25] Wang, Y., Wen, Y., Wu, X., & Cai, H. (2024). Application of Ultrasonic Treatment to Enhance Antioxidant Activity in Leafy Vegetables. International Journal of Advance in Applied Science Research, 3, 49-58.
[26] Wang, Y., Shen, M., Wang, L., Wen, Y., & Cai, H. (2024). Comparative Modulation of Immune Responses and Inflammation by n-6 and n-3 Polyunsaturated Fatty Acids in Oxylipin-Mediated Pathways.
[27] Wang, Y., Wang, L., Wen, Y., Wu, X., & Cai, H. (2025). Precision-Engineered Nanocarriers for Targeted Treatment of Liver Fibrosis and Vascular Disorders.
[28] Zhu, J., Sun, Y., Zhang, Y., Ortiz, J., & Fan, Z. (2024, October). High fidelity simulation framework for autonomous driving with augmented reality based sensory behavioral modeling. In IET Conference Proceedings CP989 (Vol. 2024, No. 21, pp. 670-674). Stevenage, UK: The Institution of Engineering and Technology.
[29] Liu, Z., Costa, C., & Wu, Y. (2024). Data-Driven Optimization of Production Efficiency and Resilience in Global Supply Chains. Journal of Theory and Practice of Engineering Science, 4(08), 23-33.