

RESEARCH

Open Access



Predicting the returns of the US real estate investment trust market: evidence from the group method of data handling neural network

Wendi Zhang^{1*} , Bin Li¹, Alan Wee-Chung Liew², Eduardo Roca¹ and Tarlok Singh¹

*Correspondence:
Jerry.zhang2@griffithuni.edu.au

¹ Department of Accounting,
Finance and Economics,
Griffith Business School, Griffith
University, Nathan Campus,
Brisbane, QLD, Australia

² Institute for Integrated
and Intelligent Systems, Griffith
University, Nathan, Brisbane,
QLD, Australia

Abstract

Purpose: The Group Method of Data Handling (GMDH) neural network has demonstrated good performance in data mining, prediction, and optimization. Scholars have used it to forecast stock and real estate investment trust (REIT) returns in some countries and region, but not in the United States (US) REIT market. The primary goal of this study is to predict the US REIT market using GMDH and then compare its accuracy with that derived from the traditional prediction method.

Design/methodology/approach: To forecast the return on the US REIT index, this study used the GMDH neural network and the generalized autoregressive conditional heteroscedasticity (GARCH) model. In this test, the training samples, testing samples, and kernel functions of the GMDH model are controlled to investigate their impact on the accuracy of the machine learning approach. Corresponding experiments were performed using the GARCH model, and the accuracies of these two approaches were compared.

Findings: Compared with GARCH, GMDH's accuracy is much higher, indicating that the machine learning approach can provide a highly accurate prediction of REIT prices. The size of the training samples and the kernel functions in the GMDH model affect the accuracy of the prediction results. In particular, the kernel function has a significant impact on prediction accuracy. The linear and linear covariance kernel functions are simple to train and yield accurate predictions, whereas the quadratic function is difficult to train. Even with small training samples, GMDH can outperform GARCH in prediction accuracy.

Research limitations/implications: Although GMDH shows good performance in predicting the US REIT return, it is still a black-box model, and the algorithm is difficult for financial analysts to develop and customize. The data used in this study come from the US REIT market, which is the world's largest and most liquid market.

Social implications: This research shows that the GMDH model outperforms the GARCH model in forecasting REIT returns. Hence, investors can use the machine learning approach to make more accurate predictions of the target REITs' returns and

thus better investment decisions. Future investors and researchers may use GMDH to forecast the performance of REITs in other markets.

Originality/value: This is the first study to apply the GMDH neural network to the US REIT market and determine the impact of the two factors on its performance. For example, this research first discusses the impact of kernel functions on the US REIT market using the GMDH neural network. It also includes short-term daily prediction returns that were not previously considered, making it a valuable reference for financial industry analysts.

Keywords: REIT prediction, Machine learning, GMDH, GARCH, Accuracy

Introduction

REIT is the shorthand for real estate investment trust, which is an open-end trust run by a company that allows people to invest in real estate in a non-physical purchase. The US Congress built the first REIT and introduced four main regulations to protect investors and regulate REIT. First, REIT contains at least 75% of assets in real estate, government bonds, and cash. Second, they must have more than 50% of the total shareholding of more than five people combined. Third, more than 75% of income should come from renting or selling real estate, or from mortgage interest. Fourth, at least 95% of taxable income must be distributed as dividends each year (Irem et al. 2020; Brueggeman & Fisher 2022).

Based on their characteristics, REITs are classified into three types: equity-REIT (EREIT), mortgage-REIT (MREIT), and hybrid-REIT (HREIT). EREIT is a publicly traded company whose primary business is the acquisition, management, renovation, maintenance, and, in some cases, sale of real estate properties. MREITs issue and hold loan and other debt instruments backed by real estate, and their dividend yield is generally higher than MREIT. Meanwhile, HREIT combines the features and operations of EREIT and MREIT (Block 2012; Hansz et al. 2017).

REITs have become a popular financial product over the last half-century, and more than 40 countries have now established REIT markets to securitize their real estate assets. The global REIT market is now worth over \$2 trillion. REITs have become a vital investment option for many investors today. Ott et al. (2005) indicated that REITs have become one of the most popular investment options for investors, and the market has reached maturity in many markets and regions outside the United States, such as Australia and Canada. Furthermore, because the majority of REIT earnings are distributed to investors, they are considered a more profitable financial product than common stock portfolios (Mori & Ziobrowski 2011). A new investment trend is to construct investment portfolios by integrating common stocks and REITs; such activities increase investment portfolios' diversification and profitability (Anderson et al. 2015). REIT selection and return prediction, like stocks, are two important strategies for REIT investment that affect principal and investment returns and help investors optimize their portfolio returns (Lee & Pai 2010).

REIT prices are important in the investment process. Predicting the returns of various REIT products can assist in determining whether they should be added to the portfolio; additionally, it can assist in determining the best trading operations, such as buying, holding, or selling these REIT products (Cici et al. 2011). Researchers

have attempted to develop and apply various traditional and machine learning (ML) approaches to improve the selection and price prediction of REIT products.

The fintech and ML approaches have become popular across diverse research areas, such as stock market analysis, users' behavior identification, biometrics, customer service, cyber defense, bankruptcy prediction, and peer-to-peer network analysis (Livingston 2005; Maxwell et al. 2018; Kou et al. 2021a, b; Li et al. 2022). Conventional methods have a fundamental difference from the ML approach in that they are model-based and rely on a parametric model constructed based on domain knowledge. For example, the underlying process of the GARCH model is a plausible mathematical model. The data are then used to fit the parametric model to find the optimal parameter settings. By contrast, ML is purely data-driven and non-parametric. A learning algorithm (support vector machine (SVM), neural network, etc.) is selected to apply to the data to learn a function that best maps inputs to outputs. There is no requirement that this learned function be a close approximation of the true (but unknown) function underlying the data—in most cases, one simply chooses a convenient ML algorithm that is expressive enough and then uses data to learn an ML model that can best map inputs to outputs on all possible observations in the training set (Dixon et al. 2020).

Nevertheless, the ML method in REIT prediction remains relatively new. Existing studies mainly focused on using the ML approach to construct portfolios for REIT investment. This method can forecast REIT performance in the future and assist investors in making better investment decisions to optimize their REIT portfolio or control the proportion of REIT in multi-asset portfolios. For example, Li et al. (2017) studied the performance of the GMDH to predict the long-term return trend of REITs, whereas Loo (2019) used an artificial neural network (ANN) to estimate the Hong Kong REIT market's future performance.

The Group Method of Data Handling (GMDH) is used in this study to forecast US REIT returns. Using the GMDH has several advantages. First, the GMDH relies less on the input because of its self-organization programming. Second, GMDH automatically determines the number of hidden layers, ensuring that the training process selects the most relevant variables (Li et al. 2017). In their study related to the electricity market, Yang et al. (2018) found that GMDH can process complex datasets in a highly effective manner in forecasting the daily electricity usage. Similarly, Srinivasan (2008) claimed that GMDH shows more precise prediction results than traditional regression models in short-to-medium-term energy consumption.

The following factors motivate our research. First, REIT investment has grown in popularity; thus, investigating the effectiveness of ML in REIT prediction is worthwhile, as it is one of the most popular data mining techniques in recent years. Second, when compared with traditional data analysis techniques, the application of ML in REIT investment is still in its early stages. Although many scholars have studied the implementation of ML in REIT selection in recent years, the accuracy of the ML technique in REIT prediction remains questionable (Li et al. 2017; Loo 2019). Thus, it is worthwhile to compare the accuracy of ML and traditional models. Third, when dealing with systemic financial risk, such as the one caused by COVID-19, REIT is an important asset. The outbreak of COVID-19 causes a setback for the REIT and the stock market, with both losing 30% of their market capitalization in a short time (Hui & Chan 2022). However, stock indexes,

such as S&P 500 and Dow Jones Industrial Average, which took six months to complete a U-shaped recovery, it took only four months for the REIT markets to bounce back. REITs seem to have a stronger ability to hedge the systematic financial risk and thus are valuable in portfolio diversification.

We contribute to the literature by introducing a new method for predicting the US REIT market, despite the fact that a previous study applied GMDH to REIT in some non-US markets and showed that it is an effective tool for REIT prediction (Li et al. 2017). Nevertheless, the previous study's results may not be representative because the GMDH algorithm was not applied to the US REIT, which constitutes 2/3 of the REIT market in the world. The US REIT is important to the global REIT market not only because of its market share, but also because it is the country with the largest variety of REITs. Furthermore, our study discusses in detail the impact of different kernel functions on the GMDH neural network for predicting the US REIT market. Our out-of-sample test yields day-by-day prediction results using various kernel functions. The discovery of significant differences in prediction accuracy between kernel functions for different prediction horizons is novel in the literature.

Furthermore, based solely on the GMDH model's analysis, it is difficult to ascertain whether the GMDH model outperforms traditional models in predicting REIT returns (Li et al. 2017; Loo 2019). To ensure the robustness of the evidence, this study extends the analysis and compares the prediction of the GMDH model with that of the traditional GARCH model.

The remainder of the study is organized as follows: "Literature review: REIT prediction with traditional and ML methods" section reviews the literature and summarizes empirical studies conducted over the last two decades in both traditional and ML settings. "Model specifications and data" section presents the GMDH and GARCH models and discusses the variables and data used in the study. "Empirical results and analysis" section presents and analyzes the empirical results. "Discussion" section provides a discussion. Finally, "Conclusions" section concludes the paper.

Literature review: REIT prediction with traditional and ML methods

Previous research has applied and developed many different traditional and ML methods to predict the REITs' return and price. Traditional methods dominated the REIT price and return studies before 2007.

An earlier study notes that a simple linear regression model is difficult to obtain good out-of-sample testing results in the US REIT return prediction (Ling et al. 2000). Researchers use multifactor models to respond, and they find that REIT return is highly correlated with dividend, return history, price momentum, liquidity, and profitability factors (Chui et al. 2003; Cheng & Roulac 2007; Olanrele et al. 2014). Furthermore, Li & Lei (2011) found that the macroeconomy plays a significant role in determining REIT's performance; meanwhile, the REIT market's performance has an impact on the macroeconomy. Using Fama–Macbeth approach, Shen (2021) determined that distressed REIT securities earn high future returns, whereas Shen et al. (2021) showed that low-beta REITs deliver a significantly higher risk-adjusted return than high-beta REITs. Swinkels (2023) noted that the real estate index has a significant impact on the value of real estate tokens, which are digital tokens similar to REITs.

The vector autoregression model is another traditional model commonly used in REIT research. According to Lu and So (2001), the REIT return is negatively related to the economy's inflation rate. Ling and Naranjo (2006) found that new fund flow has a greater effect on REIT price than trading volume. Moreover, some researchers attempt to predict REIT returns using publicly available information; for example, Ling and Naranjo (2015) used publicly available information to predict REIT returns, but their results are unsatisfactory. Sirmans et al. (2006) reported that management changes cannot be used to predict REIT performance. According to Siew (2015), random events may cause a shift in the development trend of the Australian REIT market. Although many researchers have used traditional approaches to study REIT, they acknowledge that ML may be a more efficient method of research (Cheng & Roulac 2007; Siew 2015). Several researchers have also advocated for the use of ML and emerging algorithms in financial analysis (Kou et al., 2022).

Since 2007, ML has been widely used in REIT research. Lertwachara (2007) compared the return of a randomly chosen portfolio with that of an ANN-based portfolio; the ANN outperforms the randomly chosen portfolio by 26.05% higher return. Feng & Li (2014) created an REIT portfolio with 20% above market returns via an SVM. Wang et al. (2016) employed SVM and back-propagation neural network in the Singaporean REIT market; they report that ML approaches predict REIT performance more accurately than simple regression models, and that the selection of variables affects the simulation outcomes. According to Li et al. (2017), the GMDH neural network can accurately predict REIT performance in following countries and region: Australia, Hong Kong, Italy, and Turkey. Furthermore, Hausler et al. (2018) captured and analyzed investor sentiment via the SVM; they reported that, in addition to financial and environmental factors, news-based sentiment has a significant impact on the returns of REITs and security markets.

