**Enhancing REITs Portfolios Through Advanced Machine Learning Analysis (Using Fundamental Data)**

Professor Liu Lili

National University of Singapore, lily.liu@nus.edu.sg

Leong Wei Ming

National University of Singapore, leongwm@u.nus.edu

Abstract:

This paper's main contributions are the successful combination of Genetic Algorithms to search for outperforming Alphas, thereby creating a portfolio optimisation model that is more profitable and less risky than the market index. The paper also extends past research by incorporating fundamental analysis into ML models and implements a trade execution logic for portfolio allocation for ML models.

Keywords: Portfolio optimization, stock price prediction, fundamental analysis, trade execution

Enhancing REITs Portfolios Through Advanced Machine Learning Analysis (Using Fundamental Data)

*Completed Research Paper*

# **Introduction**

Due to the evolving nature of the financial markets, traditional time-series forecasting models which are static in nature are less effective than Machine Learning models in identifying the best investments (Sheth & Manan, 2023). The success of Machine Learning has led to numerous research papers applying a myriad of Machine Learning techniques to predict stock prices (Obthong et.al., 2020). However, fewer papers apply these approaches in the Real Estate sector and these papers largely focus on future stock price prediction rather than portfolio allocation as a whole (Habab & Kampouridis, 2024). Many of these papers also use price-volume data without fundamental financial data as inputs. In 2022, the global real estate sector was worth more than $380 trillion and worth more than the global equity and bond markets combined (Tostevin & Rushton, 2023), with approximately 893 listed Real Estate Investment Trusts (REITs) (Nareit, 2024). Hence, addressing the research gaps in this sector is particularly valuable.

## ***Problem Definition***

This paper aims to develop an effective real estate portfolio optimisation model that maximises profits, minimises risks and outperforms the market indexes with a sharpe ratio of more than 2. Three Machine Learning (ML) models, namely, Multiple Linear Regression (MLR), Neural Networks (NN) and Long-Short Term Memory (LSTM) are used to construct models. A trading agent is also constructed to allocate funds to the portfolio based on the results from the models. The input data to the models will be extended beyond price-volume data to include fundamental data. The models will be trained on 10 years of real market data. Then, their performances will be compared with traditional financial models and the market indexes.

## ***Motivations***

Interviews of traders have revealed that both technical and fundamental analysis are used for forecasting investments, for shorter and medium to long term investment horizons respectively (Oberlechner, 2001). Given that most investments in real estate are medium to long term, the lack of fundamental data inputs for REITs stock price predictions reveals a pressing need for fundamental data to be incorporated.

## ***Contribution and Creativity***

This project analyses REITs stock datasets with more than 150 features of technical and fundamental indicators. It expands the dimensionality of the input training data significantly beyond price-volume features to incorporate REIT company data from financial statements. The comprehensive dataset covers 150 out of the total worldwide population of 893 REITs and includes up to 20 years of historical data. It also contributes an implementation of a trading agent or trade execution logic that acts on individual REIT price predictions to allocate funds to selected REITS in a portfolio. Overall, this research paper has extended extant research on the application of Machine Learning models in Optimising Real Estate Portfolios.

# **Background**

In order to better communicate ideas, the key financial terminologies are synthesised from Stephen Ross' Corporate Finance textbook (2021) and contextualised to this paper. Particularly the difference between a stock, REIT and a portfolio, and technical versus fundamental analysis. This builds the foundation for further discussions on the measures of portfolio performance, namely, profits/losses and the Sharpe ratio. Other methods of analysis such as sentiment and macro-economic analysis are also briefly discussed.

## ***Key Terms***

A stock also called a share, represents ownership of a company's assets. Each stock is represented in the market by a unique identifier called a ticker e.g. Apple's ticker is "AAPL". Over time, share prices can rise and fall. The daily movement of the share prices over time will be represented in this paper by the formula:

A math equation with numbers and symbols

Description automatically generated with medium confidence

* Prices: P is the price of a specific stock at a particular time
* Date: T in this paper is the daily date
* Ticker: i is the stock ticker for a specific stock

A REIT company invests in properties. Typically, these companies invest or own a mix of housing, commercial or industrial properties. Buying shares of REITs allow investors to own a share of the real estate that these REITS are managing and indirectly invest their monies in the real estate sector. Each of the 893 REITs across the world are different companies which own a wide range of different properties. In this paper, a REIT is used to refer to the stock of a REIT company. A REIT is a type of stock.

A portfolio is a collection of stocks i.e a portfolio is a combination of more than one stock. For example, the portfolio in Figure 2.1 consists of 10% of stock A, 11% of stock B and 9% of stock C. The relative proportions of each stock in a portfolio can be represented by weights. This paper will represent a portfolio of stocks with the vector and matrix below:



A group of black letters

Description automatically generated

* Weight: W is the percentage value of a ticker in the portfolio.

