

Macroeconomic Predictability and Machine Learning in REIT Returns: Out-of-Sample Evidence and Property-Sector Heterogeneity

Mahsa Khoshnoud

mkhoshnoud@stetson.edu

Stetson University

Research Article

Keywords: REIT returns, machine learning, macroeconomic predictors, XGBoost, out-of-sample forecasting, forecast-comparison inference, property-sector heterogeneity

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11196364/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

**Macroeconomic Predictability and Machine Learning in REIT
Returns:
Out-of-Sample Evidence and Property-Sector Heterogeneity**

September 2026

Abstract

This study examines whether nonlinear machine-learning models improve out-of-sample forecasts of U.S. Real Estate Investment Trust (REIT) returns and whether any predictive gains vary across forecast horizons and property sectors. Using 47,075 firm-month observations for 383 REIT securities over 2000–2023, we compare Gradient Boosting, Random Forest, and XGBoost with pooled OLS and a historical-mean benchmark under a strictly chronological forecasting design. Machine-learning models provide little improvement at short horizons but generate economically meaningful gains for medium-horizon cumulative returns. At the six-month horizon, Gradient Boosting and XGBoost achieve out-of-sample R^2 values of approximately 9.3%, while pooled OLS remains slightly negative. Formal forecast-comparison tests, however, do not establish statistically significant aggregate superiority over the linear or historical-mean benchmarks after multiple-testing adjustment. Further analysis shows that the nonlinear forecasting signal is concentrated in the common time-series component of REIT returns rather than in cross-sectional security selection. Predictive importance is driven primarily by real activity, labor-market conditions, policy rates, and long-term real interest rates. Results also vary substantially across property sectors, with the strongest statistically supported nonlinear predictability found among health-care REITs. Overall, the findings show that the usefulness of machine learning for REIT return forecasting depends on the horizon, regime, and sector, and that improvements in conventional forecast-performance measures should be distinguished from statistically established predictive dominance.

Keywords: REIT returns; machine learning; macroeconomic predictors; XGBoost; out-of-sample forecasting; forecast-comparison inference; property-sector heterogeneity

1. Introduction

Publicly traded real estate investment trusts (REITs) occupy a distinctive position in financial markets. Their securities trade alongside conventional equities, yet their underlying cash flows and valuations remain closely tied to property-market fundamentals, financing conditions, inflation, economic activity, and labor-market conditions. Early research documents exposure to both equity- and bond-market risks (Titman and Warga, 1986; Chan, Hendershott, and Sanders, 1990), while subsequent studies relate REIT returns to interest rates, inflation, output, and other macroeconomic variables (McCue and Kling, 1994; Chen, Hsieh, and Jordan, 1997; Chen, Hsieh, Vines, and Chiou, 1998). More recent evidence emphasizes that these relationships need not be stable, with REIT sensitivities varying across macroeconomic regimes and property sectors (Demiralay and Kılınçarslan, 2022; Connolly and Stivers, 2024).

These features make REITs a natural setting for nonlinear forecasting methods. Linear predictive regressions impose a fixed functional form, even though the implications of a given macroeconomic variable may depend on the surrounding economic state. Higher interest rates, for example, may have different implications when accompanied by strong real activity than when they reflect persistent inflation or restrictive monetary policy. Similar interactions can arise among unemployment, inflation, growth, and market conditions. Machine-learning methods can accommodate such nonlinearities without requiring them to be specified *ex ante*. Their flexibility, however, also increases the risk that apparent predictive gains reflect overfitting, model selection, or unusual features of a particular evaluation period rather than persistent out-of-sample information.

The broader asset-pricing literature shows that nonlinear methods can improve return prediction when model selection and out-of-sample evaluation are carefully disciplined (Gu, Kelly, and Xiu, 2020; Bianchi, Buechner, and Tamoni, 2020; Leippold, Wang, and Zhou, 2021; Kelly, Malamud, and Zhou, 2023). Recent REIT studies have brought these methods directly into listed real estate. Leow and Lindenthal (2025) document the forecasting usefulness and investor relevance of machine-learning models for REIT returns, while Jenett et al. (2025) use XGBoost and explainable machine-learning techniques to study time-varying drivers of REIT returns and volatility. These studies provide evidence that nonlinear methods can extract economically relevant information

from REIT data. They also sharpen the next empirical question: whether apparent machine-learning gains remain statistically defensible under strict out-of-sample evaluation, what component of REIT returns those models actually predict, and how stable the resulting evidence is across horizons, regimes, and property sectors.

This paper addresses those questions using a deliberately parsimonious macro-finance information set. The predictors include the federal funds rate, the real 10-year interest rate, median CPI, real GDP per capita, unemployment, broad equity-market returns, changes in macroeconomic variables, three lags of REIT total returns, and firm size. Pooled ordinary least squares (OLS) provides the primary linear benchmark and is compared with Gradient Boosting, Random Forest, and XGBoost. The forecasting design is explicitly chronological. Models are estimated on 2000–2015, tuned exclusively on 2016–2019, refit through 2019 with the selected hyperparameters frozen, and evaluated once on 2020–2023. Forecast observations are assigned according to target-end dates so that no future information enters the estimation or validation samples. The nonlinear models are evaluated against both OLS and a historical-mean forecast, with firm fixed effects considered as a stronger linear benchmark in robustness analysis.

The first finding is that forecastability is strongly horizon dependent. At the one-, two-, and three-month direct horizons, every model produces negative out-of-sample R^2 . The results change at cumulative horizons. Gradient Boosting reaches an out-of-sample R^2 of 5.6% at three months compared with -4.0% for pooled OLS, while Gradient Boosting and XGBoost reach approximately 9.3% at six months and pooled OLS remains slightly negative. RMSE and MAE show corresponding improvements. Judged only by conventional performance metrics, the six-month results would suggest economically meaningful machine-learning gains.

Formal forecast-comparison inference, however, yields a more qualified conclusion. Squared-error loss differentials are averaged across REITs within forecast target-end months and evaluated using Diebold–Mariano-type statistics with HAC adjustment for overlapping cumulative-return windows and the Harvey–Leybourne–Newbold small-sample correction. The resulting comparisons are then adjusted using the Holm procedure across the aggregate forecast-comparison family. Under this inferential standard, none of the nonlinear models establishes statistically significant aggregate superiority over pooled OLS or the historical mean. Thus, sizeable

improvements in out-of-sample R^2 , RMSE, and MAE do not necessarily imply statistically established predictive dominance. This distinction is central to the paper.

The second finding concerns the economic nature of the predictability captured by the nonlinear models. SHAP evidence for the six-month XGBoost specification assigns the greatest predictive importance to real activity, labor-market conditions, policy rates, changes in policy rates, and long-term real interest rates rather than to simple return persistence or firm size. More importantly, the signal is concentrated in the common time-series component of REIT returns. The monthly cross-sectional average of realized six-month returns has a correlation of 0.530 with the corresponding XGBoost forecast, and a Mincer–Zarnowitz regression yields a slope of 2.52, although predicted returns are substantially smoother than realizations. By contrast, the average monthly cross-sectional Spearman correlation between predicted and realized individual-REIT returns is approximately zero, and the subsequent return spread between the highest and lowest predicted quintiles is economically small and statistically insignificant. The machine-learning gains therefore reflect aggregate timing information rather than cross-sectional security-selection ability.

The third finding is that nonlinear predictability is conditional rather than uniform. An earlier pseudo-out-of-sample exercise, using a 2014–2016 holdout, preserves substantial forecast-error improvements for nonlinear models relative to pooled OLS but does not reproduce positive out-of-sample R^2 relative to the historical mean. The relative advantage of nonlinear specifications over the pooled linear model is therefore more persistent than their absolute advantage over the unconditional benchmark. Property-sector evidence reinforces this conclusion. Using a pre-specified six-month XGBoost specification rather than selecting the best-performing model after observing sector-level test outcomes, health-care REITs provide the strongest evidence of nonlinear predictability: out-of-sample R^2 reaches 11.5%, and the improvement relative to the historical mean remains statistically significant after adjustment across sector comparisons. Retail REITs exhibit sizeable positive point estimates but do not reject equal predictive accuracy in formal forecast-loss tests, while diversified and residential REITs favor OLS.

These findings contribute to the emerging REIT machine-learning literature in three respects. First, the paper shifts the focus from whether nonlinear models can generate positive out-of-sample fit to whether those gains survive formal forecast-comparison inference and multiple-testing

adjustment. This distinction matters because the conclusion implied by conventional out-of-sample performance metrics is materially stronger than the conclusion supported by corrected loss-differential tests. Second, the paper identifies the component of REIT returns being forecast. The evidence points to a smooth, macro-sensitive common return component rather than cross-sectional stock-selection ability. Third, the paper documents explicit boundary conditions for machine-learning usefulness: predictive gains emerge primarily at medium cumulative horizons, vary across macroeconomic regimes, and differ substantially across property sectors.

These contributions are complementary to, but distinct from, the closest recent REIT machine-learning studies. Leow and Lindenthal (2025) emphasize the forecasting usefulness and economic relevance of machine learning for REIT investors, whereas Jenett et al. (2025) use explainable machine learning to characterize nonlinear and time-varying drivers of REIT returns and volatility. The present study instead focuses on whether nonlinear forecast improvements remain statistically defensible after strict chronological model selection, formal forecast-loss comparison, multiple-testing adjustment, stronger linear benchmarks, temporal holdouts, and sector-level inference. The objective is therefore not to establish that machine learning can be applied to REITs, but to determine when its apparent predictive advantage survives stricter statistical and economic scrutiny.

The resulting interpretation is narrower than a claim of universal machine-learning superiority. Nonlinear models produce economically meaningful improvements in medium-horizon forecast-performance metrics and stronger evidence in selected property sectors. At the aggregate level, however, those improvements are not statistically distinguishable from benchmark forecasts after multiple-comparison adjustment and do not translate into cross-sectional security-selection ability. The evidence therefore suggests that machine learning can reveal useful predictive structure in listed real estate, but that the strength of this conclusion depends critically on the forecast horizon, benchmark, return component, economic regime, property sector, and inferential criterion used to define predictive success.

2. Related Literature

2.1 REIT returns and macroeconomic conditions

The empirical studies of REIT returns are rooted in the broader asset-pricing literature. The Capital Asset Pricing Model links expected returns to systematic market risk (Sharpe, 1964), while subsequent work demonstrates that a single market beta does not fully describe expected-return variation (Fama and MacBeth, 1973; Fama and French, 1992). Ross's (1976) Arbitrage Pricing Theory provides a direct theoretical foundation for allowing multiple systematic factors to affect returns, and Chen, Roll, and Ross (1986) show that innovations in industrial production, inflation, the term structure, and credit conditions can be priced in equity markets. Fama and French (1989) similarly document variation in expected stock and bond returns over the business cycle, while Fama and Schwert (1977) show that inflation is closely related to asset returns. These findings provide the economic motivation for using macroeconomic variables as predictors of a hybrid asset class such as REITs.

Early REIT studies establish that listed real estate is exposed to several sources of systematic risk. Titman and Warga (1986) show that measured REIT risk and performance depend on the asset-pricing specification, while Chan, Hendershott, and Sanders (1990) document exposure to stock-, bond-, and real-estate-related factors. Chen and Tzang (1988) specifically examine interest-rate sensitivity, reinforcing the view that financing conditions are central to REIT valuation. This evidence is consistent with the economic structure of REITs: their securities trade in equity markets, but their cash flows and discount rates are also shaped by leverage, property-market fundamentals, and monetary conditions.