The review finds that ML methods can improve portfolio selection as well as REIT performance prediction. The return of portfolios constructed using ML approaches is generally higher than the return of portfolios constructed using traditional methods for the construction of REIT portfolios (Le 2006; Lertwachara 2007; Feng & Li 2014; Loo 2020). In terms of REIT return prediction, the accuracy of results obtained from ML models is higher than that obtained from non-ML methods (Wang et al. 2016; Li et al. 2017; Loo 2019; Antunes 2021). In summary, this review suggests that it is still a relatively new area of research to use ML techniques to predict REIT returns. Most researchers continue to focus on using ML to build portfolios for REIT investment.

However, the review of the literature reveals some challenges in implementing ML approaches in REIT prediction. First, the GMDH neural network can predict the return of the REIT in four countries and region in a timely and accurate manner (Li et al. 2017). However, the study overlooked the US REIT market, which is the world's largest. The US Congress created the world's first REIT in the 1960s and established REIT-related regulations. Other countries and regions have established REITs in accordance with US regulations. The US REIT market capitalization reached \$1,352.41 billion at the end of 2021, accounting for 67.4% of the global market. In contrast, the total capitalization of REITs in these four markets is only \$124.59 billion, accounting for only 6.21% of the global market (EPRA 2022). The REIT market in the United States is more iconic and

representative of the global REIT trend. As a result, the GMDH neural network should be assessed on the US REIT market. Second, previous research has not addressed the impact of the sample size used in ML on the accuracy of REIT prediction results. Time-series model research shows that REIT prediction results are related to regression analysis datasets. In this case, the study seeks to investigate whether this effect exists in ML approaches.

Based on the above reviews and discussions, we developed the following two testable hypotheses:

Hypothesis 1. The GMDH neural network predicts the US REIT market more efficiently and accurately than the GARCH model.

Hypothesis 2. GMDH neural network can predict both short- and long-term returns in the US REIT market.

Model specifications and data

This section describes how GMDH, an ANN-based technique, predicts US REIT prices. It goes over the definitions of the variables as well as the data sources used to estimate the model. The traditional model for predicting REIT returns is the generalized autoregressive conditional heteroscedasticity (GARCH). The results of GMDH and GARCH were compared to assess the relative effectiveness of these two techniques in predicting REIT returns.

The ANN model

Theoretical framework of the ANN model

The general expression of the output of a neuron j is given by:

$$y_j = f(\text{net}_j)$$

$$\text{net}_j = \sum_n w_{jn}x_n + w_{j0} = \mathbf{w}_j\mathbf{x} + b \quad (1)$$

where \mathbf{x} is a vector of inputs to the j th neuron, b the bias term, y_j is the output of the j th neuron, net the network and \mathbf{w} the weights. Figure 1 shows the fundamental structure of a neuron.

The feedforward structure, which includes the input layer, multiple hidden layers, and the output layer, is the most common type of ANN. Each layer's neuron receives signals from the previous layer and generates signals for the next layer. The first and last layers are referred to as the input and output layers, respectively, while the other intermediate layers are referred to as the hidden layers. The neurons in two adjacent layers are completely linked (Hornik et al. 1989; Svozil et al. 1997; Liu & Chen 2020).

The training process starts the ANN with random weights and iteratively updates the weights using the back-propagation (BP) algorithm. The BP algorithm propagates the error computed at the output during training back through the layers, adjusting the weight of each neuron accordingly until the network converges to a local optimum. Following the completion of the training process, the network's performance is evaluated

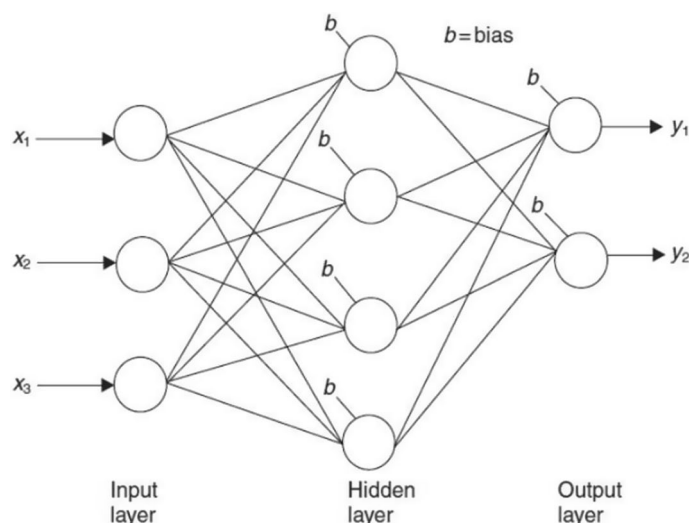


Fig. 1 Basic structure of a neuron

using the testing dataset. ANNs have been demonstrated to be effective algorithms and are widely used in complex nonlinear function mapping, image processing, and pattern recognition (Svozil et al. 1997).

Our study employs GMDH, as proposed by Li et al. (2017), to forecast US REIT returns. GMDH has a number of advantages that make it appealing for REIT research. First, the self-organization and control characteristics of GMDH ensure that the output has few effects from the input. Second, because GMDH algorithms do not require many assumptions, they are relatively easy to develop. Third, the GMDH method enables each layer to obtain the best structure in the training process to find the most relevant variables and eliminate irrelevant variables automatically. Fourth, GMDH can automatically determine the number of network layers and the neurons contained within them (Li et al. 2017).

The structure of the GMDH neural network is depicted in Fig. 2. The GMDH network is made up of a series of identical modules that are linked together. The computing process of one module, for example, the process from $x_i^{(k)}$ to $y_i^{(k)}$ can be explained as follows. The $x_i^{(k)}$ is a series of input variables and $y_i^{(k)}$ is a series of the output of the k -th module. From $x_i^{(k)}$ to $y_i^{(k)}$, a kernel function is selected to connect the input variables and the outputs. In general, the kernel function can be expressed as $y = f(x_1, x_2, \dots, x_i, \dots, x_j, \dots, x_n)$. In this equation, x_i and x_j indicate the input variables, and n represents the number of variables. The three kernel functions used are as follows: a linear function, a linear covariance function, and a quadratic function. With various models, these kernel functions can resolve underfitting and overfitting issues. The most commonly used kernel function for GMDH is shown in Eq. (4). Here, w represents the weights assigned to these variables during the process of obtaining outputs from the input layer to the output layer.

$$y = w_0 + w_1x_i + w_2x_j \tag{2}$$

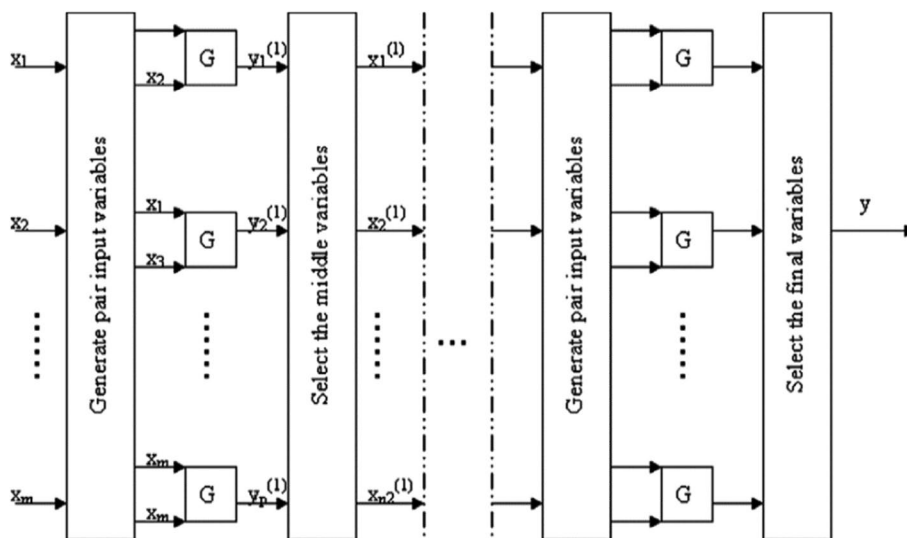


Fig. 2 Structure of the GMDH model

$$y = w_0 + w_1x_i + w_2x_j + w_3x_ix_j \tag{3}$$

$$y = G(x_i, x_j) = w_0 + w_1x_i + w_2x_j + w_3x_ix_j + w_4x_i^2 + w_5x_j^2 \tag{4}$$

The results are labeled as *G* in Fig. 2. To select the appropriate variables from the intermediate outputs to feed to the next module, a filtering procedure is used. When the appropriate variables are chosen, the next module’s calculation repeats the operations of the previous module until the final outputs are obtained.

Variables used in the ANN model

Several variables are directly inputted into the training process, allowing the GMDH method to select the appropriate variables. Table 1 defines the variables used in the training test. This empirical analysis takes into account four types of variables: volatility, trend, momentum, and return rates. These variables are, in general, standard technical indicators used in the stock market to forecast stock prices. They were chosen to forecast the return on the REIT index in the US market. “Appendix 1” contains detailed definitions and explanations of these variables.

Computation pipeline

The ML method’s computation pipeline is divided into six steps: data collection, technical indicator measurements, training, test, accuracy evaluation, and discussion. Each step is depicted in Fig. 3.

GARCH model

The GARCH model was developed around 40 years ago (Bollerslev 1986). Previous studies compared the GARCH model’s performance in predicting stock returns and volatility with that of other traditional financial models, such as the Capital Asset Pricing Model (CAPM) and the Stochastic Volatility Model. Prior studies have demonstrated

Table 1 List and definitions of the variables used in GMDH neural network

Category	Abbreviation	Variable	Description
Volatility	ATR	Average True Range	An indicator used to measure the volatility of the prices in the past two weeks
	BB	Bollinger Bands	An indicator related to the volatility of the price, proposed by John Bollinger, contains three-track lines, high (BBH), low (BBL), and median (BBM)
	KC	Keltner Channel	An indicator related to the volatility of the price, proposed by Chester W. Keltner, contains three-track lines, high (KCH), low (KCL), and center (KCC)
	DC	Donchian Channel	An indicator related to the volatility of the price, proposed by Richard Donchian, includes three track lines, high (DCH), low (DCL), and median (DCM)
Trend	MACD	Moving Average Convergence Divergence	A typical indicator that illustrates the securities and stocks' price trend, guiding investors when to buy and sell. It shows the strength of the stock and securities to investors and the changing orientation of the prices
	ADX	Average Directional Movement Index	An indicator that measures the average values of a stock or security's prices with an expanding price range. Overall, it illustrates the possible change trends' strength, letting investors know whether the price is more likely to rise or decrease
	VI	Vortex Indicator	An indicator to present the changing trend of the prices of stocks. It can identify the prices' current trends or identify the current trends' reversal
	MI	Mass Index	An indicator related to the trend is mainly used to identify the current price trend's possible reversals
	CCI	Commodity Channel Index	An indicator that is used to evaluate the trend of stock prices and the strength of the trend. Simultaneously, it can also be used to identify the scenario of being oversold or overbought
	DPO	Detrended Price Oscillator	An indicator related to the trend is used to evaluate the change cycle of a stock's prices
	PSR	Parabolic Stop and Reverse	An indicator that is similar to VI, which identifies the potential trend or reversals in the prices
	Return	DR	Daily Return
DLR		Daily Log Return	The logarithm value of the daily return rate
CR		Cumulative Return	The return rate of stock in a period

Table 1 (continued)