This paper aims to develop an effective Real Estate Portfolio Optimisation Model that maximises profits and minimises risks, outperforming the market indexes with a sharpe ratio of more than 2.

Conventionally, we need to own a certain stock before we can sell it. However, market makers also allow investors to borrow stocks to sell them first, with the contractual obligation for the investor to repurchase the stocks from the market at a later date to return to the market maker. This mechanism of selling first and buying back later is called shorting and is represented by a negative weight.

## ***Evaluating Investments with Data***

This subsection covers two schools of thought in investment analysis, technical and fundamental analysis.

### ***Technical Analysis with Price Volume Data***

Technical analysis is one method of evaluating investments in stocks and identifying prospective stocks to invest in. It focusses primarily on identifying short term trends in price and volume data. Table 2.1 provides the key features required for technical analysis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Open** | **High** | **Low** | **Close** | **Volume** |
| 31/08/2023 | 125.36 | 125.91 | 123.89 | 124.20 | 3277600 |
| 01/09/2023 | 125.40 | 125.57 | 124.05 | 124.59 | 1591000 |
| **Table 2.1: Daily Price Volume Data for Ticker PLD** | | | | | |

The features Open, High, Low and Close are derived from the Price of a stock at different points in the day and their trends can indicate future movements, while volume signals the strength of a price trend.

### ***Risk and Volatility***

Volatility is a statistical measure of how varied share price and returns on an investment are over a period of time. It is a proxy of the risk of an investment, the likelihood of an investment underperforming.

|  |
| --- |
| A diagram of a graph  Description automatically generated |
| **Figure 2.3: Prices and Profits from Stock A, Purchasing on Different Days** |

Figure 2.3a shows the distribution of prices of Stocks A and B over a period of time. Stock A has a lower standard deviation σA than Stock B σB, lower volatility and lower price risk. Because this paper seeks to minimise risk, the variance of prices and returns need to be considered.

### ***Sharpe Ratio***

Deciding between stocks or portfolios with different expected returns and risk like in Figure 2.3b where Stock B has higher returns but also higher risk than Stock A, requires a formula that evaluates these two measures together. The Sharpe ratio combines both measures of profit/loss and risk into a single formula:

A math equation with black text

Description automatically generated

* Rp: The annualised return of portfolio
* Rf : The risk free rate (return from a zero risk investment)
* σp: The standard deviation of portfolio return

In this paper, the Sharpe ratio is used as the overall measure of performance of the portfolio optimization.

### ***Fundamental Analysis with Financial Statements***

Fundamental analysis is a second method of investment analysis that is widely used by finance practitioners but less often applied by Machine Learning researchers. It involves examining the assets and incomes of the companies behind the different stock tickers by looking at their financial statements. Every quarter, every company that has a Ticker issues comprehensive financial statements publicly like in Figure 2.4.

|  |
| --- |
| A table with numbers and text  Description automatically generated |
| **Figure 2.4: Highly Abbreviated 3 Main Financial Statements** |

The predictive value of each statement is as follows:

* Income Statement: Profitability of the company’s business and the rate of growth of its value
* Balance Sheet: The asset value and financial stability of the company
* Cash Flow Statement: How cash is generated by the company

Incorporating both technical and fundamental data into machine learning models can provide a holistic picture and better inform these models with an extended set of short, medium and long-term features.

# **Literature Review**

## ***Comparison of Traditional and Machine Learning Techniques***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 3: Comparison of Techniques in Extant Research** | | | | | |
| Methods | Data Inputs | Scope for REITs | Strengths | Weaknesses | References |
| Autoregressive  Integrated  Moving  Average  (ARIMA) | Daily Price Volume Data | Stock Price Prediction | +Suitable for short-term time-series analysis  + Widely used in  finance for prediction  + More explanable than  complex ML models | -Unsatisfactory for long term prediction  -Higher RMSE than ML models | (Ariyo et.al., 2014), (Habbab & Kampouridis, 2022), (Obthong et. al., 2020) |
| Generalised  AutoRegressive Conditional Hetero-skedasticity (GARCH) | Daily Commodities  Price-Volume  data | Stock Volatility Prediction | +Outperforms ARIMA for price prediction and  volatility forcasting | - Model assumes volatility  can be predicted based on  past returns, but volatility  in reality is very unpredictable | (Fiszeder & Chung,  2020), (Lama et. al.,  2015), (Yuan et. al,  2017 |
| Multiple Linear Regression (MLR) | Daily Price- Volume Data | Stock Price Predic- tion | +Relatively low RMSE error for stock price predic- tion  + Relatively easy to implement | - Only considers input features and not historical  price data | (Obthong et. al.,  2020), (Shakhla et.  al., 2023) |
| Recurrent  Neural Networks (RNN) | Daily Price Volume Data |  | +Relatively low RMSE error for stock price prediction  + Remembers historical  stock prices | - Susceptible to vanishing and exploding gradient  problem | (Dey et. al., 2021) |
| Long-Short  Term Memory  (LSTM) | Daily Price Volume Data | Stock Price Predic- tion | + One of the lowest RMSE error  + Remembers historical  stock prices  + Solves the vanishing and  exploding gradient problem | - Mixed results when predicting REIT stock prices | (Axelsson & Song, 2023), (Habbab & Kampouridis, 2022), (Obthong et. al., 2020) |
| Random Forest (RF) | Daily Price- Volume Data,  Fundamental  Data |  | + One of the few papers  to incorporate fundamental  data into machines learning models | - Relatively low sharpe ratio compared to market indexes | (Cao, 2021), (Huang et. al., 2021) |