Direct macroeconomic evidence largely supports this multifactor view. McCue and Kling (1994) examine REIT returns in relation to prices, nominal interest rates, output, and investment and find a substantial role for interest rates. Chen, Hsieh, and Jordan (1997) compare observable macroeconomic variables with statistically derived factors and find that macro variables contain meaningful explanatory information for REIT returns. However, the evidence is not uniform across specifications. Chen, Hsieh, Vines, and Chiou (1998) show that the significance of several macroeconomic variables weakens after firm characteristics such as size and book-to-market are introduced, suggesting that REIT returns reflect both aggregate conditions and security-specific attributes.

Return predictability is likewise mixed. Liu and Mei (1992) document predictability in equity REIT returns using variables such as capitalization rates, dividend yields, and Treasury bill yields and show substantial co-movement with other assets. Li and Wang (1995), however, find that once time-varying risk premia and market integration are taken into account, REIT returns are not necessarily more predictable than conventional equities. Gyourko and Keim (1992) further highlight the information contained in public real estate securities for understanding real estate returns, while Clayton and MacKinnon (2001) show that the relation between REITs, direct real estate, and other financial assets changes over time.

More recent work emphasizes state dependence. Demiralay and Kılınçarslan (2022) examine sector-specific REITs in a regime-switching setting and show that uncertainty effects vary across regimes. Connolly and Stivers (2024) document substantial variation in the relation between public real estate returns and inflation shocks, reinforcing the idea that REITs are not characterized by a single stable inflation exposure. Nayar, Price, and Shen (2023) show that the predictive importance of macroeconomic uncertainty differs between liquid public property portfolios and less liquid direct real estate. International evidence also points to meaningful macroeconomic linkages, although the strength and direction of those relationships depend on institutional and market structure (Olanrele et al., 2020; Cohen, Burinskas, and Fateye, 2026). Collectively, these studies motivate a forecasting framework that allows macroeconomic effects to vary across horizons and states rather than imposing a single constant linear response.

Related REIT research also documents changing risk, inflation, and interest-rate sensitivities over time (Park, Mullineaux, and Chew, 1990; Liang, Prudential, and Webb, 1995; Mueller and Pauley, 1995; Chandrashekar, 1999; Simpson, Ramchander, and Webb, 2007; Rosa, 2024). Together with the more recent regime-based evidence, these studies reinforce the possibility that a single constant linear response may be too restrictive for forecasting across materially different macroeconomic states.

2.2 Machine learning and return prediction

The modern machine-learning literature approaches return predictability from a different direction. Rather than imposing a fixed functional form, machine-learning methods allow data-driven nonlinearities and interactions, with regularization and validation used to control model complexity. Mullainathan and Spiess (2017) emphasize that the principal advantage of machine

learning in applied economics is predictive rather than causal. This distinction is important for the present setting: the objective is to determine whether macroeconomic and market information improves forecasts of REIT returns and whether nonlinear models reveal stable predictive structure, not to interpret every model contribution as a causal effect.

Gu, Kelly, and Xiu (2020) provide a central benchmark for empirical asset pricing. Comparing linear models with penalized regressions, trees, random forests, gradient boosting, and neural networks, they show that nonlinear methods can materially improve equity-return prediction. The gains are associated in part with interactions among predictors that conventional specifications do not capture easily. Kelly, Malamud, and Zhou (2023) similarly emphasize the value of model complexity in return prediction. Evidence from European, Chinese, emerging, and global equity markets indicates that these gains are not limited to the U.S. market, although performance is highly sensitive to the forecasting design and the characteristics of the sample (Drobetz and Otto, 2021; Leippold, Wang, and Zhou, 2021; Hanauer and Kalsbach, 2023; Azevedo, Kaiser, and Mueller, 2023; Choi, Jiang, and Zhang, 2025). Related work also emphasizes the difficulty of stock-return forecasting and the potential gains from flexible methods in other return settings (Bali et al., 2023; Rapach and Zhou, 2023).

Other asset classes provide additional evidence that macroeconomic information can interact with nonlinear models in economically meaningful ways. Bianchi, Buechner, and Tamoni (2020) show that machine-learning methods improve bond risk-premium forecasts and that the importance of economic predictors differs across maturities. This finding is relevant for REITs because interest-rate exposure and the effect of macroeconomic fundamentals may also vary with the return horizon. At the same time, the literature emphasizes that flexible models are vulnerable to overfitting. Avramov, Cheng, and Metzker (2022) show that statistical improvements in return forecasts need not translate one-for-one into economically profitable strategies after accounting for trading frictions and economic restrictions. Mekelburg and Strauss (2024) show that pooling and winsorization can improve forecast stability in high-dimensional settings, while Rasekhschaffe and Jones (2019) stress the importance of disciplined validation in financial applications. These findings support the use of a transparent OLS benchmark and careful model comparison when applying machine learning to REIT returns.

2.3 Machine learning in REIT research and explainability

The direct application of modern machine learning to REITs is comparatively recent. Leow and Lindenthal (2025) show that machine-learning methods can improve REIT return forecasts relative to conventional linear approaches and can generate economically meaningful gains for REIT investors. Their evidence also suggests that predictability varies across periods and property sectors. Jenett et al. (2025) examine the time-varying drivers of REIT returns and volatility using XGBoost together with explainable machine-learning techniques. They find that the relative importance of predictors changes over time and that macroeconomic shocks can alter the relationships among return and risk drivers. This evidence is consistent with the broader macro-REIT literature, but it also demonstrates the value of models that can accommodate nonlinearity and changing predictor importance.

Explainability is especially important in this setting because a more accurate forecast is of limited economic value if the source of the prediction cannot be understood. SHAP-based methods decompose predictions into variable-level contributions (Lundberg and Lee, 2017; Lundberg et al., 2020) and can therefore help identify whether a nonlinear model relies primarily on broad equity-market conditions, macroeconomic fundamentals, or lagged return dynamics. Related finance research emphasizes both the promise and the limitations of such tools (Leung et al., 2021; Breitung, 2023). The interpretation remains predictive rather than causal, but it provides a useful bridge between the transparency of linear regressions and the flexibility of nonlinear forecasting models.

The present design differs from the two closest REIT machine-learning studies in emphasis. Leow and Lindenthal (2025) focus on the forecasting usefulness and investor relevance of machine learning across REIT settings, whereas Jenett et al. (2025) use XGBoost and explainability tools to characterize time-varying drivers of REIT returns and volatility. This paper instead isolates a deliberately parsimonious macro-finance information set, compares forecastability across direct and cumulative horizons under a strict chronological holdout, and evaluates whether nonlinear gains survive formal forecast tests, stronger linear benchmarks, and property-sector inference. The contribution is therefore not to claim that machine learning is new to REIT research, but to identify when macroeconomic nonlinearities add out-of-sample predictive value and where that evidence is statistically most credible.

3. Data

The sample consists of U.S. listed REITs observed monthly from January 2000 through December 2023. REIT-level returns and security characteristics are obtained from CRSP. The primary return measures are total return (RET) and return excluding distributions (RETX). Market capitalization is measured as the absolute value of price times shares outstanding and is expressed in millions of dollars.

The raw CRSP/FRED panel contains 47,405 records for 383 PERMNOs. Because CRSP distribution events can generate multiple records for the same security-month, the forecasting panel is cleaned to one observation per PERMNO-month, leaving 47,075 unique firm-month observations. The 330 removed records are duplicate security-month observations generated by CRSP distribution-event reporting, in which the same PERMNO-month can appear more than once around distribution-related events; only one observation per security-month is retained before constructing lags and forward targets. Return lags and macroeconomic first differences are reconstructed after this cleanup rather than inherited from event-level records. Forward return targets are formed only when consecutive calendar months are available, which prevents accidental horizon compression around duplicate or missing records.

Market-wide equity conditions are measured with the CRSP value-weighted return excluding distributions (VWRETX); the equal-weighted return (EWRETX) is used as a robustness control. Macroeconomic variables are obtained from FRED and include the effective federal funds rate, unemployment, median CPI, real GDP per capita, and the real 10-year interest rate. Both levels and first differences are included so that the model can distinguish the prevailing macroeconomic state from changes in that state. The real 10-year series is available for 278 of the 288 sample months. No macroeconomic observations are imputed; estimation and forecast evaluation use complete cases after the panel merge, so months lacking the real-rate observation do not enter specifications that require that predictor.

Table 1 reports descriptive statistics. Individual REIT returns are substantially more volatile than diversified market-index returns, and market capitalization is strongly right-skewed. The sample also spans materially different monetary, inflation, labor-market, and growth regimes, including the Global Financial Crisis, the low-rate post-crisis period, the COVID-19 shock, and the subsequent tightening cycle.

[Table 1]

Table 2 reports pairwise correlations. RETX is positively correlated with both CRSP market indices, while correlations with lagged REIT returns are small. The two market indices are highly correlated, and the federal funds rate is also strongly correlated with the real 10-year rate. For this reason, the main linear specification includes VWRETX rather than both market indices simultaneously, and alternative market controls are examined separately. The correlation structure also reinforces the value of comparing a transparent linear benchmark with models that allow interactions and nonlinear conditional responses.

[Table 2]

4. Empirical Methodology and Main Results

4.1 Forecast targets and linear benchmark

The forecasting exercise considers five dependent variables: one-, two-, and three-month-ahead RETX, and cumulative three- and six-month RETX. Cumulative returns are compounded from strictly forward monthly observations. The common predictor set contains the federal funds rate, median CPI, the real 10-year rate, real GDP per capita, unemployment, VWRETX, three lags of CRSP total return, first differences of the five macroeconomic variables, and log market capitalization.

The OLS benchmark is a pooled predictive regression of the future REIT return on the current information set. Inference uses two-way clustered standard errors by REIT and calendar month to accommodate serial dependence within securities and common shocks across REITs. The linear regressions are used as a transparent benchmark rather than as a structural causal model; coefficients are interpreted as conditional predictive associations.

[Table 3]

Table 3 shows that linear explanatory power increases with the forecast horizon. Short-horizon adjusted R-squared values are approximately 2%, while the three- and six-month cumulative specifications reach roughly 4% and 8%, respectively. Several macroeconomic coefficients are horizon dependent. Median CPI is negative and statistically significant for

cumulative horizons, unemployment becomes positive and significant at six months, and changes in the federal funds rate and real GDP per capita also become more informative at longer horizons. Lagged returns provide limited evidence of persistence, while log market capitalization is negative for cumulative returns.

4.2 Machine-learning models and chronological validation

The nonlinear models are Gradient Boosting, Random Forest, and XGBoost. Each model receives the same predictor set as OLS. Hyperparameters are chosen exclusively by validation-period RMSE using a deliberately limited candidate grid, reported in Appendix Table A1. Tree models are estimated in the original predictor units; no full-sample scaling transformation is applied.

The forecasting design preserves time order. The estimation sample covers 2000–2015, the validation sample covers 2016–2019, and the final test period covers 2020–2023. Hyperparameters are selected once on the validation sample and then frozen. Each selected model is refit on the combined 2000–2019 data without any additional tuning and is evaluated once on the 2020–2023 test sample. Forecast windows are assigned by their target end dates. The primary performance measure is out-of-sample R^2 relative to the historical mean estimated from pre-test observations, $R^2_OOS = 1 - SSE_model / SSE_historical\ mean$; RMSE and MAE provide complementary scale-dependent measures.

A multilayer perceptron was examined in preliminary work but is not treated as a headline model because its apparent advantage did not survive the chronological train-validation-test design. The main analysis therefore focuses on the three tree-based models and OLS.