Category	Abbreviation	Variable	Description
Momentum	MFI	Money Flow Index	An indicator that is used to identify the signals related to oversold and overbought based on the trading volume and prices
	RSI	Relative Strength Index	An indicator related to momentum, measuring the changing speed of the prices of stocks
	TSI	True Strength Index	An indicator that is used to illustrate the changing orientation of the prices
	UO	Ultimate Oscillator	An indicator that is used to identify the momentum of the price of an asset in different periods
	SR	Stochastic Oscillator	An indicator related to the momentum that identifies the gaps between the close price and the average price in a particular period
	WR	Williams %R	An indicator that is used to identify an asset's close price to judge the oversold and overbought conditions
	AO	Awesome Oscillator	An indicator that is used to measure the overall momentum of the stock market
	KAMA	Kaufman's Adaptive Moving Average	An indicator that is developed to identify the level of noise in the market
	ROC	Rate of Change	An indicator that measures the changing rate of the prices of an asset
	TRIX	Triple Exponential Average	An indicator related to momentum, proposed by Jack Hutson, shows the percentages of changes in the triple smoothed moving average. It is used to filter the changes in prices, which are regarded as unimportant, enabling investors to focus on those crucial changes
	KST	Know Sure Thing Oscillator	A momentum oscillator that is used to help investors or stock traders understand the changes in rates. It is based on the measurement of the ROC
	Ichimoku	Ichimoku Kinko Hyo	An indicator developed to monitor the changes of momentum and to identify the resistance and support for future changes

the outperformance of GARCH models over other models (Ng 1991; Fung et al. 2014; Maneemaroj et al. 2021). In addition, the GARCH model supports a significantly positive relationship between risk and return in several markets (Darrat et al. 2011). Therefore, this study employs the GMDH model to forecast US REIT and stock market returns and compares its performance with that of the traditional financial data analysis model, GARCH. The general form of the GARCH model is as follows (Lee & Pai 2010; Cho & Elshahat 2011; Zhou & Kang 2011):

$$r_t = \ln(p_t) - \ln(p_{t-1}) \quad (5a)$$

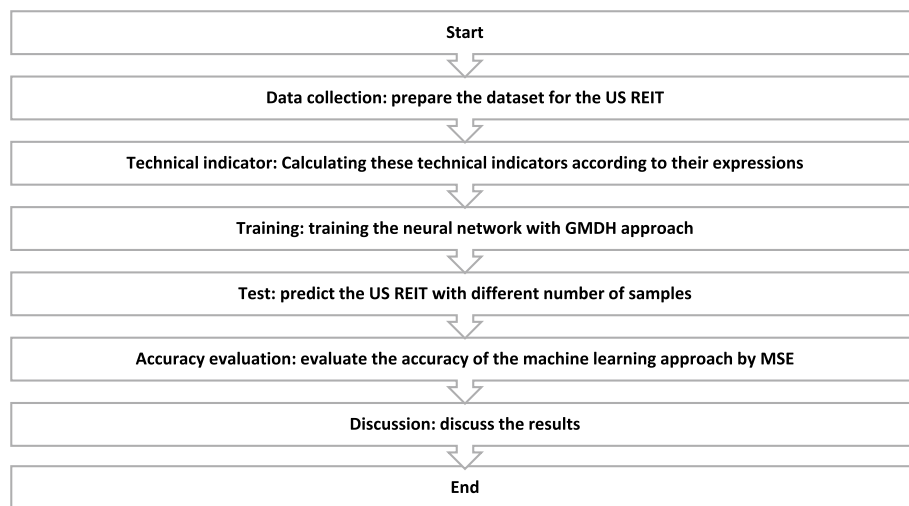


Fig. 3 Computation pipeline of machine learning process

$$r_t = \mu + \alpha r_{t-1} + \varepsilon_t \tag{5b}$$

$$\sigma_t^2 = \omega + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 \tag{5c}$$

The expression (5a) shows how the dependent variable is measured (Lee & Pai 2010; Zhou & Kang 2011). The p_t and p_{t-1} are the prices of the REIT at t and $t-1$, respectively. r_t indicates REIT return at time t , and \ln the logarithm value of price. The daily log return used in the GMDH neural network can be used directly in the estimation process.

The expression (5b) shows how the previous instance’s information affects the REIT return. In this expression, r_{t-1} indicates the return value in the previous instance, and ε_t represents how the returns are innovated; it depends on the return variance and a distribution function. Equation (6a) presents the expression of ε_t (Lee & Pai 2010). According to the expressions shown in Eq. (6), the value of ε_t depends on the variance value and the traditional distribution value at the instance t (Lee & Pai, REIT volatility prediction for skew-GED distribution of the GARCH model, 2010; Zhou & Kang 2011). Here, the density function of the error term z_t is a standard normal distribution, shown in Eq. (6b) (Lee & Pai 2010).

$$\varepsilon_t = \sigma_t z_t \tag{6a}$$

$$f(z_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_t^2}{2}} \tag{6b}$$

Equation (5c) shows the calculation of the variance in the prediction process (Lee & Pai 2010; Zhou & Kang 2011). This equation shows that the variance at time t depends on its previous value and the innovation value in the last instance.

The other variables shown in Eq. (5), including μ , α , β , γ , and ω , are the coefficients obtained from the estimation process. They can be constants (Lee & Pai, REIT

volatility prediction for skew-GED distribution of the GARCH model, 2010), or they can be changed to meet proper distribution function (Cho & Elshahat 2011); this is one way to improve the GARCH model's performance and accuracy. In this empirical study, no distribution functions are used to measure these coefficients, and their values are obtained directly from regression estimation.

When the REIT's return is predicted by Eq. (5b), its price at time t can be recalculated by transforming the Eq. (5b).

$$p_t = p_{t-1}e^{r_t} \quad (7)$$

Data sources

Our daily data come from DataStream and Bloomberg. The US REIT index includes 135 of EREIT, MREIT, and HREIT stocks. These REITs invest in the following industries: residential, industries, retail, health care, offices, hotel & resort, and mortgages. Our variables, such as momentum, volatility, and trend, are calculated from S&P500. The total number of daily observations used in this study was over 2 million.

The sample period for the US REIT daily data is from October 6, 2016, to July 30, 2021. The US REIT index includes 135 of EREIT, MREIT, and HREIT stocks. These REITs invest in the following industries: residential, industries, retail, health care, offices, hotel & resort, and mortgages. The S&P500 is used to calculate our variables, such as momentum, volatility, and trend. The number of training and test samples is controlled to investigate GMDH's accuracy under different scenarios. The training cases considered are 30, 100, 200, 300, and 600 observations, with testing samples of 30 and 60 observations. A sample is an observation at a specific time point in this context. To compare the performance of ML and traditional approaches, we estimate the GARCH model with the same number of observations.

Empirical results and analysis

The three kernel functions mentioned above are tested when using the GMDH model to predict REIT returns. Meanwhile, cases of various training and testing samples are being tested. The accuracy of these cases' results is then compared to determine which case produces the best prediction results. In addition, the GMDH model's results are compared with the GARCH model's results to determine which model produces more accurate REIT prediction results.

Estimates of the GMDH model

Linear function

In this empirical study, 17 cases were tested. Table 2 contains information about these cases, such as the number of training samples, the number of testing samples, the kernel functions used, and the input variables.

The sample period for these cases with 30 testing samples is from May 13, 2021, to June 30, 2021. The sample period for these cases with 60 testing samples is from March 31, 2021, to June 30, 2021. These cases' training periods are the transaction days preceding their testing period. The training period of the case linear-30/30, for example, is the 30 transaction days preceding the testing period.

Table 2 Details of the machine learning cases tested in this study

Case	Training Samples	Testing samples	Kernel function	Input variables
Linear-30/30	30	30	linear function	All variables
Linear-100/30	100	30	linear function	All variables
Linear-200/30	200	30	linear function	All variables
Linear-300/30	300	30	linear function	All variables
Linear-600/30	600	30	linear function	All variables
Linear-300/60	300	60	linear function	All variables
Linear-cov-30/30	30	30	linear covariance function	All variables
Linear-cov-100/30	100	30	linear covariance function	All variables
Linear-cov-200/30	200	30	linear covariance function	All variables
Linear-cov-300/30	300	30	linear covariance function	All variables
Linear-cov-300/60	300	60	linear covariance function	All variables
Quadratic-30/30	30	30	quadratic function	All variables
Quadratic-100/30	100	30	quadratic function	All variables
Quadratic-200/30	200	30	quadratic function	All variables
Quadratic-300/30	300	30	quadratic function	All variables
Quadratic-600/30	600	30	quadratic function	All variables
Quadratic-300/60	300	60	quadratic function	All variables

Note: The linear function, linear covariance function, quadratic function indicates the kernel functions described in Eqs. 2, 3, and 4, respectively

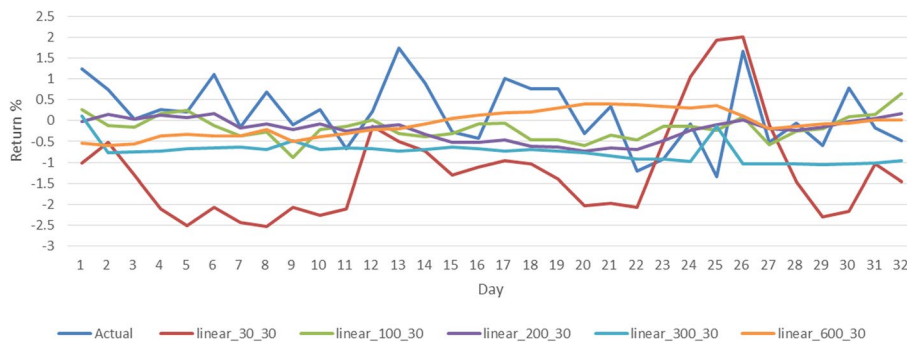


Fig. 4 Actual versus predicted returns [Case A: 30 testing samples under linear function]

Figure 4 depicts a comparison of actual returns and cases with 30 testing samples when the kernel function is the linear function. Except for the case linear-30/30, the other cases can generally depict the development trend of US REIT returns during the testing period, but the prediction results are almost always lower than the actual returns.

Table 4 presents the results that compare the gaps between the actual and predicted returns of these cases under the three kernel functions. Table 3 shows that the gaps between actual and predicted returns are much smaller for linear-100/30 and linear-200/30 than for the other cases. However, the gaps between actual and predicted returns for linear-30/30 are the largest of the five cases. Thus, the linear-100/30 and linear-200/30 prediction results outperform the other cases.

The mean squared error (MSE) is used to determine the predictability of the results. It computes the mean squared value of the difference between predicted and actual values, as shown below (Pai & Lin, 2005; Wang & Bovik 2009):

Table 3 Gaps between actual and predicted returns under the linear function [The presented value = predicted returns—actual returns]

Day	Linear-30/30	Linear-100/30	Linear-200/30	Linear-300/30	Linear-600/30
1	-2.27	-0.98	-1.26	-1.15	-1.79
2	-1.26	-0.87	-0.61	-1.52	-1.36
3	-1.35	-0.19	-0.01	-0.80	-0.60
4	-2.39	-0.11	-0.15	-1.01	-0.64
5	-2.72	0.04	-0.14	-0.88	-0.53
6	-3.18	-1.22	-0.94	-1.75	-1.48
7	-2.29	-0.21	-0.02	-0.47	-0.20
8	-3.23	-0.97	-0.78	-1.39	-0.92
9	-1.97	-0.79	-0.11	-0.38	-0.39
10	-2.54	-0.49	-0.36	-0.96	-0.67
11	-1.44	0.52	0.42	0.02	0.35
12	-0.38	-0.23	-0.40	-0.91	-0.45
13	-2.24	-2.04	-1.85	-2.48	-1.94
14	-1.64	-1.28	-1.23	-1.59	-0.98
15	-1.04	-0.04	-0.26	-0.36	0.32
16	-0.69	0.34	-0.11	-0.26	0.54
17	-1.98	-1.08	-1.48	-1.74	-0.83
18	-1.80	-1.23	-1.37	-1.46	-0.54
19	-2.15	-1.23	-1.40	-1.49	-0.46
20	-1.72	-0.29	-0.43	-0.47	0.70
21	-2.31	-0.69	-1.00	-1.19	0.05
22	-0.85	0.76	0.53	0.29	1.60
23	0.41	0.79	0.44	0.01	1.27
24	1.14	-0.05	-0.15	-0.90	0.38
25	3.29	1.12	1.26	1.22	1.72
26	0.33	-1.58	-1.65	-2.71	-1.56
27	0.43	-0.05	0.34	-0.52	0.33
28	-1.39	-0.18	-0.17	-0.96	-0.08
29	-1.72	0.40	0.42	-0.47	0.51
30	-2.94	-0.68	-0.80	-1.82	-0.83
31	-0.86	0.32	0.23	-0.83	0.20
32	-0.98	1.13	0.65	-0.46	0.51
MSE	3.69	0.72	0.70	1.45	0.87

$$MSE = \frac{\sum_{i=1,n} (y_i^{real} - y_i^{predicted})^2}{n} \tag{8}$$

A lower MSE indicates that the prediction is more accurate. However, preliminary research does not suggest that there are fixed criteria for determining what value of MSE is acceptable when the MSE is used to assess prediction accuracy. Table 4 summarizes the findings from studies on stock prediction and REIT prediction to determine the acceptable level of MSE. When different models were used to predict stock and REIT returns, different studies obtained different MSE values.