## ***Predictive Accuracy***

|  |
| --- |
| A graph of different colored squares  Description automatically generated with medium confidence |
| **Figure 3.1: RMSE for various ML Models for REITs (Habbab & Kampouridis, 2024)** |

A recent paper aimed at predicting REIT share prices through 5 machine learning algorithms and 3 traditional benchmarks, which includes many of the methods in Section 3.2 above, found that machine learning models (MLR, LSTM) appear more accurate than traditional benchmarks (ARIMA). The algorithms were trained solely on daily price data, and Figure 3.1 shows the RMSE results for out-of sample predictions over different time periods (30, 60, 90, 120 and 150 days).

## ***Gaps in Past Research to be Addressed***

From the literature review, there are the following gaps that can be further explored in this paper:

1. From the ”Data Inputs” column, it appears most papers use price-volume (technical) data without fundamental data to make predictions; and
2. From the ”Objectives” column, most papers are focussed on predicting stock prices and rarely address portfolio allocation/optimisation (i.e. to use predictions to inform buy/sell or find the optimal weights of each stock in a portfolio to maximise the Sharpe ratio).

# **Datasets**

## ***Extended Fundamental Data Features***

While many papers use price-volume data from Yahoo Finance’s API (yfinance), this data only contains 10 or so price-volume indicators. This paper goes further by obtaining a comprehensive list of 161 fundamental features from FinancialModelingPrep’s (FMP) API.

## ***Feature Exploration with Decision Trees***

A 25 layer Decision Tree of the features reveals the strongest predictors for the next day’s closing price.

|  |
| --- |
| A diagram of a tree  Description automatically generated |
| **Figure 4.2: Top 4 Layers of Decision Tree to Predict T+1 Close for PLD Ticker** |

It appears from Figure 4.2 that many of the balance sheet, income and cash flow features are indeed strong predictors of share price, coupled with technical features like Volume and Percentage Change in Share Price.

***Feature Selection***

Many of the top features can be encapsulated by other features. For example, Other Current Liabilities is a component of Total Current Liabilities and Expenses are part of Net Income. Given that the tree was generated from only one stock ticker, to make the feature selection generalisable to all 150 tickers, the results are contextualised with financial knowledge to derive the final 22 Selected Features.

# **Methodology**

A screenshot of a diagram

Description automatically generated

Figure 5.1: The Overall Workflow For This Paper

## ***Machine Learning for REITs Portfolio Optimisation***

In Figure 5.1, the first approach of ML consists of two parts, A) Price Prediction and B) Trade Execution.

### ***MLR / NN / LSTM Predictions with Extended Features***

At the pre-processing stage of all three models, data for each ticker is split into the training and test sets with a 80:20 ratio. The close price for each row is also shifted one day forward so that all 21 features can work on predicting the next day’s close. By and large, the implementation for all the models are relatively similiar except the lines of code where the model is actually used. Each ticker will have an output prediction dataframe like Figure 5.2:

### ***Trade Execution Logic***

The second part of the machine learning approach requires implementing the logic to decide how to trade on these predictions. All the predicted prices for all the dataframes are appended into a single large data frame, the parameters in Listing 5.4 are set and the green columns in Figure 5.2 are computed.

A screenshot of a computer

Description automatically generated

|  |
| --- |
| A white grid with black text  Description automatically generated |
| **Figure 5.2: Processing Predicted Prices to Determine Trade Amount** |

* trade: If Abs % Change Predicted exceeds the action threshold parameter, trade an amount equal to % Change Predicted / trade factor denominator in millions. If shorting is disabled, does not take action on negative % Change Predicted.

Next, a transaction class is defined to store the transaction details until the position is closed

A screenshot of a computer screen

Description automatically generated

A white paper with blue text

Description automatically generated

For every non-zero Trade amount in Figure 5.2, the transaction object is created in yellow.