[Table 4]

Table 4 reveals a pronounced difference between short- and medium-horizon predictability. At the one-, two-, and three-month direct horizons, none of the models produces positive out-of-sample R^2 . Random Forest performs best at one month (−5.1%), OLS is closest to the historical mean at two months (−0.5%), and XGBoost performs best at three months (−6.7%), but all remain below the unconditional benchmark.

The results change for cumulative returns. At the three-month cumulative horizon, Gradient Boosting produces an out-of-sample R^2 of 5.6% compared with −4.0% for OLS, while XGBoost

reaches 3.2%. Gradient Boosting also reduces RMSE from 0.2230 to 0.2125 and MAE from 0.1518 to 0.1455.

The strongest point estimates occur at six months. OLS remains slightly below the historical-mean benchmark ($R^2_{\text{OOS}} = -0.6\%$), whereas Gradient Boosting and XGBoost both reach approximately 9.3% and Random Forest reaches 6.5%. XGBoost records RMSE of 0.2930 and MAE of 0.2055, compared with 0.3087 and 0.2188 for OLS. These magnitudes indicate material forecast improvement in out-of-sample R^2 , RMSE, and MAE terms relative to the historical mean, but they do not by themselves establish statistical dominance over OLS; that issue is addressed in Table 5.

The pattern therefore supports a narrower conclusion than universal machine-learning superiority: nonlinear forecasting gains are concentrated in medium-horizon cumulative REIT returns. One economic interpretation is that slowly evolving monetary, labor-market, inflation, and real-activity information is more informative over horizons long enough to affect financing costs, leasing conditions, cash flows, and discount-rate expectations.

4.3 Statistical comparison of forecast accuracy

To assess whether differences in squared forecast errors are statistically distinguishable, squared-error loss differentials are averaged across REITs within each forecast target-end month. Diebold–Mariano-type statistics are then computed on the 48-month loss-differential series with HAC adjustment of $h-1$ lags for overlapping cumulative targets and the Harvey–Leybourne–Newbold small-sample correction. Positive loss improvements indicate smaller squared errors for the machine-learning model.

[Table 5]

Table 5 qualifies the forecast-error improvements reported in Table 4. None of the three-month cumulative comparisons is statistically significant. At six months, all three nonlinear models reduce average squared forecast loss relative to OLS, but the corresponding unadjusted two-sided p-values are 0.494 for Gradient Boosting, 0.576 for Random Forest, and 0.368 for XGBoost. Relative to the historical mean, the smallest unadjusted p-value is 0.063 for XGBoost. However, this comparison does not survive Holm adjustment across the 12 forecast-comparison tests in Table 5; its adjusted p-value is 0.756, and none of the aggregate comparisons remains

statistically significant after familywise correction. The aggregate results therefore provide evidence of material forecast-error reductions at medium horizons, but not conventional statistical evidence of model superiority.

The absence of conventional statistical significance should also be interpreted in light of the limited time-series dimension of the test sample. Each forecast-comparison test is based on only 48 target-end months, and the six-month cumulative-return comparisons require HAC adjustment with five lags because adjacent forecast windows overlap. Consequently, the effective amount of independent time-series information is substantially smaller than the nominal 48 observations, reducing power against moderate differences in predictive accuracy. Failure to reject equal predictive accuracy therefore should not be interpreted as evidence that the models have identical population forecast performance. For this reason, the paper evaluates the DM evidence jointly with the magnitude of out-of-sample R^2 , RMSE and MAE differences and, where applicable, bootstrap uncertainty.

The empirical analysis distinguishes between primary inferential comparisons and exploratory robustness exercises. The main aggregate forecast-comparison family is reported in Table 5 and is adjusted jointly using the Holm procedure, while the confirmatory sector-level XGBoost tests in Table 13 form a separate pre-specified family that is likewise adjusted across sectors. The remaining analyses—including alternative return definitions, market controls, subperiods, firm fixed effects, and the pre-2020 pseudo-out-of-sample exercise—are interpreted as robustness and diagnostic evidence rather than additional confirmatory hypothesis tests. Accordingly, statistical significance in those exercises is not used to redefine the paper’s primary claims.

4.4 Economic interpretation of the six-month forecasts

[Table 6]

Table 6 reports global SHAP and gain-based importance for the six-month XGBoost model. The importance ranking is dominated by macroeconomic variables. Real GDP per capita accounts for 27.6% of total absolute SHAP importance, followed by unemployment at 14.6%, the federal funds rate at 13.2%, the change in the federal funds rate at 10.0%, VWRETX at 9.3%, and the real 10-year rate at 8.4%. Lagged returns and firm size contribute comparatively little.

The result is important for the economic interpretation of the forecasts. The model is not simply reproducing broad equity-market movements or short-run return persistence. Instead, medium-horizon predictions combine information about real activity, labor-market conditions, monetary policy, and long-term real discount rates. The difference between SHAP and tree-gain rankings also cautions against relying on a single feature-importance metric.

SHAP values measure contributions to fitted predictions rather than causal effects (Lundberg and Lee, 2017; Lundberg et al., 2020). Because the macroeconomic variables are jointly determined and correlated with the business cycle, the analysis is interpreted as evidence about predictive structure within the observed state space rather than structural comparative statics.

[Figure 1]

Figure 1 plots SHAP dependence relationships for real GDP per capita, unemployment, the federal funds rate, and the real 10-year interest rate. The fitted model assigns different forecast contributions across regions of the observed macroeconomic state space. Unemployment contributions become more positive at unusually high levels, while higher policy and long-term real rates are associated with more negative forecast contributions.

These patterns help reconcile the weak or horizon-dependent OLS coefficients with the stronger six-month tree-based forecasts, but they should not be interpreted as stable structural thresholds. The SHAP dependence plots are evaluated over 2020–2023, an unusual sequence spanning pandemic disruption, near-zero rates, elevated inflation, and rapid monetary tightening. The apparent nonlinearities therefore describe how the fitted model behaved within that test-period state space and may not generalize unchanged to other macroeconomic regimes.

Apparent kinks or threshold-like features in the SHAP dependence plots may also reflect tree-partition boundaries in relatively sparse regions of the predictor space rather than economically structural breakpoints, and should therefore not be interpreted as evidence of discrete regime thresholds.

A related limitation arises from the extrapolation properties of tree-based models. Because decision-tree ensembles partition the predictor space observed during estimation rather than extrapolate smoothly beyond its support, predictions at policy-rate states near or outside those well represented in the pre-2020 training sample should be interpreted cautiously. The unusually large

movement in the federal funds rate during 2020–2023 therefore further limits the external validity of the SHAP dependence patterns at the boundary of the training distribution.

5. Robustness Analysis

5.1 Total returns including distributions

Because REITs are income-producing securities, the main RETX results are repeated using CRSP total return (RET), which includes distributions. Three- and six-month targets are compounded from consecutive forward monthly RET observations using the same chronological estimation, validation, and test design.

[Table 7]

Table 7 shows that the medium-horizon conclusion is robust to the return definition. At three months, Random Forest produces the strongest positive out-of-sample R-squared, while OLS remains negative. At six months, XGBoost and Gradient Boosting both reach approximately 9% out-of-sample R-squared and reduce RMSE and MAE relative to OLS. The similarity to the RETX results indicates that the forecasting evidence is not an artifact of excluding distributions.

5.2 Subperiod robustness

[Table 8]

Table 8 evaluates the frozen six-month RETX forecasts across alternative portions of the test period using forecast target-end dates consistently. Over the full 2020–2023 test period, Gradient Boosting and XGBoost produce out-of-sample R^2 values of 9.3%. Excluding target-end months from March through December 2020, both models remain positive at approximately 5.5%. During 2020–2021 target-end months, Gradient Boosting reaches 14.8% and XGBoost 10.2%; during 2022–2023, XGBoost remains positive at 7.3% while Gradient Boosting and Random Forest turn slightly negative. The medium-horizon result therefore is not confined to the most extreme pandemic months, although model rankings remain regime dependent.

5.3 Alternative market controls

[Table 9]

Table 9 replaces the value-weighted CRSP market return with the equal-weighted return and, separately, removes the aggregate market-return control while retaining the frozen six-month model settings. Under the EWRETX specification, out-of-sample R^2 remains positive for Gradient Boosting (8.4%), Random Forest (4.4%), and XGBoost (8.7%). Without a market-return control, the corresponding values are 11.3%, 7.4%, and 9.1%. The baseline VWRETX row reproduces Table 4 exactly. These results support the interpretation that the medium-horizon signal is not simply a repackaged contemporaneous equity-market return.

5.4 Realized versus predicted six-month returns

[Figure 2]

Figure 2 compares the monthly cross-sectional average realized six-month RETX with the corresponding frozen XGBoost forecast, indexed by forecast target-end month. The two monthly series have a correlation of 0.530, although uncertainty around this estimate is substantial: a circular moving-block bootstrap using six-month blocks yields an approximate 95% confidence interval of [0.07, 0.71]. The predicted series is also substantially smoother than the realized series, with a standard deviation of 0.0418 compared with 0.1988 for realized average returns. Thus, the model captures only about one-fifth of realized aggregate volatility.

This smoothing does not fully account for the forecasting gains. At the monthly aggregate level, XGBoost reduces RMSE from 0.2032 for the pre-test historical mean to 0.1801, corresponding to an aggregate out-of-sample R^2 of 21.4%. A Mincer–Zarnowitz regression of realized average returns on the XGBoost forecast yields a slope of 2.52, which is significantly different from zero using HAC(5) standard errors ($p = 0.005$). The magnitude of the slope is economically informative: a value substantially above one indicates pronounced under-dispersion, meaning that the forecasts contain aggregate directional information but systematically attenuate the amplitude of realized movements. This calibration evidence reinforces the interpretation of the model as capturing a smooth common return component rather than reproducing the full volatility of REIT returns.

The evidence is different cross-sectionally. The average monthly Spearman correlation between predicted and realized individual-REIT returns is approximately zero (0.0005; HAC $p = 0.991$), and the average subsequent return spread between the highest and lowest predicted

quintiles is only 0.23 percentage points over six months (HAC $p = 0.944$). The forecasting gains should therefore be interpreted primarily as time-series predictability in the common component of REIT returns rather than cross-sectional security-selection ability.

5.5 Firm fixed-effects linear benchmark

A stronger linear benchmark allows each REIT to have a time-invariant intercept. Firm effects are estimated using only pre-test observations, and the out-of-sample comparison is restricted to REITs with sufficient pre-2020 history to estimate an ex ante firm effect. Pooled OLS is re-estimated on exactly the same restricted test observations for comparability.

[Table 12]

Table 12 shows that fixed effects do not overturn the main result. At six months, firm fixed effects raise out-of-sample R^2 from -1.0% for pooled OLS to 3.1% , indicating that persistent cross-sectional heterogeneity contains some medium-horizon forecasting information. On the identical 7,079-observation test sample, however, Gradient Boosting reaches 9.7% and XGBoost 9.5% , with lower RMSE and MAE than the firm-FE benchmark. Thus, the medium-horizon machine-learning gains cannot be attributed solely to omitted time-invariant REIT-specific effects. The restriction excludes 632 of 7,711 six-month test observations from 24 REITs that lack usable pre-test histories.

6. Property-Sector Heterogeneity

6.1 Ziman merge and sector composition

The aggregate analysis treats listed REITs as a single asset class even though the underlying property sectors differ in tenant demand, lease structure, capital intensity, and sensitivity to macroeconomic conditions. To examine whether these differences matter for forecasting, the cleaned CRSP/FRED panel is merged with the CRSP/Ziman monthly REIT file using PERMNO and calendar month.