Compared with the MSE values of the five cases under the linear function shown in Table 3, case linear-200/30 has the lowest value, 0.70, while case linear-100/30 has a very

Table 4 MSE obtained by other studies in predicting returns of stocks and REITs

Author(s)	Objective	MSE
Pai and Lin (2005)	Stock prediction	0.0114–1.4676, and below 0.3001 is accepted
Panda et al. (2007)	Stock prediction	0.0001–0.1
Kogan et al. (2009)	Stock and REIT prediction	0.1365–5.6778, and below 0.3 is accepted
Schumaker and Chen (2009a)	Stock prediction	0.04261–0.17944
Schumaker and Chen (2009b)	Stock prediction	0.1954
Lee and Pai (2010)	REIT prediction	2.5727–2.7067
Ma et al. (2010)	Stock prediction	5.1176e-4 – 0.0478
Xiao et al. (2012)	REIT prediction	0.0011–0.0027
Braun (2016)	REIT prediction	8.6277–11.9349
Dietzel (2016)	REIT prediction	0.082
Feng et al. (2019)	Stock prediction	2.28e-4–5.20e-4

close MSE value, 0.72. These results show that the test accuracy in the two cases is higher than in the other three cases, which is consistent with the analysis of the gaps between actual and predicted returns. Furthermore, when the MSE values of the five cases are compared with the previous works, the linear function's accuracy is similar to the findings of Kogan et al. (2009). However, this study can accept prediction results when the MSE is less than 1.45 because the MSE value is still small. In other words, except for linear-30/30, the results of all cases can be accepted.

Linear covariance function

When the linear covariance function is used as the kernel function in the GMDH model, four cases with the same linear function condition setting are studied. Table 5 describes the details of these cases, which are labeled linear-cov-30/30, linear-cov-100/30, linear-cov-200/30, and linear-cov-300/30. Figure 6 compares the prediction results under the linear covariance function with the actual returns. This graph shows that, with the exception of case linear-cov-30/30, the prediction results are close to the actual results for all cases under the linear covariance function. The results in Fig. 5 and Table 5 show that the prediction results of the other three tests under the linear covariance function are acceptable.

Table 5 shows that with an MSE value of 0.69, the gap between the actual and predicted returns for case linear-cov-200/30 is much smaller than the gaps for the other cases under the linear covariance function. Cases linear-cov-100/30 and linear-cov-300/30 have acceptable gaps of around 1%, with MSE values of 1.05 and 1.08, respectively.

Quadratic function

This study also tests cases where the quadratic function is implemented as the kernel function. In total, six cases are tested: quadratic-30/30, quadratic-100/30, quadratic-200/30, quadratic-300/30, quadratic-600/30, and quadratic-300/60. Table 6 shows the results of these cases. All cases, with the exception of quadratic-300/60, test 30 samples. Figure 6 compares the actual returns in these cases with the predicted results.

Figure 6 shows that three cases—quadratic-30/30, quadratic-100/30, and quadratic-200/30—have invalid results due to outliers in their predicted returns and the

Table 5 Gaps between actual and predicted returns under the linear covariance function

Day	Linear-cov-30/30	Linear-cov-100/30	Linear-cov-200/30	Linear-cov-300/30
1	-1.01	-0.95	-0.57	-0.62
2	1.62	-0.29	0.19	-0.32
3	-0.29	0.28	0.48	0.25
4	-0.10	0.48	0.48	0.09
5	-0.54	0.82	0.53	0.07
6	0.70	0.14	0.42	-0.60
7	-1.11	1.53	0.85	0.24
8	-1.94	0.47	0.18	-0.64
9	-1.87	1.08	0.76	-0.22
10	-2.06	1.21	0.59	-0.56
11	-1.20	2.14	1.61	0.28
12	-2.80	0.91	0.89	-0.55
13	-4.33	-1.84	-0.40	-2.12
14	-5.81	-1.67	0.11	-0.76
15	-11.85	0.21	0.81	0.25
16	-9.40	0.56	0.87	0.46
17	-8.42	-0.88	-0.44	-1.86
18	-17.04	-0.70	-0.21	-2.14
19	-29.16	-0.71	-0.34	-1.06
20	8.07	0.38	0.72	-0.48
21	-11.14	-0.19	0.07	-2.87
22	-5.71	1.35	1.60	-0.93
23	-5.57	1.01	1.23	-0.02
24	-4.74	0.49	0.28	-0.61
25	-2.10	1.86	0.91	0.46
26	-2.63	-1.74	-2.00	-2.11
27	-2.04	0.89	-0.17	-0.76
28	-2.86	-0.15	-0.41	-0.97
29	-2.03	0.37	0.03	-0.25
30	-2.63	-1.35	-0.68	-1.33
31	-1.30	0.17	0.51	-0.46
32	0.31	-0.19	1.86	0.31
MSE	57.22	1.05	0.69	1.08

absence of a similar trend as the actual return. Meanwhile, the predicted returns of quadratic-300/30 and quadratic-600/30 are close to the actual returns. As a result, the three cases, quadratic-30/30, quadratic-100/30, and quadratic-200/30, are deemed invalid, and no further investigation is conducted for them.

Table 6 shows the differences between the actual and predicted returns for quadratic-300/30 and quadratic-600/30. Compared with the gaps between actual and predicted quadratic-300/30 results, the gaps between actual and predicted quadratic-600/30 results are much smaller. The respective MSE values for the two cases are 4.35 and 0.62. Thus, only the predicted quadratic-600/30 is acceptable when the quadratic function is used as the kernel function.

Table 6 Gaps between actual and predicted returns under the quadratic function

Day	Quadratic-300/30	Quadratic-600/30
1	-2.85	-1.36
2	-4.42	-0.86
3	-3.15	-0.16
4	-3.25	-0.39
5	-2.88	-0.34
6	-4.32	-1.23
7	-1.99	0.03
8	-2.57	-0.83
9	-0.74	-0.10
10	-1.01	-0.39
11	0.07	0.58
12	-0.92	-0.34
13	-2.18	-1.82
14	-1.14	-0.57
15	0.15	1.18
16	0.20	1.12
17	-1.22	-0.49
18	-0.97	0.64
19	-0.87	0.58
20	0.50	0.25
21	-0.14	-0.69
22	0.63	0.87
23	-0.72	0.50
24	-0.28	-0.56
25	1.01	0.41
26	-1.92	-1.88
27	-0.10	0.26
28	-0.86	-0.12
29	-0.61	0.41
30	-2.52	-0.95
31	-1.31	-0.02
32	-5.18	0.34
MSE	4.35	0.62

Note: The gaps between the actual returns and the predicted returns of quadratic-30/30, quadratic-100/30, and quadratic-200/30 are not calculated because the results of these three cases are not acceptable

Long-term predictions

The above cases have predicted 30 days' returns of the US REIT. In general, a 30-day period is relatively short, and when the training samples meet the proper conditions, the GMDH model effectively predicts the short-term returns of REIT under the three kernel functions. Under these circumstances, it is worthwhile to determine whether the GMDH model is also capable of predicting REIT long-term returns. Hence, the effectiveness of the GMDH model in predicting 60-day returns was tested in this study, with 300 training samples. The testing period runs from March 31, 2021 to June 30, 2021, with 300 training samples coming before that. In this study, three more cases are tested: linear-300/60, linear-cov-300/60, and quadratic-300/60. Figure 7 depicts a comparison of the three cases' actual returns and prediction results. Directly, the

Table 7 Gaps between actual and long-term predicted returns [Under linear and linear covariance and quadratic functions]

Day	Linear [300/60]	Linear covariance [300/60]	Quadratic [300/60]	Day	Linear [300/60]	Linear covariance [300/60]	Quadratic [300/60]
1	-0.23	0.42	0.65	32	-1.96	-1.36	-1.73
2	-2.06	-1.57	-1.41	33	-1.20	-0.54	-0.67
3	-0.84	0.04	0.34	34	-1.21	-0.59	-0.68
4	-1.45	-0.76	-0.76	35	-1.00	-0.55	-0.47
5	-1.06	-0.35	-0.28	36	-1.74	-1.31	-1.28
6	-0.64	0.04	0.35	37	-0.71	-0.23	0.07
7	-0.52	-0.03	-0.01	38	-1.46	-0.90	-0.75
8	-0.99	-0.44	-0.43	39	-0.82	-0.29	0.06
9	-1.77	-1.09	-0.91	40	-0.97	-0.59	-0.37
10	0.08	0.57	0.99	41	-0.07	0.30	0.57
11	-1.40	-1.32	-1.48	42	-0.84	-0.50	-0.44
12	-0.42	0.07	0.49	43	-2.69	-2.04	-1.77
13	-0.70	-0.52	-0.04	44	-2.02	-1.38	-0.77
14	-1.52	-1.14	-0.55	45	-0.72	-0.34	0.53
15	-1.23	-0.88	0.11	46	-0.46	-0.13	0.52
16	-0.06	0.24	1.83	47	-2.02	-1.53	-0.63
17	-0.72	-0.46	-1.40	48	-2.02	-1.45	91.50
18	-0.57	-0.08	-0.50	49	-2.10	-1.48	-1.15E+09
19	-0.92	-0.43	0.10	50	-1.20	-0.55	-1.66E+29
20	-0.57	-0.01	0.74	51	-1.78	-1.03	-4.86E+13
21	-1.69	-1.15	-2.64	52	-0.50	0.54	-8.40E+88
22	-1.48	-0.82	1.06	53	-0.44	0.43	272.20
23	-0.44	0.34	8.56	54	-1.43	-0.37	-3.89E+51
24	-0.24	0.20	-0.14	55	-0.02	0.97	-3.79E+71
25	1.11	1.35	-8.96	56	-3.08	-1.95	-3.82E+46
26	-1.36	-0.81	-9.03	57	-1.61	0.05	-536.51
27	-2.52	-1.52	-6.65	58	-1.51	-0.20	-6.71E+10
28	-0.70	0.26	-2.41	59	-1.07	0.36	-328.70
29	0.01	0.62	-2.01	60	-2.32	-0.97	61.35
30	1.70	2.20	-0.82	61	-1.53	-0.03	-42.26
31	-1.52	-1.38	-3.03	62	-0.76	0.44	-163.85
MSE	Linear-300/60	1.83	Linear-cov-300/60	0.82	Quadratic-300/60	1.138E+176	

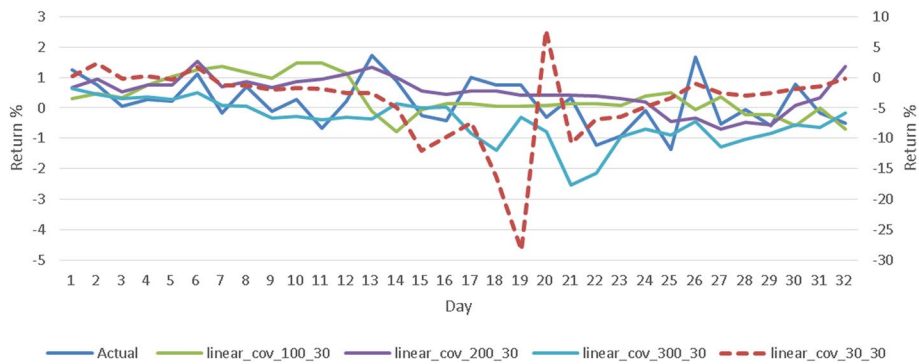


Fig. 5 Actual versus predicted returns [Case B: 30 testing samples under linear covariance function]

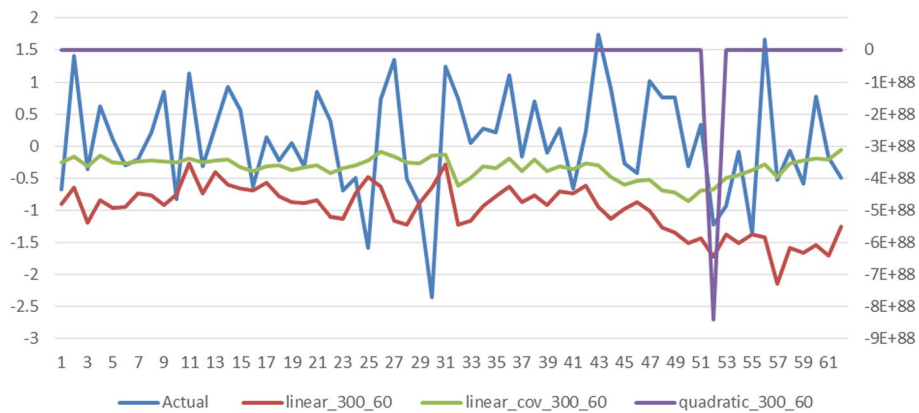


Fig. 6 Actual versus predicted returns [Case C: 30 testing samples under the quadratic function]

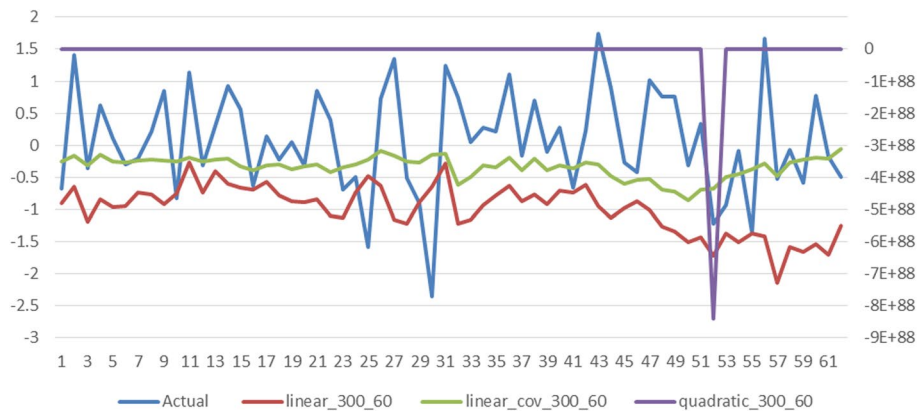


Fig. 7 Actual versus long-term predicted returns [Under linear, linear covariance and quadratic functions]

results of the case quadratic-300/60 are unacceptable because the magnitude of this model’s prediction results is not at the same level as the actual returns.

Meanwhile, the other two cases can show the trend of actual returns, but there are still significant gaps between the prediction results and the actual returns. The gaps between the two cases and the actual returns, as well as their MSE values, are calculated to determine whether the two cases should be accepted. Table 7 shows the differences between the actual returns and linear-300/60 and linear-cov-300/60.

The results in Table 7 show that the gaps between actual returns and linear-cov-300/60 are much smaller than those between actual returns and linear-300/60. The MSE values for linear-cov-300/60 and linear-300/60 are 0.82 and 1.83, respectively, indicating that linear-cov-300/60 is more accurate in predicting long-term returns than linear-300/60.

The following points can be summarized based on an analysis of the results of these tests with different kernel functions. First, the linear kernel function is robust to training model use in that the prediction is very stable and can achieve stable returns for both long-term and short-term goals. Second, the linear covariance kernel function is robust to the use of the training model. When compared with the linear kernel, the linear covariance kernel function fits the data better and predicts more accurate

results, but it is less stable, which requires further consideration (e.g., long-term goals). Third, the quadratic kernel function is difficult to train, necessitating a large number of samples (300+) to facilitate learning. A trained quadratic kernel is excellent for short-term goals but fails miserably for long-term goals.

The preceding analysis shows that kernel functions play an important role in determining the outcomes of the GMDH approach in predicting REIT returns. Scholars find that, aside from the GMDH model, kernel functions impact the outcomes of other ML approaches, such as the SVM model (Yuan et al. 2010). Yuan et al. (2010) have implemented the SVM model to test the internet traffic classification, and they investigated the effectiveness of four kernel functions, including RBF (radial basis function), poly, linear, and sigmoid, and discovered that RBF has the highest accuracy, followed by poly. This study, like Yuan et al. (2010), finds that different kernel functions are accurate in different cases. When the training sample is 100 and the testing sample is 30, the linear function is more accurate than the other two functions.

For the following pairs of training and testing samples, the linear covariance function is more accurate than the other two functions: 200/30, 300/30, and 300/60. The quadratic function is more accurate than the other two functions when the training sample is 600, and the testing sample is 30. Therefore, when implementing the GMDH model to forecast REIT returns, it is important to control the kernel functions in order to improve the accuracy of this ML method.

Estimates of the GARCH model

The GARCH model is also used to predict REIT returns. To compare the results of the GARCH model and the GMDH model, the numbers of observations used in the GARCH model and the numbers of predicted returns were controlled, and then six scenarios were tested, including Scenarios G-300/60, G-600/30, G-300/30, G-200/30, G-100/30, and G-30/30. These scenarios are explained as follows: Scenario G-600/30 means that the 300 samples are used in the GARCH model to estimate coefficients in Eqs. (5a), (5b), and (5c), and 60 returns are predicted. The meanings of the other scenarios are similar to Scenario G-600/30, but the number of samples used in the simulation and prediction differ. Table 8 depicts the specifics of these scenarios. Meanwhile, Table 9 displays the simulation results for these scenarios using the GARCH model.

Table 8 Descriptions of the scenarios by the GARCH model

Scenarios	Number of observations	Number of predicted returns
G-300/60	300	60
G-600/30	600	30
G-300/30	300	30
G-200/30	200	30
G-100/30	100	30
G-30/30	30	30

Table 9 Coefficients of these scenarios based on the GARCH model

Scenarios	$r_t = \mu + \alpha r_{t-1} + \varepsilon_t$		$\sigma_t^2 = \omega + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2$			MSE
	μ	α	ω	β	γ	
G-300/60	0.0710***	0.0691**	0.0493***	0.1311***	0.7613***	1.9317
G-600/30	0.0711***	0.0693***	0.0491***	0.1307***	0.7615**	1.6315
G-300/30	0.0703**	0.0652***	0.0474***	0.1265***	0.7521***	1.8723
G-200/30	0.0723***	0.0673***	0.0511**	0.1329***	0.7631***	2.2102
G-100/30	0.0812***	0.0686***	0.0525***	0.1341**	0.7671**	3.7935
G-30/30	0.0896**	0.0690**	0.0562***	0.1368**	0.7713***	4.9346

Note: *** and ** significant at the 1% and 5% levels, respectively

Table 10 MSE values of the two approaches

Training/testing	GARCH Model	GMDH Model		
		Linear kernel	Linear covariance kernel	Quadratic kernel
30/30	4.9346	3.69	57.22	–
100/30	3.7935	0.72	1.05	–
200/30	2.2102	0.70	0.69	–
300/30	1.8723	1.45	1.08	4.35
600/30	1.6315	0.87	–	0.62
300/60	1.9317	1.93	0.82	–

Note: The training and testing numbers of samples for the GARCH model indicate the number of samples used for calculating the coefficients and then the number of returns predicted by the model. "–" indicates that these cases are not accepted in this study because their predicted results are greatly impacted by outliers

GMDH versus GARCH predictions: a comparison

The MSEs of the two approaches were shown in Table 10. Given enough training samples, the GMDH neural network outperforms the GARCH model in terms of accuracy. Moreover, the GMDH model always outperforms the GARCH model in terms of accuracy when the kernel function is linear. When the number of training samples is greater than 100, the accuracy of the linear covariance function is greater than that of the GARCH model. When the kernel function is the quadratic function, its accuracy is higher than the GARCH model only for one case with 600 training samples, quadratic-600/30. Unlike the GARCH model, whose accuracy increases as the number of samples increases, the GMDH model’s accuracy is much more negatively affected by a lack of training samples.

When compared with a traditional GARCH-based approach, ML-based GMDH outperforms the GARCH model significantly. In other words, GMDH can predict more accurate results with less data and obtain results faster than GARCH models.

Discussion

The results shown above support the first hypothesis that the GMDH neural network outperforms the GARCH model in predicting the US REIT market. The GMDH neural network outperforms GARCH in predicting the US REIT for two reasons. First,

GMDH's self-organization can sort the effects of all possible model variables to find the best solution. Second, GMDH can analyze the structure of nonlinear and complex systems and avoid overfitting through using appropriate kernel functions.

Table 10 lists the comparison results of the two approaches for different number of training and testing samples. Our findings are consistent with Li et al. (2017) in that the GMDH neural network and ANN have better performance in predicting stocks and REITs than other traditional models, such as CAPM and double exponential smooth models. Furthermore, our results also lend support to the second hypothesis that GMDH can predict the US REIT market in both short- and long-term. Tables 3, 5, and 6 illustrate the short-term prediction results, while Table 7 shows the long-term prediction results of GMDH using various kernel functions. Given enough training data, our empirical results suggest that the GMDH approach can accurately predict REIT prices. Finally, we find that the GMDH produces better prediction results by avoiding model underfitting and overfitting issues through the use of different kernel functions. For example, using a linear kernel function can avoid the overfitting with simple data, while the quadratic function can better fit complex data.

Previous research applies two ML approaches to predict four REIT markets—Australia, Hong Kong, Italy, and Turkey—illustrating that the GMDH neural network stands out from the simple BP neural network (Li et al. 2017). However, the world's largest REIT market, the US, has been omitted. This study uses the GMDH neural network to forecast short- and long-term return forecasts for the US REIT market, as well as daily return forecasts. Furthermore, it demonstrates how the kernel function influences GMDH prediction results, which has previously been overlooked.