The merge matches 44,526 of 47,075 cleaned firm-month observations, a 94.6% match rate, and covers 366 of the 383 PERMNOs in the CRSP/FRED sample. The main sector analysis is restricted to equity REITs ($rtype = 2$, In the Ziman taxonomy, $rtype = 2$ denotes equity REITs; mortgage REITs are excluded from the main sector analysis.), yielding 34,797 firm-month

observations and 283 unique equity REITs. Broad property types are defined from the Ziman ptype code, with the broad industrial/office group further separated using psub where the constituent structure permits a clear split.

[Table 10]

Table 10 shows substantial representation in retail, industrial/office-related, residential, diversified, lodging, and health-care REITs. Self-storage has a long time series but relatively few distinct firms, while the specialty category is economically heterogeneous. These sample differences motivate caution in interpreting small-sector estimates and make the larger retail, residential, office, and diversified groups particularly informative. In this table, The single firm appearing in the mortgage/real-estate-finance property-type category within the equity-REIT sample reflects the Ziman property-type coding and is not analyzed as a separate sector.

6.2 Sector-level forecasting results

[Table 11]

Table 11 reports the six-month forecasting results separately within major equity REIT sectors and shows all four model families rather than selecting a single winner from the test period. The point estimates reveal substantial heterogeneity: XGBoost is positive for health care and retail, Random Forest is positive for lodging and office, and OLS is strongest for diversified and residential REITs. Because several sector cells contain relatively few firms, these model-by-sector point estimates are treated as descriptive rather than definitive evidence.

The contrast with the linear benchmark remains economically informative. Diversified and residential REITs favor OLS, whereas several more cyclical sectors display positive nonlinear point estimates. However, the differences in Table 11 do not by themselves establish that a given model-family advantage is statistically reliable, particularly in health care, lodging, office, self-storage, and specialty categories with thin firm counts.

Accordingly, the main sector claim is based on the pre-specified XGBoost inference in Table 13 rather than on choosing the best-performing machine-learning family after observing test outcomes.

6.3 Formal sector-level inference

To assess the reliability of sector heterogeneity without post-test model-family selection, Table 13 evaluates the pre-specified six-month XGBoost model across sectors and reports both formal forecast-loss tests and bootstrap uncertainty. The XGBoost specification and hyperparameters are fixed before the 2020–2023 test period and are not re-tuned by sector using test observations.

[Table 13]

The evidence is strongest for health-care REITs. XGBoost produces an out-of-sample R^2 of 11.5%, with both month-block and REIT-cluster 95% confidence intervals excluding zero. Its forecast-loss improvement relative to the historical mean is statistically significant ($p = 0.002$) and remains below the 5% level after a Holm adjustment across the nine sector comparisons. Retail REITs also show material predictive gains in out-of-sample R^2 and bootstrap intervals. XGBoost reaches 10.3%, and both bootstrap confidence intervals remain positive, although the Diebold–Mariano tests versus OLS ($p = 0.261$) and the historical mean ($p = 0.118$) do not reject equal predictive accuracy.

Evidence for lodging and office is more fragile. Their XGBoost point estimates are positive but the month-block confidence intervals include zero, while self-storage is too thin to support a headline conclusion. Industrial/data-center and specialty REITs provide no convincing positive nonlinear evidence. Residential and diversified REITs remain important counterexamples in which OLS performs better. Sector heterogeneity is therefore real but more qualified than the raw model rankings alone suggest.

[Figure 3]

Figure 3 summarizes the confirmatory sector evidence by comparing OLS with the pre-specified XGBoost model rather than the ex post best-performing machine-learning family. The XGBoost markers include 95% month-block bootstrap confidence intervals. Health care and retail stand out as the clearest positive nonlinear cases, whereas diversified and residential REITs favor OLS and several smaller sectors have wide uncertainty. This visualization therefore emphasizes both heterogeneity and sampling uncertainty.

7. Pre-2020 pseudo-out-of-sample robustness.

To assess whether the relative advantage of nonlinear models is unique to the unusual 2020–2023 macroeconomic environment, the forecasting exercise is repeated using an earlier chronological split. Models are initially estimated through 2010, hyperparameters are selected exclusively on 2011–2013 validation observations, and the selected specifications are refit through 2013 before evaluation on forecast target-end months from 2014–2016. The earlier holdout does not reproduce the positive out-of-sample R^2 values observed in 2020–2023: all models remain below the historical-mean benchmark. Nevertheless, the relative model ranking is informative. At the three-month cumulative horizon, Random Forest reduces RMSE from 0.2039 for pooled OLS to 0.1166, while at six months Gradient Boosting and XGBoost reduce RMSE from 0.3332 to approximately 0.1910. Thus, the relative advantage of nonlinear specifications over pooled OLS is already visible in a pre-COVID holdout, although their ability to outperform the unconditional historical-mean benchmark is regime-dependent. This result is depicted in appendix table A2.

8. Conclusion

This paper evaluates whether a parsimonious set of macroeconomic and market variables can forecast U.S. REIT returns more effectively when modeled nonlinearly. The evidence draws a clear distinction between short- and medium-horizon predictability. Direct one- to three-month returns remain difficult to forecast, while cumulative returns become more predictable at longer horizons. In the 2020–2023 final holdout, Gradient Boosting and XGBoost produce six-month out-of-sample R^2 values of approximately 9.3% relative to the historical mean, compared with -0.6% for pooled OLS. Formal forecast-comparison tests, however, do not establish statistical dominance over OLS or the historical mean after adjustment for multiple comparisons. Moreover, a pre-2020 pseudo-out-of-sample exercise does not reproduce positive out-of-sample R^2 relative to the historical mean, although nonlinear models continue to outperform pooled OLS in relative forecast-error terms. The aggregate evidence is therefore best interpreted as showing material but regime-dependent medium-horizon forecast improvement rather than a universally stable nonlinear advantage.

The explainability results indicate that real GDP per capita, unemployment, policy rates, changes in policy rates, and long-term real interest rates account for most of the six-month XGBoost model's predictive importance. These SHAP patterns are descriptive of the fitted model rather than causal effects, and the apparent thresholds should be interpreted cautiously because the

2020–2023 test period spans an unusually sharp sequence of pandemic disruption, near-zero rates, inflation, and rapid monetary tightening.

The property-sector extension sharpens the economic contribution but also shows why statistical uncertainty matters. Formal inference is strongest for health-care REITs, where the pre-specified XGBoost model reaches 11.5% out-of-sample R^2 and remains statistically significant relative to the historical mean after adjustment for multiple sector comparisons. Retail REITs also show positive and bootstrap-robust predictive gains, although their DM tests are not significant at conventional levels. Evidence for lodging and office is more suggestive, while residential and diversified REITs favor OLS. Thus, model complexity is useful selectively rather than uniformly across property types.

Several additional diagnostics reinforce this narrower interpretation. Firm fixed effects improve the six-month linear benchmark to 3.1% out-of-sample R^2 , but the tree-based models remain near 9.5–9.7% on the same sample. The aggregate XGBoost forecasts are substantially smoother than realized returns yet retain significant timing information; by contrast, cross-sectional forecast rankings contain essentially no security-selection signal. The results should therefore be interpreted as forecasting evidence about the common, macro-sensitive component of medium-horizon REIT returns rather than causal identification or a ready-made stock-picking strategy. Future work can extend the out-of-sample period, add credit and property-market predictors, and evaluate investable market-timing rules under transaction costs. Overall, the evidence supports a focused conclusion: nonlinear machine learning can add value at medium horizons and in selected property sectors, but the magnitude of that advantage is regime-dependent, and its usefulness is conditional rather than universal.

References

- Avramov, D., Cheng, S., & Metzker, L. (2022). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science*. <https://doi.org/10.1287/mnsc.2022.4449>
- Azevedo, V., Kaiser, G., & Mueller, S. (2023). Stock market anomalies and machine learning across the globe. *Journal of Asset Management*. <https://doi.org/10.1057/s41260-023-00318-z>
- Bali, T. G., Beckmeyer, H., Mörke, M., & Weigert, F. (2023). Option return predictability with machine learning and big data. *Review of Financial Studies*. <https://doi.org/10.1093/rfs/hhad017>
- Bianchi, D., Buechner, M., & Tamoni, A. (2020). Bond risk premiums with machine learning. *Review of Financial Studies*. <https://doi.org/10.1093/rfs/hhaa062>
- Breitung, C. (2023). Automated stock picking using random forests. *Journal of Empirical Finance*. <https://doi.org/10.1016/j.jempfin.2023.05.001>
- Chan, K. C., Hendershott, P. H., & Sanders, A. B. (1990). Risk and return on real estate: Evidence from equity REITs. *Real Estate Economics*. <https://doi.org/10.1111/1540-6229.00531>
- Chandrashekar, V. (1999). Time-series properties and diversification benefits of REIT returns. *Journal of Real Estate Research*. <https://doi.org/10.1080/10835547.1999.12090963>
- Chen, K., & Tzang, D. (1988). Interest-rate sensitivity of real estate investment trusts. *Journal of Real Estate Research*. <https://doi.org/10.1080/10835547.1988.12090561>
- Chen, N.-F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*. <https://doi.org/10.1086/296344>
- Chen, S.-J., Hsieh, C., & Jordan, B. D. (1997). Real estate and the arbitrage pricing theory: Macrovariables vs. derived factors. *Real Estate Economics*. <https://doi.org/10.1111/1540-6229.00725>
- Chen, S.-J., Hsieh, C., Vines, T., & Chiou, S.-N. (1998). Macroeconomic variables, firm-specific variables and returns to REITs. *Journal of Real Estate Research*. <https://doi.org/10.1080/10835547.1998.12090954>
- Choi, D., Jiang, W., & Zhang, C. (2025). Alpha go everywhere: Machine learning and international stock returns. *Review of Asset Pricing Studies*. <https://doi.org/10.1093/rapstu/raaf005>
- Clayton, J., & MacKinnon, G. (2001). The time-varying nature of the link between REIT, real estate and financial asset returns. *Journal of Real Estate Portfolio Management*. <https://doi.org/10.1080/10835547.2001.12089632>
- Cohen, V., Burinskas, A., & Fateye, T. B. (2026). Reevaluating macroeconomic drivers for European REIT (EU-REIT) index: An extended exploration of performance factors. *Journal of European Real Estate Research*. <https://doi.org/10.1108/jerer-03-2025-0035>
- Connolly, R. A., & Stivers, C. T. (2024). Public real estate returns and inflation shocks: The central role of inflation nonneutrality. *Real Estate Economics*. <https://doi.org/10.1111/1540-6229.12495>
- Demiralay, S., & Kılınçarslan, E. (2022). Uncertainty measures and sector-specific REITs in a regime-switching environment. *The Journal of Real Estate Finance and Economics*. <https://doi.org/10.1007/s11146-022-09898-w>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. <https://doi.org/10.1080/07350015.1995.10524599>
- Drobetz, W., & Otto, T. (2021). Empirical asset pricing via machine learning: Evidence from the European stock market. *Journal of Asset Management*. <https://doi.org/10.1057/s41260-021-00237-x>
- Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*. [https://doi.org/10.1016/0304-405x\(89\)90095-0](https://doi.org/10.1016/0304-405x(89)90095-0)
- Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*. <https://doi.org/10.1111/j.1540-6261.1992.tb04398.x>
- Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*. <https://doi.org/10.1086/260061>
- Fama, E. F., & Schwert, G. W. (1977). Asset returns and inflation. *Journal of Financial Economics*. [https://doi.org/10.1016/0304-405x\(77\)90014-9](https://doi.org/10.1016/0304-405x(77)90014-9)

- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*. <https://doi.org/10.1093/rfs/hhaa009>
- Gyourko, J., & Keim, D. B. (1992). What does the stock market tell us about real estate returns? *Real Estate Economics*. <https://doi.org/10.1111/1540-6229.00591>
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291. [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4)
- Hanauer, M. X., & Kalsbach, T. (2023). Machine learning and the cross-section of emerging market stock returns. *Emerging Markets Review*. <https://doi.org/10.1016/j.ememar.2023.101022>
- Jenett, H., Nagl, C., Nagl, M., Price, S. M., & Schaefers, W. (2025). Dynamics of REIT returns and volatility: Analyzing time-varying drivers through an explainable machine learning approach. *Journal of Real Estate Finance and Economics*. <https://doi.org/10.1007/s11146-025-10016-9>
- Kelly, B., Malamud, S., & Zhou, K. (2023). The virtue of complexity in return prediction. *The Journal of Finance*. <https://doi.org/10.1111/jofi.13298>
- Leippold, M., Wang, Q., & Zhou, W. (2021). Machine learning in the Chinese stock market. *Journal of Financial Economics*. <https://doi.org/10.1016/j.jfineco.2021.08.017>
- Leow, K., & Lindenthal, T. (2025). Enhancing real estate investment trust return forecasts using machine learning. *Real Estate Economics*. <https://doi.org/10.1111/1540-6229.12527>
- Leung, E., Lohre, H., Mischlich, D., Shea, Y., & Strohm, M. (2021). The promises and pitfalls of machine learning for predicting stock returns. *The Journal of Financial Data Science*. <https://doi.org/10.3905/jfds.2021.1.062>
- Li, Y., & Wang, K. (1995). The predictability of REIT returns and market segmentation. *Journal of Real Estate Research*. <https://doi.org/10.1080/10835547.1995.12090800>
- Liang, Y., Prudential, W., & Webb, J. (1995). Intertemporal changes in the riskiness of REITs. *Journal of Real Estate Research*. <https://doi.org/10.1080/10835547.1995.12090795>
- Liu, C. H., & Mei, J. (1992). The predictability of returns on equity REITs and their co-movement with other assets. *Journal of Real Estate Finance and Economics*. <https://doi.org/10.1007/bf00174808>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2, 56–67. <https://doi.org/10.1038/s42256-019-0138-9>
- McCue, T. E., & Kling, J. L. (1994). Real estate returns and the macroeconomy: Some empirical evidence from real estate investment trust data, 1972–1991. *Journal of Real Estate Research*. <https://doi.org/10.1080/10835547.1994.12090754>
- Mekelburg, E., & Strauss, J. (2024). Pooling and winsorizing machine learning forecasts to predict stock returns with high-dimensional data. *Journal of Empirical Finance*. <https://doi.org/10.1016/j.jempfin.2024.101538>
- Mueller, G., & Pauley, K. (1995). The effect of interest-rate movements on real estate investment trusts. *Journal of Real Estate Research*. <https://doi.org/10.1080/10835547.1995.12090793>
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*. <https://doi.org/10.1257/jep.31.2.87>
- Nayar, N., Price, S. M., & Shen, K. (2023). Macroeconomic uncertainty and predictability of real estate returns. *Journal of Real Estate Research*. <https://doi.org/10.1080/08965803.2023.2211812>
- Olanrele, O. O., Fateye, T. B., Adegunle, T. O., Ajayi, C. A., Said, R., & Baaki, K. (2020). Causal effects of macroeconomic predictors on real estate investment trust's (REIT's) performance in Nigeria. *Pacific Rim Property Research Journal*. <https://doi.org/10.1080/14445921.2020.1857498>
- Park, J. Y., Mullineaux, D. J., & Chew, I. K. (1990). Are REITs inflation hedges? *The Journal of Real Estate Finance and Economics*. <https://doi.org/10.1007/bf00153708>
- Rapach, D. E., & Zhou, G. (2023). Forecasting stock returns. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4403635>

- Rasekhschaffe, K. C., & Jones, R. C. (2019). Machine learning for stock selection. *Financial Analysts Journal*.
<https://doi.org/10.1080/0015198x.2019.1596678>
- Rosa, C. (2024). The impact of monetary policy on REITs: Evidence from FOMC announcements. *The Journal of Real Estate Finance and Economics*. <http://dx.doi.org/10.1007/s11146-024-09992-1>
- Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*.
[https://doi.org/10.1016/0022-0531\(76\)90046-6](https://doi.org/10.1016/0022-0531(76)90046-6)
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>
- Simpson, M. W., Ramchander, S., & Webb, J. R. (2007). The asymmetric response of equity REIT returns to inflation. *The Journal of Real Estate Finance and Economics*. <https://doi.org/10.1007/s11146-007-9023-0>
- Titman, S., & Warga, A. (1986). Risk and the performance of real estate investment trusts. *Real Estate Economics*.
<https://doi.org/10.1111/1540-6229.00395>

Tables

Table 1. Descriptive Statistics

Panel A. REIT-level variables

Variable	N	Mean	Std. Dev.	25th pct.	Median	75th pct.
RETX	46,894	0.0045	0.1096	-0.0391	0.0052	0.0480
RET	46,894	0.0100	0.1091	-0.0336	0.0104	0.0532
Market capitalization (\$ millions)	47,037	3,167.44	7,211.65	334.65	1,097.78	2,974.31

Panel B. Market and macroeconomic variables

Variable	N	Mean	Std. Dev.	25th pct.	Median	75th pct.
VWRETX	288	0.0049	0.0461	-0.0214	0.0104	0.0331
EWRETX	288	0.0065	0.0559	-0.0229	0.0078	0.0371
Federal funds rate	288	1.7853	1.9504	0.1400	1.0950	2.6150
Unemployment rate	288	5.7649	1.9708	4.3000	5.2000	6.7000
Median CPI	288	2.7682	1.3580	2.0636	2.5380	3.1753
Real GDP per capita	288	56,859.80	4,931.15	53,645.00	55,493.50	60,242.75
Real 10-year interest rate	278	1.0666	0.7840	0.4903	1.0159	1.7138
Δ Federal funds rate	285	-0.0009	0.1879	-0.0100	0.0000	0.0300
Δ Unemployment rate	285	0.0007	0.6733	-0.1000	0.0000	0.1000
Δ Median CPI	285	-0.0050	0.8032	-0.5379	-0.0471	0.5228
Δ Real GDP per capita	285	61.22	452.92	0.00	0.00	18.00
Δ Real 10-year interest rate	275	-0.0043	0.2075	-0.1230	0.0042	0.1046

Notes: Table 1 reports descriptive statistics for the variables used in the empirical analysis over January 2000–December 2023. Panel A is based on REIT firm-month observations. RET is the CRSP total return, while RETX is the return excluding distributions. Market capitalization is calculated as the absolute value of CRSP price multiplied by shares outstanding and is expressed in millions of dollars. Panel B is based on 288 unique calendar months. VWRETX and EWRETX are the CRSP value-weighted and equal-weighted market returns excluding distributions. The federal funds rate, unemployment rate, median CPI, real GDP per capita, and real 10-year interest rate are obtained from FRED. The prefix Δ denotes the first difference of the corresponding macroeconomic variable. The smaller number of observations for the real 10-year interest rate reflects missing observations in the underlying series.

Table 2. Correlations among REIT Return, Market, and Macroeconomic Variables

Panel A. REIT Returns, Market Returns, and Lagged Returns

Variable	RETX	VWRETX	EWRETX	RET lag 1	RET lag 2	RET lag 3
RETX	1.000					
VWRETX	0.392	1.000				
EWRETX	0.410	0.894	1.000			
RET lag 1	−0.031	−0.005	0.045	1.000		
RET lag 2	−0.045	−0.071	−0.060	−0.037	1.000	
RET lag 3	0.021	0.049	0.020	−0.047	−0.036	1.000

Panel B. Market Returns and Macroeconomic Levels

Variable	VWRETX	EWRETX	Fed Funds	Unemployment	Median CPI	Real GDP p.c.	Real 10Y Rate
VWRETX	1.000						
EWRETX	0.897	1.000					
Federal funds rate	−0.075	−0.065	1.000				
Unemployment rate	0.162	0.196	−0.576	1.000			
Median CPI	−0.110	−0.144	0.348	−0.553	1.000		
Real GDP per capita	0.058	−0.070	−0.116	−0.345	0.510	1.000	
Real 10-year rate	−0.142	−0.081	0.763	−0.366	0.157	−0.423	1.000

Panel C. Changes in Macroeconomic Variables

Variable	Δ Fed Funds	Δ Unemployment	Δ Median CPI	Δ Real GDP p.c.	Δ Real 10Y Rate
Δ Federal funds rate	1.000				
Δ Unemployment rate	−0.241	1.000			
Δ Median CPI	0.049	−0.037	1.000		
Δ Real GDP per capita	0.217	−0.672	0.133	1.000	
Δ Real 10-year rate	0.119	−0.010	−0.008	0.019	1.000

Notes: Panel A is based on REIT firm-month observations. RETX denotes the CRSP return excluding distributions. VWRETX and EWRETX denote the CRSP value-weighted and equal-weighted market returns excluding distributions. RET lag 1 through RET lag 3 are lags of CRSP total return (RET). Panels B and C are based on one observation per month from January 2000 through December 2023. The real 10-year interest-rate series contains fewer observations because of missing values. The prefix Δ denotes the first difference of the corresponding macroeconomic variable.

Table 3. Baseline OLS Predictive Regressions for REIT Returns

Variable	1-Month Ahead	2-Month Ahead	3-Month Ahead	3-Month Cumulative	6-Month Cumulative
Constant	-0.0976 (-0.89)	0.0455 (0.45)	0.0908 (0.88)	0.1683 (0.97)	0.3033 (1.51)
Federal funds rate	-0.0018 (-0.44)	0.0035 (0.91)	0.0047 (1.26)	0.0077 (1.21)	0.0090 (0.99)
Median CPI	-0.0046 (-0.99)	-0.0028 (-0.61)	-0.0055 (-1.12)	-0.0122** (-1.99)	-0.0156* (-1.86)
Real 10-year rate	0.0121 (0.93)	-0.0085 (-0.72)	-0.0126 (-1.17)	-0.0198 (-1.03)	-0.0262 (-1.14)
Real GDP per capita / 1,000	0.0016 (0.98)	-0.0006 (-0.40)	-0.0010 (-0.63)	-0.0021 (-0.75)	-0.0046 (-1.43)
Unemployment rate	0.0051* (1.65)	0.0043 (1.54)	0.0021 (0.68)	0.0093* (1.88)	0.0236*** (3.75)
VWRETX	0.0912 (0.71)	-0.1599 (-1.51)	-0.0121 (-0.12)	-0.0694 (-0.39)	-0.1315 (-0.52)
RET lag 1	-0.0584* (-1.70)	0.0296 (0.97)	0.0315 (0.95)	0.0187 (0.35)	-0.0651 (-0.76)
RET lag 2	0.0194 (0.58)	0.0455 (1.44)	-0.0008 (-0.02)	0.0789 (1.60)	0.0462 (0.69)
RET lag 3	0.0411 (1.29)	0.0146 (0.44)	-0.0446 (-1.21)	0.0164 (0.32)	-0.0001 (-0.00)
Δ Federal funds rate	-0.0180 (-0.58)	0.0266 (1.03)	0.0299 (1.06)	0.0487 (0.99)	0.0950* (1.65)
Δ Median CPI	-0.0031 (-0.60)	-0.0011 (-0.17)	0.0131** (2.24)	0.0063 (0.78)	0.0018 (0.17)
Δ Real 10-year rate	-0.0060 (-0.31)	0.0313 (1.64)	0.0303* (1.66)	0.0541* (1.74)	0.0278 (0.74)
Δ Real GDP per capita / 1,000	0.0147 (1.19)	0.0026 (0.15)	-0.0106 (-0.93)	0.0116 (0.42)	0.0422** (1.99)
Δ Unemployment rate	-0.0029 (-0.42)	0.0095 (1.21)	-0.0063 (-0.66)	0.0036 (0.28)	-0.0040 (-0.29)
Log market capitalization	-0.0014* (-1.68)	-0.0013* (-1.69)	-0.0012 (-1.62)	-0.0038** (-2.44)	-0.0070*** (-2.73)
Observations	43,650	43,241	42,833	42,833	41,623
REIT clusters	368	368	367	367	362
Calendar-month clusters	276	275	274	274	271
R ²	0.0217	0.0200	0.0226	0.0390	0.0763
Adjusted R ²	0.0214	0.0196	0.0223	0.0387	0.0760