Conclusions

In recent years, ML has been widely used in the study of finance and economics, and it has also been practically applied to the field of finance. For example, JP Morgan devoted 280 pages of its investor report to reporting on how JP Morgan applies ML, deep learning, and neural networks to quantitative investment and algorithmic trading. Our study employs the ML-based GMDH model to forecast REIT returns in the United States and assesses the method's practical utility. The primary contribution of this study is the application of a GMDH neural network to forecast the trend of the US REIT market. The accuracy of the GMDH neural network is compared with that of a traditional approach, the GARCH model, in an empirical study. Several points can be identified by comparing the results of the two models. First, GMDH outperforms GARCH in predicting REIT returns. The ML method is used to test 17 cases while controlling the training samples, testing samples, and kernel functions. The test results reveal that the ML approach's accuracy is high in most cases, except when the training sample size is 30. Except for the quadratic kernel function, the linear and linear covariance kernel functions can be easily trained to obtain stable prediction results. Compared with the GARCH method's accuracy, the MSE values of most of these cases for the GMDH method are lower, indicating that the ML approach can easily obtain more accurate prediction results. However, the accuracy of the GMDH method is influenced by the kernel functions used when the training and testing samples are changed. Hence, when using the GMDH method to forecast REIT returns, the kernel function must be adjusted when changing the training

and testing samples. Based on the results of this study, the GMDH method produces more accurate predictions of REIT returns in the US market, even with small sample sizes. When investors decide to enter the REIT market, they can use the GMDH method to predict the short-term trend of the US REIT market based on the previous year's historical data.

Although this study found that the ML approach is more significant than traditional methods—by comparing the accuracy of the GMDH neural network with that of the GARCH model in an empirical study—this research still has some limitations. First, although the GMDH neural network is an excellent data analysis algorithm, it is a model that financial analysts will find more difficult to develop and implement. Second, the US REIT index was used in this empirical analysis. Although the US REIT market capitalization accounts for 61% of the global market, comparison results based on the US REIT index cannot accurately represent the overall situation of all REIT stocks. Therefore, further analysis is needed to improve the representativeness of the findings.

The following research directions can be considered for future studies. First, more ML and non-ML approaches can be tested and compared to determine the efficiency and effectiveness of ML approaches in predicting REIT returns. In the future, researchers can compare the efficiency of other ML and deep learning methods, such as SVM and Random Forest, in predicting REIT returns with the results of the GMDH method. Second, out-of-sample studies can collect data on REITs from multiple markets, as opposed to only the US market. Third, future research can explore which ML approaches can be used to effectively build REIT portfolios to increase returns.

Appendix 1: Definitions of the variables used in GMDH model

(1) Average True Range (ATR)

The expression to calculate the ATR of the REIT can be described by the following equation (Panapongpakorn & Banjerdpongchai, 2019):

$$\text{ATR} = \frac{\sum_{i=1}^N (P_{d-i} - V_{d-i})}{N} \quad (9)$$

In this expression, P_{d-i} and V_{d-i} are the highest and lowest prices in the period $d - i$, where d is forecasted day and i is the initial day, and N is the total number of the selected period. In this case, $P_{d-i} - V_{d-i}$ measures that price range in the period $d - i$.

(2) Bollinger Bands (BB)

The calculation formula of the BB can be expressed as follows (Bollinger 1992):

$$\text{Upperband(BBH)} : \bar{X} + 2\sigma \quad (10a)$$

$$\text{Middleband(BBM)} : \bar{X} \quad (10b)$$

$$\text{Lowerband(BBL)} : \bar{X} - 2\sigma \tag{10c}$$

In this case, the value of \bar{X} and σ are: $\bar{X} = \frac{\sum_{i=1}^N x_i}{N}$, $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N}}$, representing the mean value and derivation of the observed samples. When these formulas are used, the observed samples are the closing prices in a period and the value \bar{X} indicates the mean value in the selected duration, while the upper band has added two deviations, and the lower band has deducted two deviations.

(3) Keltner Channel (KC)

The calculation of the KC depends on the calculation of the EMA (Exponential Moving Average) and the ATR, and the details of formulas are listed as follows (Ruiz-Franco et al. 2018):

$$\text{The upper band (KCH)} : \text{EMA}(n_1) + k \times \text{ATR}(n_2) \tag{11a}$$

$$\text{Middle band (KCC)} : \text{EMA}(n_1) \tag{11b}$$

$$\text{The lower band (KCL)} : \text{EMA}(n_1) - k \times \text{ATR}(n_2) \tag{11c}$$

Here, EMA is calculated by the following formula:

$$\text{EMA}_n = \frac{2}{k + 1} \times P_{t-1} + \left(1 - \frac{2}{k + 1}\right) \times \text{EMA}_{n-1} \tag{12}$$

The EMA is implicitly calculated, depending on the value of the previous period (EMA_{n-1}). P_{t-1} indicates the closing price of the asset of the previous period, and n is the sample of the current period, and k indicates the number of periods.

(4) Donchian Channel (DC)

Formulas do not measure the upper band and lower band of the DC. In contrast, they are determined by the observations. The upper band indicates the highest prices in a period, while the lower band indicates the lowest prices in the same period. The middle band indicates the mean value of the gaps between the upper band and the lower band. Hence, it can be measured by Patil et al. (2020):

$$\text{The upper band (DCH)} : \text{The highest values in N periods} \tag{13a}$$

$$\text{Middleband(DCM)} : \text{DCM} = \frac{\text{DCH} - \text{DCL}}{2} \tag{13b}$$

$$\text{The lower band (DCL)} : \text{The lowest values in N periods} \tag{13c}$$

(5) Moving Average Convergence Divergence (MACD)

MACD is also an indicator based on the calculation of EMA, and its expression is shown (Du Plessis 2012):

$$\text{MACD} = \text{EMA}(12 - \text{dayclosingprices}) - \text{EMA}(26 - \text{dayclosingprices}) \tag{14}$$

This is a particular situation used in this research. The general formula of MACD can be expressed as follows (Hung, 2016):

$$MACD_t = EMA(s)_t - EMA(l)_t \tag{15}$$

In this formula, t indicates the moment, $EMA(s)$ and $EMA(l)$ are two states of EMA , determined by the lag lengths, and s and l represent short and long lag lengths, and the short lag length represents the fast-moving average while the long lag length represents the slow-moving average (Hung, 2016). In this empirical analysis, the short lag length is 12 days, and the long lag length is 26 days, which are two standard lag lengths used by researchers (Du Plessis 2012).

(6) Average Directional Movement Index (ADX)

The ADX is an indicator depending on the measurement of the ATR. Because it is used to identify the direction, it depends on the judgment of the directional indicator (DI) and directional movement (DM) (Gurrib 2018). The details of the calculation formula of the ADX can be expressed as follows (Gurrib 2018).

First, the DM is calculated, and it is divided into two types, $DM(+)$ and $DM(-)$ (in simple, $+DM$ and $-DM$).

$$DM_t = \begin{cases} DM_t(+) = high_t - high_{t-1} & \text{if } high_t - high_{t-1} > low_t - low_{t-1} \\ DM_t(-) = low_t - low_{t-1} & \text{if } low_t - low_{t-1} > high_t - high_{t-1} \end{cases} \tag{16}$$

When the value of DM is identified, the trader must measure the DI based on the 14-period-DM. The 14-period-DM is measured by the following formula (Gurrib 2018):

$$14 - \text{period} - DM = \sum_{i=1,14} DM_i \tag{17a}$$

$$\text{Subsequent} - 14 - \text{period} - DM = 14 - \text{period} - DM_{t-1} - \left(\frac{\sum_{i=1,14} DM_i}{14} \right)_{t-1} + DM_t \tag{17b}$$

Then the value of DI can be calculated by:

$$DI_t = \begin{cases} DI_t(+) = \frac{smoothed\ DM_t(+)}{ATR_t} = \frac{100}{14 - period - TR_t} 14 - period - DM_t(+) \\ DI_t(-) = \frac{smoothed\ DM_t(-)}{ATR_t} = \frac{100}{14 - period - TR_t} 14 - period - DM_t(-) \end{cases} \tag{18}$$

When the values of DI are obtained, the values of the directional index (DX) can be obtained (Gurrib 2018).

$$DX_t = \left| \frac{DI_t(+) - DI_t(-)}{DI_t(+) + DI_t(-)} \right| \tag{19}$$

Furthermore, ADX is the average value of DX in a 14-period (Gurrib 2018).

$$ADX_t = \frac{\sum_{i=1,14} DX_i}{14} \tag{20a}$$

$$\text{Subsequent ADX}_t = \frac{13\text{ADX}_{t-1} + \text{DX}_t}{14} \tag{20b}$$

According to the above equations from 17 to 20, the ADX estimates the trend in recent periods by integrating the temporary directional movements.

(7) Vortex Indicator (VI)

The calculation of the VI can be divided into four steps, and it is also divided into two trends, VI(+) and VI(-) CITATION Žbi15 \l 2052 (Žbikowski 2015).

The first step is to calculate the true range (TR), measured by Eq. (21). It measures the maximum value among the three items, including the gaps between the current high and current low prices, the gaps between the current high and previous close prices, and the gaps between the current low and previous close prices.

$$\text{TR}_t = \max\{\text{high}_t - \text{low}_t, \text{high}_t - \text{close}_{t-1}, \text{low}_t - \text{close}_{t-1}\} \tag{21}$$

The second step is to measure the two-directional movements, the downtrend (VM(-)) and the uptrend (VM(+)). The expression is shown in Eq. (22).

$$\text{VM}_t = \begin{cases} \text{VM}_t(+) = |\text{high}_t - \text{low}_{t-1}| \\ \text{VM}_t(-) = |\text{low}_t - \text{high}_{t-1}| \end{cases} \tag{22}$$

The third step is to calculate the special n periods' movement, and the formulas are shown in Eq. 23. Here, a period can be 14 days or 30 days, representing two weeks or a month. In this empirical, 14 days is selected.

$$\text{VM}_t^n = \begin{cases} \sum_{i=0,n} \text{VM}_{t-i}(+) = \sum_{i=0,n} |\text{high}_{t-i} - \text{low}_{t-1-i}| \\ \sum_{i=0,n} \text{VM}_{t-i}(-) = \sum_{i=0,n} |\text{low}_{t-i} - \text{high}_{t-1-i}| \end{cases} \tag{23}$$

And the final step is to calculate the two trends VI(+) and VI(-), shown in Eq. 24 CITATION Žbi15 \l 2052 (Žbikowski 2015). Like other indicators related to trends, the VI also depends on the true range in the past periods.

$$\text{VI}_t = \begin{cases} \text{VI}_t(+) = \frac{\sum_{i=0,n} \text{VM}_{t-i}(+)}{\sum_{i=0,n} \text{TR}_{t-i}} \\ \text{VI}_t(-) = \frac{\sum_{i=0,n} \text{VM}_{t-i}(-)}{\sum_{i=0,n} \text{TR}_{t-i}} \end{cases} \tag{24}$$

(8) Mass Index (MI)

The MI depends on the EMA calculation expressed in Eq. 11 and the simple moving average (SMA). The SMA calculation is shown in Eq. 25, which indicates the mean values of the closing prices in a period of 14 days. Here, PC indicates the closing price of a trading day.

$$\text{SMA}_t = \frac{\sum_{d=1,D} \text{PC}_d}{D} \tag{25}$$

The MI is calculated by Eq. 26, referring to the sum of the ratio between EMA and SMA. In this case, a period for SMA and EMA is 14 days, and a variety of periods are considered, including 2 periods, 6 periods, 20 periods, and 40 periods.

$$MT_t = \sum_{n=1, N} \frac{EMA_{k-day, n}}{SMA_{k-day, n}} \tag{26}$$

(9) Commodity Channel Index (CCI)

The CCI is calculated by Eq. (27), which depends on the moving average of price in the past days (Srinivasan et al. 2018).

$$CCI_t = \frac{Typical\ price_t - MA_t}{0.015 \times mean\ deviation_t} \tag{27}$$

Here, the typical price refers to the sum of the average value of the high, low, and close price in a series of periods, and MA and mean deviation indicates the moving average and deviation of these typical prices. The expressions of them are shown as follows (Srinivasan et al. 2018).

$$Typical\ price_t = \frac{high_t + low_t + close_t}{3} \tag{28a}$$

$$MA_t = \frac{\sum_{i=1, N} Typical\ price_{N-i}}{N} \tag{28b}$$

$$mean\ deviation_t = \frac{\sum_{i=1, N} (Typical\ price_t - MA_t)}{N} \tag{28c}$$

(10) Detrended Price Oscillator (DPO)

The DPO is calculated by Eq. 29 (Oyewola et al. 2019). It indicates the gaps between the current closing price and the moving average in the last $(\frac{n}{2} + 1)$ days. In this empirical analysis, a period is 14 days, hence $(\frac{n}{2} + 1)$ indicates 8 days.

$$DPO_t = close_t - SMA_{\frac{n}{2}+1} \tag{29}$$

(11) Parabolic Stop and Reverse (PSR)

The PSR can be divided into two trends, downtrend, and uptrend, and they can be calculated by the same expression (Chen & Metghalchi 2012).

$$PSR_t = PSR_{t-1} + AF \times (EP_{t-1} - PSR_{t-1}) \tag{30}$$

In this formula, PSR_t and PSR_{t-1} represent the current and the previous PSR of the asset, AF is an acceleration factor, changing in the range [0.02, 0.2]. The factor's best value is 0.2; hence, this empirical analysis uses 0.2 in the computing process. EP indicates an extreme price, either the highest or the lowest prices. For the downtrend condition, the lowest price is used. For the uptrend, the highest price is used (Chen & Metghalchi 2012).