Notes: Table 3 reports pooled OLS predictive regressions of future REIT returns on current macroeconomic conditions, market returns, lagged REIT total returns, and firm size. The dependent variables are one-, two-, and three-month-ahead RETX, as well as compounded three- and six-month future RETX. RETX is the CRSP return excluding distributions. VWRETX is the CRSP value-weighted market return excluding distributions. RET lag 1 through RET lag 3 are reconstructed from CRSP total return after the panel is reduced to one observation per PERMNO-month and require consecutive calendar months. Real GDP per capita and its change are divided by 1,000 for readability. Market capitalization enters in logarithmic form. Forward return targets require consecutive calendar months. Standard errors are two-way clustered by REIT and calendar month to account for serial dependence within securities and common shocks across REITs. t-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 4. Out-of-Sample Forecasting Performance

Panel A. Out-of-Sample R^2

Model	1-Month	2-Month	3-Month	3-Month Cumulative	6-Month Cumulative
OLS	-0.1475	-0.0050	-0.1061	-0.0403	-0.0065
Gradient Boosting	-0.1029	-0.0577	-0.1138	0.0556	0.0930
Random Forest	-0.0513	-0.0673	-0.1454	-0.0041	0.0645
XGBoost	-0.1300	-0.0268	-0.0674	0.0317	0.0932

Panel B. RMSE

Model	1-Month	2-Month	3-Month	3-Month Cumulative	6-Month Cumulative
OLS	0.1472	0.1379	0.1446	0.2230	0.3087
Gradient Boosting	0.1443	0.1415	0.1451	0.2125	0.2930
Random Forest	0.1409	0.1421	0.1471	0.2191	0.2976
XGBoost	0.1461	0.1394	0.1420	0.2151	0.2930

Panel C. MAE

Model	1-Month	2-Month	3-Month	3-Month Cumulative	6-Month Cumulative
OLS	0.0978	0.0902	0.0953	0.1518	0.2188
Gradient Boosting	0.0938	0.0955	0.0965	0.1455	0.2079
Random Forest	0.0917	0.0950	0.0975	0.1519	0.2111
XGBoost	0.0947	0.0936	0.0936	0.1475	0.2055

Notes: Table 4 reports chronological out-of-sample forecasting performance. The estimation sample covers 2000–2015, the validation sample covers 2016–2019, and forecast targets realized in 2020–2023 form the final test sample. Hyperparameters are selected exclusively using validation-period RMSE and are frozen before each model is refit on the combined 2000–2019 sample; no test observations are used for tuning. All models use an identical predictor set. R^2_{OOS} is measured relative to the historical-mean return estimated from pre-test observations. Forward targets require consecutive calendar months in the cleaned one-PERMNO-per-month panel. Test sample sizes are 7,874, 7,843, 7,812, 7,811, and 7,711 for the one-month, two-month, three-month, three-month cumulative, and six-month cumulative horizons, respectively.

Table 5. Diebold–Mariano Tests of Out-of-Sample Forecast Accuracy

Horizon	ML Model	Benchmark	Mean Loss Improvement	DM Statistic	p-value	Holm-adjusted p-value
3-Month cumulative	Gradient Boosting	OLS	0.00457	1.060	0.295	1.000
3-Month cumulative	Gradient Boosting	Historical mean	0.00267	1.308	0.197	1.000
3-Month cumulative	Random Forest	OLS	0.00176	0.504	0.616	1.000
3-Month cumulative	Random Forest	Historical mean	-0.00014	-0.051	0.959	1.000
3-Month cumulative	XGBoost	OLS	0.00344	0.987	0.328	1.000
3-Month cumulative	XGBoost	Historical mean	0.00154	0.691	0.493	1.000
6-Month cumulative	Gradient Boosting	OLS	0.00923	0.689	0.494	1.000
6-Month cumulative	Gradient Boosting	Historical mean	0.00885	1.281	0.206	1.000
6-Month cumulative	Random Forest	OLS	0.00657	0.564	0.576	1.000
6-Month cumulative	Random Forest	Historical mean	0.00620	1.049	0.300	1.000
6-Month cumulative	XGBoost	OLS	0.00929	0.909	0.368	1.000
6-Month cumulative	XGBoost	Historical mean	0.00892	1.902	0.063	0.756

Notes: Table 5 reports Diebold–Mariano-type tests based on differences in squared forecast errors. Positive mean loss improvements indicate smaller errors for the machine-learning model. Loss differentials are averaged across REITs within each forecast target-end month before inference. HAC standard errors use $h - 1$ lags to account for overlapping cumulative-return horizons, and the test statistic applies the Harvey–Leybourne–Newbold small-sample correction. Each comparison uses 48 target-end months. Two-sided unadjusted p-values are reported together with Holm-adjusted p-values controlling the familywise error rate across the 12 comparisons in the table.

Table 6. Variable Importance for the Six-Month XGBoost REIT Return Forecast

Panel A. Variable Importance

Variable	SHAP Rank	Mean Absolute SHAP	SHAP Importance (%)	Gain Importance (%)
Real GDP per capita	1	0.0305	27.62	9.43
Unemployment rate	2	0.0161	14.57	9.62
Federal funds rate	3	0.0146	13.18	6.33
Δ Federal funds rate	4	0.0110	9.95	7.46
CRSP value-weighted market return (VWRETX)	5	0.0103	9.31	3.86
Real 10-year interest rate	6	0.0093	8.41	16.24
Δ Unemployment rate	7	0.0049	4.39	10.63
RET lag 1	8	0.0044	3.98	11.00
Δ Real GDP per capita	9	0.0040	3.63	4.12
RET lag 3	10	0.0021	1.89	5.26
Median CPI	11	0.0017	1.54	4.40
Δ Real 10-year interest rate	12	0.0008	0.70	6.01
RET lag 2	13	0.0007	0.62	3.59
Log market capitalization	14	0.0002	0.21	2.06
Δ Median CPI	15	0.0000	0.00	0.00
Total			100.00	100.00

Panel B. Model Performance

Model performance	Value
Out-of-sample R ²	0.0932
RMSE	0.2930
MAE	0.2055
Test observations	7,711
Test period	2020–2023

Notes: Table 6 reports variable importance for the XGBoost model used to forecast six-month cumulative REIT returns. Variables are ranked according to mean absolute SHAP values calculated over the 2020–2023 out-of-sample test period. Mean Absolute SHAP measures the average magnitude of each predictor's contribution to the model's forecasts. SHAP Importance (%) expresses each variable's mean absolute SHAP value as a percentage of total SHAP importance. Gain Importance (%) measures the proportion of improvement in the XGBoost objective associated with tree splits involving each predictor. The prefix Δ denotes the first difference of the corresponding variable. SHAP values measure predictive contribution and should not be interpreted as causal effects.

Table 7. Robustness to Using CRSP Total Returns**Panel A. Three-Month Cumulative Total Return**

Model	Out-of-Sample R²	RMSE	MAE
OLS	-0.0301	0.2235	0.1525
Gradient Boosting	0.0211	0.2179	0.1507
Random Forest	0.0375	0.2161	0.1484
XGBoost	0.0186	0.2182	0.1511

Panel B. Six-Month Cumulative Total Return

Model	Out-of-Sample R²	RMSE	MAE
OLS	-0.0412	0.3184	0.2268
Gradient Boosting	0.0908	0.2975	0.2133
Random Forest	0.0598	0.3025	0.2157
XGBoost	0.0911	0.2975	0.2095

Notes: Table 7 reports out-of-sample forecasting results using CRSP total return (RET), which includes distributions, rather than RETX, which excludes distributions. Three- and six-month cumulative returns are formed by compounding consecutive forward monthly returns. Models are initially estimated using 2000–2015 observations, model specifications are selected using the 2016–2019 validation period, and final forecast performance is evaluated on 2020–2023. All models use the same predictor set as the main analysis. R²_OOS is measured relative to the historical-mean total-return forecast estimated from the pre-test sample. Positive values indicate an improvement relative to the historical-mean benchmark. Bold entries identify the best-performing model within each forecast horizon.

Table 8. Subperiod Robustness of Six-Month REIT Return Forecasts

Subsample	Model	Out-of-Sample R ²	RMSE	MAE	N
Full test period (2020–2023)	Gradient Boosting	0.0930	0.2930	0.2079	7,711
Full test period (2020–2023)	Random Forest	0.0645	0.2976	0.2111	7,711
Full test period (2020–2023)	XGBoost	0.0932	0.2930	0.2055	7,711
Exclude March– December 2020 target- end months	Gradient Boosting	0.0549	0.2717	0.1880	6,102
Exclude March– December 2020 target- end months	Random Forest	0.0315	0.2750	0.1903	6,102
Exclude March– December 2020 target- end months	XGBoost	0.0550	0.2717	0.1845	6,102
2020–2021 target-end months	Gradient Boosting	0.1476	0.3381	0.2406	3,876
2020–2021 target-end months	Random Forest	0.0990	0.3476	0.2484	3,876
2020–2021 target-end months	XGBoost	0.1015	0.3471	0.2483	3,876
2022–2023 target-end months	Gradient Boosting	-0.0422	0.2390	0.1747	3,835
2022–2023 target-end months	Random Forest	-0.0208	0.2365	0.1733	3,835
2022–2023 target-end months	XGBoost	0.0728	0.2254	0.1622	3,835

Notes: Table 8 reports six-month cumulative RETX forecast performance across alternative portions of the 2020–2023 test period. Subsamples are defined consistently by forecast target-end month. Models are estimated through 2015, hyperparameters are selected on 2016–2019, then frozen and refit through 2019. No reported subsample is used for model selection. R²_OOS is measured relative to the historical mean estimated from pre-test observations.

Table 9. Robustness to Alternative Market-Return Controls

Market-Control Specification	Model	Out-of-Sample R²	RMSE	MAE
VWRETX control	Gradient Boosting	0.0930	0.2930	0.2079
VWRETX control	Random Forest	0.0645	0.2976	0.2111
VWRETX control	XGBoost	0.0932	0.2930	0.2055
EWRETX control	Gradient Boosting	0.0840	0.2945	0.2094
EWRETX control	Random Forest	0.0439	0.3009	0.2136
EWRETX control	XGBoost	0.0869	0.2940	0.2060
No market-return control	Gradient Boosting	0.1129	0.2898	0.2045
No market-return control	Random Forest	0.0735	0.2962	0.2079
No market-return control	XGBoost	0.0914	0.2933	0.2048

Test observations: 7,711 in each specification. Notes: Table 9 reports six-month cumulative RETX forecast performance under alternative treatments of the broad equity-market control. The baseline includes VWRETX, the second specification substitutes EWRETX, and the third omits the market-return control. The six-month model settings are frozen from the validation exercise; each specification is refit on the combined pre-test data. The VWRETX row therefore reproduces the corresponding Table 4 results.

Table 10. Distribution of Equity REIT Observations by Property Type

Ziman ptype	Property Type	Firm-Month Observations	Unique REITs	Share of Equity REIT Sample (%)	Sample Period
1	Specialty / Other	1,851	17	5.32	2000–2023
2	Diversified / Mixed	3,640	28	10.46	2000–2023
3	Health Care	2,649	22	7.61	2000–2023
4	Industrial / Office-related	8,155	72	23.44	2000–2023
5	Lodging / Hospitality	3,254	28	9.35	2000–2023
6	Mortgage / Real Estate Finance*	165	1	0.47	2005–2018
8	Residential	5,180	38	14.89	2000–2023
9	Retail	8,687	68	24.96	2000–2023
10	Self-Storage	1,108	7	3.18	2000–2023
0	Unclassified	108	3	0.31	2001–2004
Total		34,797	283	100.00	2000–2023

Notes: Table 10 reports the distribution of observations in the merged CRSP/FRED–Ziman sample restricted to equity REITs (rtype = 2). Property classifications are based on the Ziman ptype variable. The CRSP/FRED and Ziman datasets are merged using PERMNO and calendar month. The overall merge rate is 94.6%. Property-type labels are recovered from the constituent firms associated with each Ziman numeric classification; where the classification is especially broad, the label is intentionally conservative. The mortgage/real-estate-finance category contains only one equity-classified REIT in the merged sample and will therefore not be analyzed separately in the sector forecasting tests. Unclassified observations are also excluded from subsequent sector-level forecasting analysis. Property-type labels are conservative constituent-based labels for the Ziman ptype/psub groupings; ambiguous categories are grouped as Specialty / Other and are not used for strong headline inference.

Table 11. Six-Month Out-of-Sample Forecasting Performance by Equity REIT Property Sector

Property Sector	Model	OOS R ²	RMSE	MAE	Test Obs.	REITs
Diversified / Mixed	OLS	0.1720	0.2166	0.1645	752	18
	Gradient Boosting	0.0279	0.2347	0.1866	752	18
	Random Forest	-0.0090	0.2391	0.1879	752	18
	XGBoost	0.0793	0.2284	0.1775	752	18
Health Care	OLS	-0.1334	0.3087	0.2090	466	13
	Gradient Boosting	-0.0237	0.2934	0.2040	466	13
	Random Forest	0.0122	0.2882	0.1919	466	13
	XGBoost	0.1201	0.2720	0.1795	466	13
Industrial / Data Centers	OLS	0.0027	0.1738	0.1375	449	10
	Gradient Boosting	-0.4024	0.2061	0.1580	449	10
	Random Forest	-0.1356	0.1854	0.1427	449	10
	XGBoost	-0.0830	0.1811	0.1418	449	10
Lodging / Hospitality	OLS	-1.1355	0.5877	0.3800	554	13
	Gradient Boosting	0.0619	0.3895	0.2817	554	13
	Random Forest	0.1366	0.3737	0.2694	554	13
	XGBoost	0.0973	0.3821	0.2699	554	13
Office	OLS	-0.4062	0.3234	0.2134	640	15
	Gradient Boosting	-0.0609	0.2809	0.2206	640	15
	Random Forest	0.0524	0.2655	0.2082	640	15
	XGBoost	0.0035	0.2722	0.2134	640	15
Residential	OLS	0.1488	0.2433	0.1680	750	17
	Gradient Boosting	-0.1230	0.2794	0.2238	750	17
	Random Forest	-0.1840	0.2869	0.2270	750	17
	XGBoost	-0.0031	0.2641	0.2035	750	17
Retail	OLS	-0.0127	0.3783	0.2780	1,184	32
	Gradient Boosting	0.1124	0.3542	0.2445	1,184	32
	Random Forest	0.0755	0.3615	0.2469	1,184	32
	XGBoost	0.1195	0.3528	0.2442	1,184	32
Self-Storage	OLS	0.0390	0.1880	0.1483	224	5
	Gradient Boosting	-0.0059	0.1924	0.1627	224	5
	Random Forest	0.0910	0.1829	0.1546	224	5
	XGBoost	-0.0007	0.1919	0.1612	224	5
Specialty / Other	OLS	-0.0613	0.3832	0.2616	390	10
	Gradient Boosting	-0.1712	0.4026	0.2750	390	10
	Random Forest	-0.0628	0.3835	0.2632	390	10
	XGBoost	-0.1853	0.4050	0.2718	390	10

Notes: Table 11 reports six-month cumulative RETX forecasting performance separately across equity REIT property sectors. Each model family is estimated within sector using the chronological estimation/validation/test design. All model-family results are shown; no ex post best-model choice from the 2020–2023 test period is used for formal sector inference. Because several sectors contain few firms, Table 13 provides the confirmatory analysis using a pre-specified XGBoost specification with bootstrap confidence intervals and forecast-loss tests.

Table 12. Out-of-Sample Forecasting Performance with Firm Fixed Effects**Panel A. Pooled OLS versus Firm Fixed-Effects OLS**

Forecast Horizon	Model	Test Obs.	REITs	OOS R²	RMSE	MAE
1-Month Ahead	Pooled OLS	7,239	169	-0.1550	0.1475	0.0979
1-Month Ahead	Firm-FE OLS	7,239	169	-0.2581	0.1540	0.1029
2-Month Ahead	Pooled OLS	7,162	166	-0.0038	0.1375	0.0899
2-Month Ahead	Firm-FE OLS	7,162	166	-0.0362	0.1397	0.0931
3-Month Ahead	Pooled OLS	7,155	166	-0.1045	0.1440	0.0949
3-Month Ahead	Firm-FE OLS	7,155	166	-0.0536	0.1406	0.0926
3-Month Cumulative	Pooled OLS	7,154	166	-0.0432	0.2251	0.1531
3-Month Cumulative	Firm-FE OLS	7,154	166	-0.1757	0.2390	0.1618
6-Month Cumulative	Pooled OLS	7,079	165	-0.0101	0.3099	0.2204
6-Month Cumulative	Firm-FE OLS	7,079	165	0.0306	0.3036	0.2165

Panel B. Six-Month Comparison on the Common Firm-FE-Eligible Test Sample

Model	OOS R²	RMSE	MAE
Pooled OLS	-0.0101	0.3099	0.2204
Firm-FE OLS	0.0306	0.3036	0.2165
Gradient Boosting	0.0968	0.2930	0.2079
Random Forest	0.0667	0.2978	0.2114
XGBoost	0.0949	0.2933	0.2060

Notes: Firm effects are estimated using only pre-test observations. Test-period REITs without sufficient pre-2020 history to estimate an ex ante fixed effect are excluded, and pooled OLS is evaluated on the same restricted observations. At the six-month horizon, 7,079 of the 7,711 baseline test observations are eligible; 632 observations from 24 REITs are excluded. Panel B evaluates the frozen tree models on that identical 7,079-observation test sample. R²_OOS is measured relative to the historical mean estimated from pre-test data.

Table 13. Statistical Inference for Six-Month XGBoost Forecasts by Equity REIT Sector**Panel A. Forecast Performance and Diebold–Mariano Tests**

Property Sector	Test Obs.	REITs	XGB OOS R ²	OLS OOS R ²	DM p vs OLS	DM p vs Hist. Mean
Diversified / Mixed	752	18	0.0945	0.1720	0.556	0.231
Health Care	466	13	0.1148	-0.1334	0.256	0.002
Industrial / Data Centers	449	10	-0.0881	0.0027	0.706	0.356
Lodging / Hospitality	554	13	0.0683	-1.1355	0.227	0.619
Office	640	15	0.0041	-0.4062	0.514	0.919
Residential	750	17	0.0034	0.1488	0.170	0.999
Retail	1,184	32	0.1029	-0.0127	0.261	0.118
Self-Storage	224	5	0.0163	0.0390	0.877	0.865
Specialty / Other	390	10	-0.1678	-0.0613	0.508	0.043

Panel B. Bootstrap 95% Confidence Intervals for XGBoost OOS R²

Property Sector	95% Month-Block CI	95% REIT-Cluster CI
Diversified / Mixed	[-0.0451, 0.1976]	[0.0210, 0.1576]
Health Care	[0.0884, 0.1381]	[0.0890, 0.1767]
Industrial / Data Centers	[-0.2636, 0.0650]	[-0.1683, -0.0267]
Lodging / Hospitality	[-0.2017, 0.2465]	[-0.0555, 0.1792]
Office	[-0.1274, 0.1095]	[-0.0155, 0.0257]
Residential	[-0.1161, 0.0859]	[-0.0439, 0.0313]
Retail	[0.0244, 0.2219]	[0.0837, 0.1304]
Self-Storage	[-0.2667, 0.2085]	[-0.1791, 0.1153]
Specialty / Other	[-0.3879, -0.0350]	[-0.2182, -0.0012]

Notes: Table 13 evaluates the pre-specified six-month XGBoost forecasting model across equity REIT sectors. The XGBoost specification and hyperparameters are fixed before the 2020–2023 test period and are not re-tuned by sector using test observations. R² OOS is calculated relative to each sector's pre-test historical mean. Diebold–Mariano-type tests use squared forecast-loss differentials averaged across REITs within each target-end month, HAC(5) adjustment for overlapping six-month returns, and the Harvey–Leybourne–Newbold small-sample correction. Two-sided p-values are reported. Month-block intervals use 5,000 circular moving-block bootstrap replications with six-month blocks; REIT-cluster intervals use 5,000 resamples of whole REIT histories.

Appendix Table A1. Hyperparameter Candidate Grid and Validation Selections

Panel A. Candidate Grid

Model	Candidate 1	Candidate 2	Candidate 3
Gradient Boosting	iter=100; lr=.03; max	iter=200; lr=.03; max	iter=100; lr=.05; max
	leaves=15; min leaf=20; L2=.1	leaves=15; min leaf=20; L2=.1	leaves=15; min leaf=20; L2=.1
Random Forest	trees=100; depth=6; min	trees=100; depth=10; min	trees=100; depth=None; min
	leaf=25; max features=.8	leaf=25; max features=.8	leaf=25; max features=.8
XGBoost	trees=100; depth=2; lr=.03; min	trees=200; depth=2; lr=.03; min	trees=100; depth=3; lr=.03; min
	child=5	child=5	child=5

Panel B. Validation-Selected Candidate by Forecast Horizon (validation RMSE in parentheses)

Horizon	Gradient Boosting	Random Forest	XGBoost
1-Month Ahead	1 (0.073181)	1 (0.072551)	3 (0.071798)
2-Month Ahead	1 (0.072980)	1 (0.071610)	2 (0.071526)
3-Month Ahead	1 (0.073119)	1 (0.071441)	1 (0.071223)
3-Month Cumulative	1 (0.123631)	2 (0.126028)	3 (0.122443)
6-Month Cumulative	1 (0.217668)	2 (0.253247)	1 (0.179044)

Notes: Hyperparameters are selected using only 2016–2019 validation-period RMSE. Selected settings are then frozen before refitting on 2000–2019 and evaluating 2020–2023. For XGBoost, subsample = 0.8, colsample_bytree = 0.8, and L2 regularization (reg_lambda) = 1 are fixed across candidates. Random seed 42 is used for stochastic models. No test-period performance is used in hyperparameter selection.

Appendix Table A2. Pre-2020 Pseudo-Out-of-Sample Forecasting Performance

Panel A.