(12) Money Flow Index (MFI)

The expression of the MFI is shown in equation (Di 2014).

$$MFI_t = 100 - \frac{100}{1 + MR_t} \tag{31}$$

In this formula, MR is the Money Ratio, indicating the percentage between the positive flow and negative money flow in a period marked as MF(+) and MF(-), respectively. $MR_t = \frac{MF_t(+)}{MF_t(-)}$. In general, money flow is calculated by the typical price multiplying the trading volume, $MF_t = Typical\ price \times volume$. The money flow is divided into positive and negative ones due to the changes in the typical price, and the relationship is shown in Eq. (32). The calculation of the typical price is shown in Eq. (28a).

$$MF_t = \begin{cases} MF_t(+), & \text{if } Typical\ price_t > Typical\ price_{t-1} \\ MF_t(-), & \text{if } Typical\ price_t < Typical\ price_{t-1} \end{cases}$$

(13) Relative Strength Index (RSI)

RSI expressions similar to MFI, shown in Eq. 32 (Crowell et al. 2016).

$$RSI_t = 100 - \frac{100}{1 + RS_t} \tag{32}$$

In this formula, RS indicates relative strength, which is calculated by: $RS_t = \frac{\sum_{k=i-n,i} PU(t_k)}{\sum_{k=i-n,i} PD(t_k)}$, where PU and PD are uptrend and downtrend prices. When the current closing price is greater than the closing price in the previous period, for example, the previous trading day, the price is marketed as a uptrend price. In contrast, it is the market as a downtrend price. Therefore, the RS means the percentage between the uptrend and downtrend prices in a particular period with n trading days. In this empirical analysis, n is 14, indicating 14 days (Crowell et al. 2016).

(14) True Strength Index (TSI)

The calculation of the TSI can be expressed (Hartle 2002).

$$TSI_t = \frac{EMA(EMA(mtm, r), s)}{EMA(EMA(|mtm|, r), s)} \times 100 \tag{33}$$

In this formula, mtm indicates the gaps between the current closing price and the previous trading day’s closing price, $mtm = closing\ price_t - closing\ price_{t-1}$. EMA indicates the exponential moving average operation shown in Eq. 11, and r and s are the length of a period and the number of periods, which are 25 and 13, respectively, indicating 25 days and 13 periods (Hartle 2002).

(15) Ultimate Oscillator (UO)

The expression of UO is shown in Eq. 34.

$$UO_t = 100 \times \frac{4 \times \frac{BP(7)}{TR(7)} + 2 \times \frac{BP(14)}{TR(14)} + \frac{BP(28)}{TR(28)}}{7} \tag{34}$$

In this formula, TR is the true range, which is calculated by Eq. 20. BP is the buying pressure, which is calculated by $BP_t = closing\ price_t - \min\{low_t, closing\ price_{t-1}\}$. Due to this, $\frac{BP(7)}{TR(7)} = \frac{\sum_{k=1,7} BP_k}{\sum_{k=1,7} TR_k}$, $\frac{BP(14)}{TR(14)} = \frac{\sum_{k=1,14} BP_k}{\sum_{k=1,14} TR_k}$, and $\frac{BP(28)}{TR(28)} = \frac{\sum_{k=1,28} BP_k}{\sum_{k=1,28} TR_k}$, which indicates the conditions of 7 days, 14 days, and 28 days before the current closing price.

(16) Stochastic Oscillator (SO)

The expression of SO is shown in Eq. 35.

$$SO_t = \frac{Closing\ price_t - L_{14}}{H_{14} - L_{14}} \times 100 \tag{35}$$

In this formula, *Closingprice_t* indicates the current closing price, *L₁₄* and *H₁₄* are the lowest and highest prices in the previous 14 trading days.

(17) Williams %R (WR)

The expression of WR is shown as follow (Li et al. 2017).

$$WR_t = \frac{H_{14} - closing\ price_t}{H_{14} - L_{14}} \times 100 \tag{36}$$

The meaning of these abbreviations, including *L₁₄* and *H₁₄*, share the same meanings of the SO.

(18) Awesome Oscillator (AO)

The AO means the gaps between the average value of 5 days before the current trading day and the average value of 34 days before the current trading day (Iovane et al. 2016). The calculation of SMA is shown in Eq. 24.

$$AO_t = SMA(5, t) - SMA(34, t) \tag{37}$$

(19) Kaufman’s Adaptive Moving Average (KAMA)

KAMA’s expressions are shown as follows (Karasu et al. 2020).

$$KAMA_t = KAMA_{t-1} + S_t \times (closing\ price_t - KAMA_{t-1}) \tag{38}$$

In this equation, *S_t* is the smoothing constant proposed by Kaufman (1995), the following formula calculates it.

$$S_t = sc^2 = ER_t \times sc_{fast} + (1 - ER_t) \times sc_{slow} \tag{39}$$

And, the formula of ER is $ER_t = \frac{D_t}{V_t} = \frac{closing\ price_t - closing\ price_{t-n}}{\sum_{i=1, n} |closing\ price_t - closing\ price_{t-i}|}$. The fast and slow smoothing constant (*sc_{fast}* and *sc_{slow}*) can be 0.6667 and 0.0645 when the periods are 2 and 30 days (Rakićević et al. 2014).

(20) Rate of Change (ROC)

The formula of the ROC is presented. In this formula, *n* indicates the number of days before the current trading day. As for daily ROC, *n* is 1.

$$ROC_t(n) = \frac{closing\ price_t - closing\ price_{t-n}}{closing\ price_{t-n}} \tag{40}$$

(21) Triple Exponential Average (TRIX)

The calculation formula of the TRIX can be expressed as follows (Di 2014). It is based on the EMA, which is shown in Eq. 11. Here, *n* indicates the number of days before the current trading day. As for the daily TRIX, *n* is 1.

$$TRIX_t = \frac{EMA(EMA(EMA(closing\ price_t)))}{EMA(EMA(EMA(closing\ price_{t-n})))} \tag{41}$$

(22) Know Sure Thing Oscillator (KST)

The calculation of KST depends on the ROC, and its expression is shown in Eq. 42 (Oyewola et al. 2019). In this formula, ROC indicates the change rate, which is measured by Eq. 29, and SMA indicates the moving average, measured by Eq. 24.

$$\begin{aligned} \text{KST}_t = & 1 \times \text{SMA}_t(\text{ROC}(10), 10) + 2 \times \text{SMA}_t(\text{ROC}(15), 10) \\ & + 3 \times \text{SMA}_t(\text{ROC}(20), 10) + 4 \times \text{SMA}_t(\text{ROC}(30), 15) \end{aligned} \quad (42)$$

(23) Ichimoku Kinko Hyo (Ichimoku)

Ichimoku's expression can be expressed as Eq. (43), which includes 5 components, Tenkan-sen, Kijun-sen, Senkou Span A, Senkou Span B, and Chikou Span (Cahyadi, 2012). Tenkan-sen focuses on the past 9 periods, Kijun-sen focuses on the past 26 periods, and Senkou Span B focuses on the past 52 periods. Chikou Span means to lag 26 days for the current prices, moving the price line to back to 26 days (Cahyadi, 2012).

$$\text{Tenkan} - \text{sen}_t = \frac{\max\{\text{high}_{t-1}, \dots, \text{high}_{t-9}\} + \min\{\text{low}_{t-1}, \dots, \text{low}_{t-9}\}}{2} \quad (43a)$$

$$\text{Kijun} - \text{sen}_t = \frac{\max\{\text{high}_{t-1}, \dots, \text{high}_{t-26}\} + \min\{\text{low}_{t-1}, \dots, \text{low}_{t-26}\}}{2} \quad (43b)$$

$$\text{Senkou SpanA}_t = \frac{\text{Tenkan} - \text{sen}_t + \text{Kijun} - \text{sen}_t}{2} \quad (43c)$$

$$\text{Senkou SpanB}_t = \frac{\max\{\text{high}_{t-1}, \dots, \text{high}_{t-52}\} + \min\{\text{low}_{t-1}, \dots, \text{low}_{t-52}\}}{2} \quad (43d)$$

$$\text{Chikou Span}_t = \text{closing price}_{t+26} \quad (43e)$$

(24) Daily Return (DR)

The DR is calculated by measuring the change ratio of the closing prices from the previous day to the current day (Li et al. 2017).

$$\text{DR}_t = \frac{\text{closing price}_t - \text{closing price}_{t-1}}{\text{closing price}_{t-1}} \quad (44)$$

(25) Daily Log Return (DLR)

The DLR is the logarithm value of the daily return (Li et al. 2017).

$$\text{DLR}_t = \ln(1 + \text{DR}_t) = \ln\left(\frac{\text{closing price}_t}{\text{closing price}_{t-1}}\right) = \ln(\text{closing price}_t) - \ln(\text{closing price}_{t-1}) \quad (45)$$

(26) Cumulative Return (CR)

The CR is similar to the DR, but it does not indicate the adjacent two days' gaps. In contrast, it focuses on the changes in the prices over a period. In the formula, n indicates the length of the period. In this empirical analysis, n is 14.

$$CR_t = \frac{\text{closing price}_t - \text{closing price}_{t-n}}{\text{closing price}_{t-n}} \quad (46)$$

Acknowledgements

Not Applicable

Author contributions

WZ designed and performed the experiments, derived the models, and analyzed the data. AL worked out the technical details. BL, ER, and TK developed the theoretical framework. All authors contributed to the interpretation of the results. WZ took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript. All authors read and approved the final manuscript.

Funding

Not Applicable.

Availability of data and materials

The data that support this finding of this study are available from database of Bloomberg and DataStream, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Financial Innovation.

Declarations

Competing interests

The authors whose name are listed immediately below certify that they have NO affiliate with involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or material discussed in this manuscript.