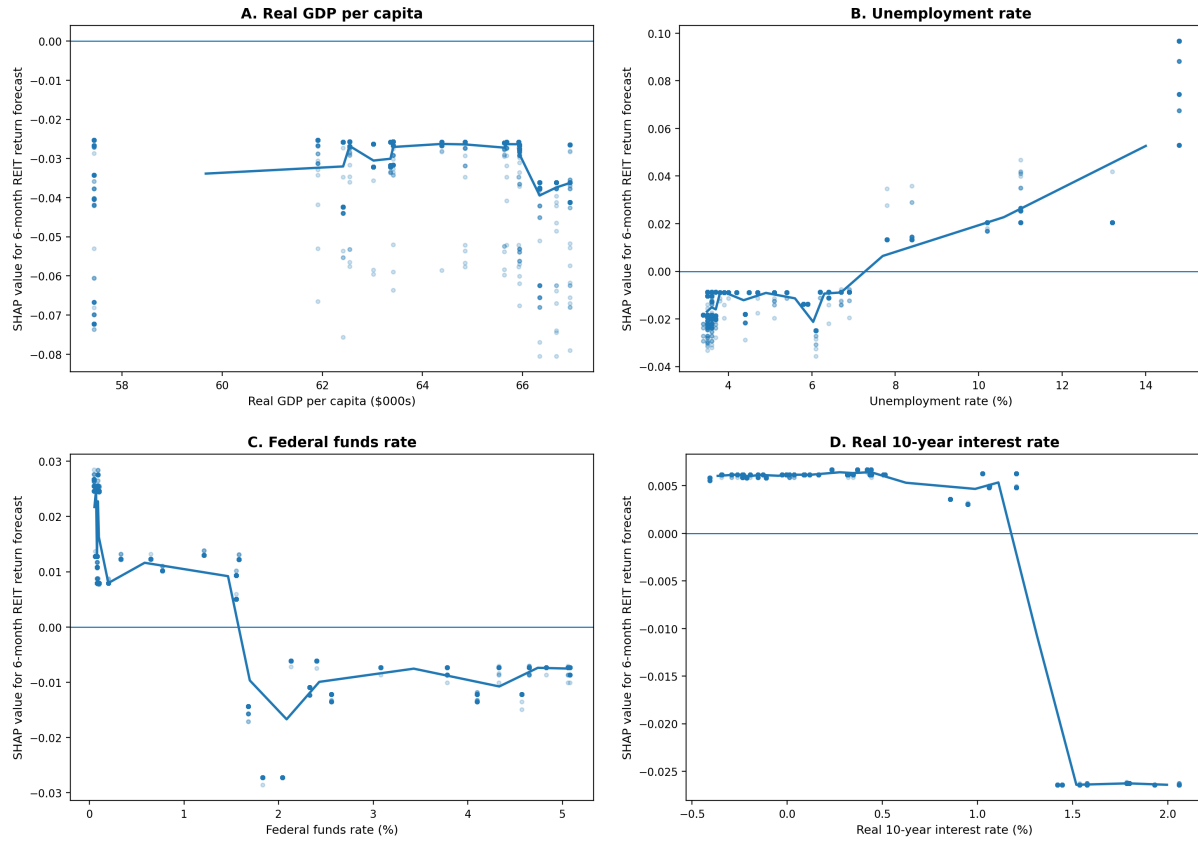
Horizon	Model	OOS R^2	RMSE	MAE	Test N
3-month cumulative	OLS	-2.1237	0.2039	0.1715	6,268
	Gradient Boosting	-0.1082	0.1215	0.0894	6,268
	Random Forest	-0.0219	0.1166	0.0843	6,268
	XGBoost	-0.1712	0.1249	0.0921	6,268
6-month cumulative	OLS	-2.9709	0.3332	0.2862	6,148
	Gradient Boosting	-0.3043	0.1910	0.1415	6,148
	Random Forest	-0.8304	0.2262	0.1795	6,148
	XGBoost	-0.3043	0.1910	0.1399	6,148

Panel B. The validation-selected candidates

Horizon	Gradient Boosting	Random Forest	XGBoost
3-month	Candidate 1	Candidate 2	Candidate 3
6-month	Candidate 1	Candidate 3	Candidate 3

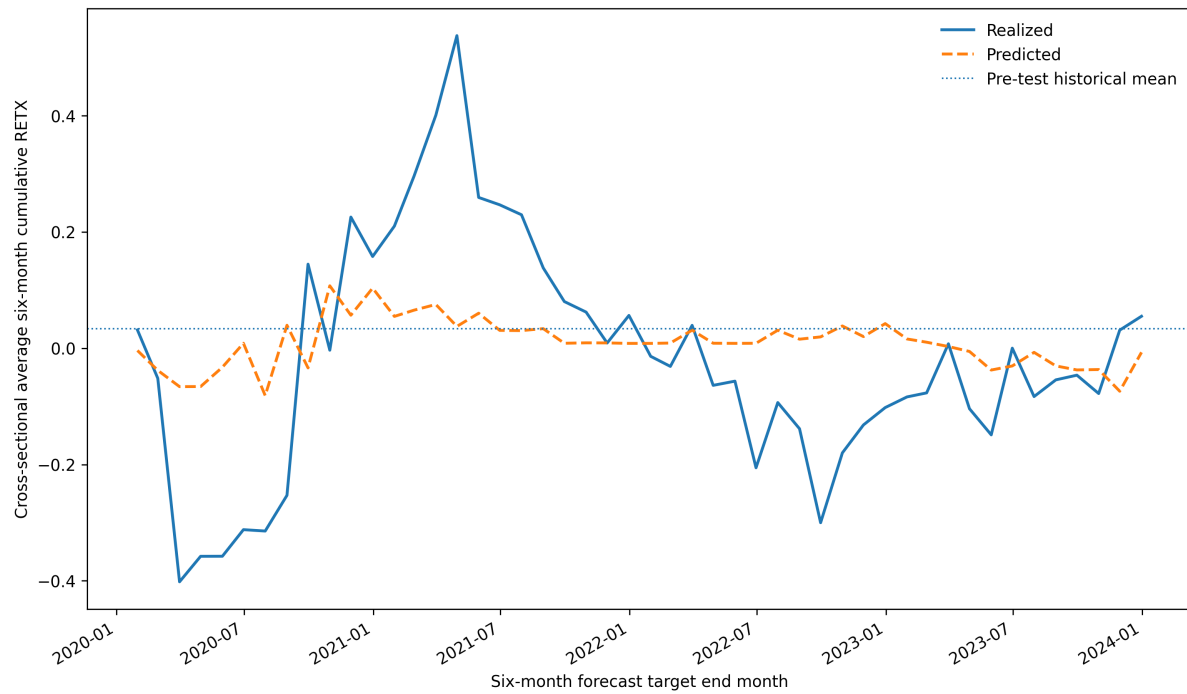
Notes: Table A2 reports a pre-2020 pseudo-out-of-sample robustness exercise for cumulative RETX forecasts. Models are initially estimated using observations with forecast target-end dates through 2010. Hyperparameters are selected exclusively using 2011–2013 validation-period RMSE and are then frozen before each model is refit through 2013. Forecast target-end months from 2014–2016 form the pseudo-out-of-sample test period. All models use the same predictor set and candidate hyperparameter grids as the main analysis. R^2_{OOS} is measured relative to the historical-mean return estimated from pre-test observations. Forward targets and return lags require consecutive calendar months. No pseudo-test observations are used for model selection.

Figure 1. SHAP Dependence for Leading Macroeconomic Predictors



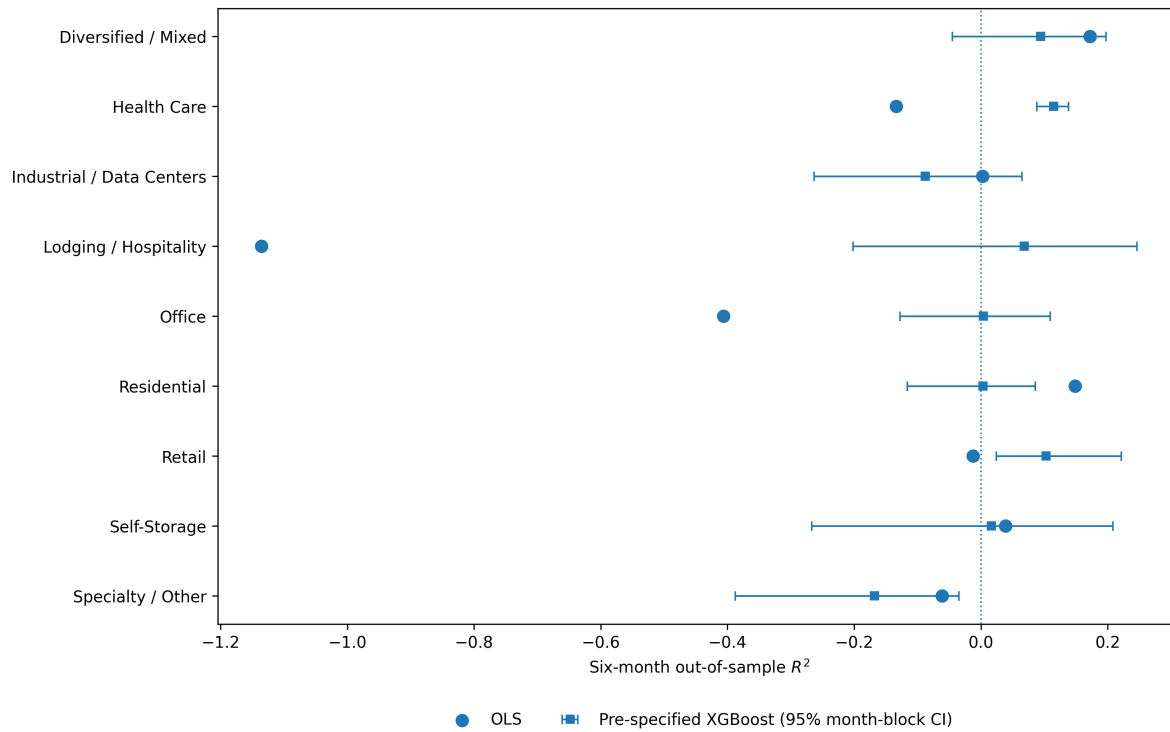
Notes: Figure 1 illustrates the nonlinear relationship between the four leading macroeconomic predictors and the six-month cumulative REIT return forecasts generated by the XGBoost model. Each point represents an out-of-sample firm-month observation from 2020–2023. The horizontal axis reports the observed value of the predictor, while the vertical axis reports that variable's SHAP contribution to the predicted six-month REIT return. Positive SHAP values indicate that the predictor increases the model's forecast relative to its baseline prediction, whereas negative SHAP values indicate that the predictor lowers the forecast. The fitted lines summarize the average nonlinear pattern in the SHAP values.

Figure 2. Realized versus Predicted Six-Month REIT Returns, 2020–2023



Notes: Figure 2 compares the cross-sectional average realized six-month cumulative RETX with the corresponding average XGBoost forecast over the 2020–2023 out-of-sample period, indexed by forecast target-end month. The dotted horizontal line is the pre-test historical mean. The realized/predicted monthly correlation is 0.530; the 95% confidence interval, based on 5,000 circular moving-block bootstrap replications with six-month blocks, is approximately [0.07, 0.71]. Predictions are substantially smoother than realizations; supplementary calibration tests show that this shrinkage does not fully account for the model's aggregate forecast improvement.

Figure 3. Sector-Level Out-of-Sample Forecasting Performance



Notes: Figure 3 compares six-month out-of-sample R^2 for OLS and the pre-specified XGBoost model across equity REIT property sectors. XGBoost is not selected ex post as the best model within each sector. Horizontal error bars are 95% circular moving-block bootstrap confidence intervals based on six-month blocks of forecast target-end months. Positive R^2 values indicate improvement relative to the sector-specific pre-test historical mean.