Received: 23 February 2022 Accepted: 23 March 2023

Published online: 01 July 2023

References

- Anderson RI, Benefield JD, Hurst ME (2015) Property-type diversification and REIT performance: an analysis of operating performance and abnormal returns. *J Econ Finance* 39(1):48–74
- Antunes JA (2021) To supervise or to self-supervise: a machine learning based comparison on credit supervision. *Financ Innov* 7(1):26
- Block RL (2012) Investing in REITs: real estate investment trusts. Wiley
- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econom* 31(3):307–327
- Bollinger J (1992) Using Bollinger bands. *Stocks Commod* 10(2):47–51
- Braun N (2016) Google search volume sentiment and its impact on REIT market movements. *J Prop Invest Finance* 34(3):249–262
- Brueggeman WB, Fisher JD (2022) Real estate finance and investments. McGraw-Hill
- Cahyadi Y (2012) Ichimoku Kinko Hyo: Keunikan Dan Penerapannya dalam Strategi Perdagangan valuta asing (Studi Kasus Pada Pergerakan USD/JPY Dan EUR/USD). *Binus Bus Rev* 3(1):480
- Chen CP, Metghalchi M (2012) Weak-form market efficiency: evidence from the Brazilian stock market. *Int J Econ Financ* 4(7):22–32
- Cheng P, Roulac SE (2007) REIT characteristics and predictability. *Int Real Estate Rev* 10(2):23–41
- Cho JH, Elshahat AF (2011) Predicting time-varying long-run variance-modified component GARCH model approach. *J Financ Econ Pract* 11(1):52–68
- Chui AC, Titman S, Wei KJ (2003) The cross section of expected REIT returns. *Real Estate Econ* 31(3):451–479
- Cici G, Corgel J, Gibson S (2011) Can fund managers select outperforming REITs? Examining fund holdings and trades. *Real Estate Econ* 39(3):455–486
- Crowell BW, Bock Y, Liu Z (2016) Single-station automated detection of transient deformation in GPS time series with the relative strength index: a case study of Cascadian slow slip. *J Geophys Res Solid Earth* 121(12):9077–9094
- Darrat AF, Gilley OW, Li B, Wu Y (2011) Revisiting the risk/return relations in the Asian Pacific Markets: New evidence from alternative models. *J Bus Res* 64(2):199–206
- Di X (2014) Stock trend prediction with technical indicators using SVM. Independent Work Report, Stanford Univ.
- Dietzel MA (2016) Sentiment-based predictions of housing market turning points with Google trends. *Int J Hous Mark Anal* 9(1):108–136
- Dixon MF, Halperin I, Bilokon P (2020) Machine learning in finance: from theory to practice. Springer, Berlin
- Du Plessis AW (2012) The effectiveness of a technical analysis strategy versus a buy-and-hold strategy on the FTSE/JSE top 40 index shares of the JSE Ltd: The case of the Moving Average Convergence Divergence Indicator. Doctoral dissertation, University of Johannesburg
- EPRA (2022) EPRA total market table. EPRA. <https://www.epra.com/research/market-research>
- Feng F, He X, Wang X, Luo C, Liu Y, Chua TS (2019) Temporal relational ranking for stock prediction. *ACM Trans Inf Syst (TOIS)* 37(2):1–30

- Feng K, Li Q (2014) Using stepwise regression and support vector regression to comprise REITs' portfolio. In: 2014 IEEE 7th Joint international information technology and artificial intelligence conference, IEEE, pp 158–162
- Fung K, Lau C, Chan K (2014) The conditional equity premium, cross-sectional returns and stochastic volatility. *Econ Model* 38:316–327
- Gurrib I (2018) Performance of the Average Directional Index as a market timing tool for the most actively traded USD based currency pairs. *Banks Bank Syst* 13(3):58–70
- Hansz JA, Zhang Y, Zhou T (2017) An Investigation into the substitutability of equity and mortgage REITs in real estate portfolios. *J Real Estate Financ Econ* 54:338–364
- Hartle T (2002) The true strength index. *Signal* 1(1):36–40
- Hausler J, Ruschinsky J, Lang M (2018) News-based sentiment analysis in real estate: a machine learning approach. *J Prop Res* 35(4):344–371
- Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Netw* 2(5):359–366
- Hui ECM, Chan KKK (2022) How does Covid-19 affect global equity markets? *Financ Innov* 8:25
- Hung NH (2016) Various moving average convergence divergence trading strategies: a comparison. *Invest Manage Financ Innovations* 13(2):363–369
- Iovane G, Amorosa A, Leone M, Nappi M, Tortora G (2016) Multi indicator approach via mathematical inference for price dynamics in information fusion context. *Inf Sci* 373:183–199
- Irem D, Piet E, Erkan Y (2020) Corporate diversification and the cost of debt. *J Real Estate Financ Econ* 61(3):316–368
- Karasu S, Altan A, Bekiros S, Ahmad W (2020) A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy* 212:1–12
- Kogan S, Levin D, Routledge BR, Sagi JS, Smith NA (2009) Predicting risk from financial reports with regression. In: Human language technologies: The 2009 annual conference of the North American chapter of the association for computational linguistics, Association for Computational Linguistics, pp 272–280
- Kou G, Olgu Akdeniz Ö, Dinçer H, Yüksel S (2021a) Fintech investments in European banks: a hybrid IT2 fuzzy multidimensional decision-making approach. *Financ Innov* 7(1):39
- Kou G, Xu Y, Peng Y, Shen F, Chen Y, Chang K, Kou S (2021b) Bankruptcy prediction for smes using transactional data and two-stage multiobjective feature selection. *Decis Support Syst* 140:113429
- Kou G, Yüksel S, Dinçer H (2022) Inventive problem-solving map of innovative carbon emission strategies for solar energy-based transportation investment projects. *Appl Energy* 311:118680
- Le A (2006) A linear multifactor model for REITs selection using gradient maximization. *Montreal: Université de Montréal, Montréal*.
- Lee YH, Pai TY (2010) REIT volatility prediction for skew-GED distribution of the GARCH model. *Expert Syst Appl* 37(7):4737–4741
- Lertwachara K (2007) Selecting stocks using a genetic algorithm: a case of real estate investment trusts (REITs). *Kaset-sart J Soc Sci* 28(1):106–116
- Li J, Lei L (2011) Determinants and information of REIT pricing. *Appl Econ Lett* 18(15):1501–1505
- Li RY, Fong S, Chong KWS (2017) Forecasting the REITs and stock indices: group method of data handling neural network approach. *Pac Rim Prop Res J* 23(2):123–160
- Li T, Kou G, Peng Y, Yu PS (2022) An integrated cluster detection, optimization, and interpretation approach for financial data. *IEEE Trans Cybern* 52(12):13848–13861
- Ling DC, Naranjo A (2006) Dedicated REIT mutual fund flows and REIT performance. *J Real Estate Finance Econ* 32(4):409–433
- Ling DC, Naranjo A (2015) Returns and information transmission dynamics in public and private real estate markets. *Real Estate Econ* 43(1):163–208
- Ling DC, Naranjo A, Ryngaert MD (2000) The predictability of equity REIT returns: time variation and economic significance. *J Real Estate Finance Econ* 20(2):117–136
- Liu L, Chen Q (2020) How to compare market efficiency? The Sharpe ratio based on the ARMA-GARCH forecast. *Financ Innov* 6(1):38
- Livingston F (2005) Implementation of Breiman's random forest machine learning algorithm. *ECE591Q Machine Learning Journal Paper*, pp 1–13
- Loo WK (2019) Predictability of HK-REITs returns using artificial neural network. *J Prop Invest Finance* 38(4):291–307
- Loo WK (2020) Performing technical analysis to predict Japan REITs' movement through ensemble learning. *J Prop Invest Finance* 38(6):551–562
- Lu C, So RW (2001) The relationship between REITs returns and inflation: a vector error correction approach. *Rev Quant Financ Acc* 16(2):103–115
- Ma W, Wang Y, Dong N (2010) Study on stock price prediction based on BP neural network. In: 2010 IEEE international conference on emergency management and management sciences, IEEE, pp 57–60
- Maneemaroj P, Lonkani R, Chingchayanurak C (2021) Appropriate expected return and the relationship with risk. *Glob Bus Rev* 22(4):865–878
- Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: an applied review. *Int J Remote Sens* 39(9):2784–2817
- Mori M, Ziobrowski AJ (2011) Performance of pairs trading strategy in the U.S. REIT market: performance of pairs trading strategy in the U.S. REIT market. *Real Estate Econ* 39(3):409–428
- Ng L (1991) Tests of the CAPM with time-varying covariances: a multivariate GARCH approach. *J Financ* 46(4):1507–1521
- Olanrele O, Said R, Bin Daud M (2014) Divided based return forecast as benchmark for REIT performance. *OIDA Int J Sustain Dev* 7(10):93–110
- Ott SH, Riddiough TJ, Yi HC (2005) Finance, investment and investment performance: evidence from the REIT sector. *Real Estate Econ* 33(1):203–235

- Oyewola DO, Dada EG, Olaoluwa OE, Al-Mustapha KA (2019) Predicting Nigerian stock returns using technical analysis and machine learning. *Eur J Electr Eng Comput Sci* 3(2):1–8
- Pai PF, Lin CS (2005) A hybrid Arima and support vector machines model in stock price forecasting. *Omega* 33(6):497–505
- Panapongpakorn T, Banjerdpongchai D (2019) Short-term load forecast for energy management systems using time series analysis and neural network method with average true range. In: 2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)
- Panda, G., Mohanty, D., Majhi, B., & Sahoo, G. (2007) Identification of nonlinear systems using particle swarm optimization technique. In: IEEE congress on evolutionary computation, IEEE, pp 3253–3257
- Patil P, Wu CS, Potika K, Orang M (2020) Stock market prediction using ensemble of graph theory, machine learning and deep learning models. In Proceedings of the the 3rd international conference on software engineering and information management, pp 85–92
- Rakićević A, Končarević R, Petrović B (2014) Comparison of moving averages for trading trends: The case of the belgrade stock exchange. In: *New business models and sustainable competitiveness*, pp 688–696.
- Ruiz-Franco L, Jiménez-Gómez M, Lambis-Alandete E (2018) Trading strategy on the Future Mini S & P 500. *Int J Appl Eng Res* 13(13):11018–11024
- Schumaker RP, Chen H (2009a) A quantitative stock prediction system based on financial news. *Inf Process Manag* 45(5):571–583
- Schumaker RP, Chen H (2009a) Textual analysis of stock market prediction using breaking financial news. *ACM Trans Inf Syst* 27(2):1–19
- Shen J (2021) Distress risk and stock returns on equity REITs. *J Real Estate Financ Econ* 62:455–480
- Shen J, Hui EC, Fan K (2021) The beta anomaly in the REIT market. *J Real Estate Financ Econ* 63:414–436
- Siew RYJ (2015) Predicting the behaviour of Australian ESG REITs using Markov chain analysis. *J Financ Manag Prop Constr* 20(3):252–267
- Sirmans S, Friday S, Price R (2006) Do management changes matter? An empirical investigation of REIT performance. *J Real Estate Res* 28(2):131–148
- Srinivasan D (2008) Energy demand prediction using GMDH networks. *Neurocomputing* 72(1–3):625–629
- Srinivasan K, Currim F, Ram S (2018) Predicting high-cost patients at point of admission using network science. *IEEE J Biomed Health Inform* 22(6):1970–1977
- Svozil D, Kvasnicka V, Pospichal J (1997) Introduction to multi-layer feedforward neural networks. *Chemom Intell Lab Syst* 39(1):43–62
- Swinkels L (2023) Empirical evidence on the ownership and liquidity of real estate tokens. *Financ Innov* 9:45
- Wang Z, Bovik AC (2009) Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Process Mag* 26(1):98–117
- Wang X, Xiao X, Xiao Z (2016) S-REITS' performance forecast using a small sample model associating support vector machine with vector auto-regression model. *Int J Innov Comput Inf Control* 12(1):15–40
- Xiao ZY, Li SM, Lin ZX (2012) A hybrid modeling approach for forecasting the volatility of REITs index in US market. In: 2012 International conference on management science & engineering 19th annual conference proceedings, IEEE, pp 1861–1867
- Yang L, Yang H, Liu H (2018) GMDH-based semi-supervised feature selection for Electricity Load Classification forecasting. *Sustainability* 10(1):217
- Yuan R, Li Z, Guan X, Xu L (2010) An SVM-based machine learning method for accurate internet traffic classification. *Inf Syst Front* 12(2):149–156
- Zbikowski K (2015) Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy. *Expert Syst Appl* 42(4):1797–1805
- Zhou J, Kang Z (2011) A comparison of alternative forecast models of REIT volatility. *J Real Estate Finance Econ* 42(3):275–294

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