The closing transaction details are then added to each transaction object by iterating through the ticker’s price movement over time and calling the Action method of the transaction class, which will close the position when the conditions in Listings 5.5 are met.

### ***Performance Evaluation***

In the evaluation phase the various dataframes will be consolidated to calculate the expected return, annualized rate of return, standard deviation followed by the Sharpe ratio.

# **Machine Learning Results**

## ***Evaluating Stock Price Predictions***

In terms of the first part of the ML approach of stock price prediction, it appears all 3 models have predicted prices that closely mirror the actual prices. However, comparing their Mean Square Errors (MSE) reveals that LSTM provides the most accurate price prediction, followed by NN and then LR, with a MSE of 212, 252 and 323 respectively in Figure 7.2.

|  |  |
| --- | --- |
| A graph with different colored bars  Description automatically generated | A graph of a graph showing a number of stocks  Description automatically generated with medium confidence |
| **Figure 7.2: Mean Squared Error of Prediction Models** | **Figure 7.5: Total Portfolio Value (Stocks + Cash) Over Time** |

It appears that even while including fundamental data features, LSTM’s memory of historical price information does aid its ability to identify trends and make more accurate predictions (Obthong et. al.,2020). And NN’s multiple layers help to further reduce the error in predictions.

In Figure 7.5, the total portfolio value was maintained consistently above the starting investment of 10 Million, although it spiked in early 2020 and fell sharply thereafter. Notably, the COVID-19 pandemic started to impact the financial markets in mid-2020 which coincided with the total fall in portfolio value.

## ***Overall Evaluation of Performances***

A table with numbers and a green box

Description automatically generated with medium confidence

From Figure 7.6, it appears that trading without shorting is superior in terms of profitability and sharpe ratio. This is possibly because short selling creates the potential for unlimited losses (Bryant, 2023).

Surprisingly, the NN approach has the highest Sharpe Ratio of 3.5 followed by LR with 2.53 and LSTM with 1.52. This could be because of how the trade execution logic is devised, which requires the change predicted to be above the action threshold percentage. Nevertheless, NN and MLR have resulted in a profitable trading strategy that minimises risk and maximises profits with a sharpe ratio greater than 2.

## ***Benchmarking Against Indexes, ARIMA, GARCH***

A graph with green and blue bars

Description automatically generated

From Figure 9.1, it appears that ML models significantly outperform traditional approaches to portfolio optimisation with more than double the sharpe ratio, compared to the market index of 0.14 or the average REIT stock of 0.6. In terms of risk and volatility, the ML models also have significantly lower volatility than the market index and ARIMA as represented by the blue line.

## ***Key Findings***

The key findings for this paper are as follows:

1. The extension of input data for ML models to include fundamental data results in acceptable predictions and portfolio returns that have a sharpe ratio of more than 2 for NN and LR approaches.
2. As whole the ML approaches adopted by this paper outperforms the traditional and market approaches.
3. It is surprising that despite LSTM’s highest accuracy in predicting the next day’s closing price, the portfolio allocation based on LSTM’s predictions perform the poorest. This emphasises that for the ML models, the trade execution logic that implements the actual portfolio allocation based on predictions has a large effect on the overall portfolio performance beyond price prediction.

## ***Future Work***

The trade execution logic has a large impact on the overall profitability of the portfolio allocation for ML price prediction approaches. Hence, more experiments can be done with regards to the various parameters input into the trade execution code, and possibly refine the logic to improve performance. Macroeconomic indicators can also be included in the features to account for events like COVID-19.

**References**

[1] Anton Aguilar-Rivera, Manuel Valenzuela-Rend´on, and J. Rodr´ıguez-Ortiz. Genetic algorithms and Darwinian approaches in financial applications: A survey. Expert Systems with Applications, 42:7684–7697, November 2015.

[2] Alexandria. Unlocking Japanese Market Sentiment with Nikkei FTRI, 2023.

[3] Adebiyi A. Ariyo, Adewumi O. Adewumi, and Charles K. Ayo. Stock Price Prediction Using the ARIMA Model. In 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pages 106–112, March 2014.

[4] Birger Axelsson and Han-Suck Song. Univariate Forecasting for REITs with Deep Learning: A Comparative Analysis with an ARIMA Model, 2023.

[5] Sean Bryant. Why You Should Never Short a Stock, 2023.

[6] Kai Cao. Fundamental Analysis via Machine Learning. 2021.

[7] Polash Dey, Emam Hossain, Md Ishtiaque Hossain, Mohammed Armanuzzaman Chowdhury, Md Shariful Alam, Mohammad Shahadat Hossain, and Karl Andersson. Comparative Analysis of Recurrent Neural Networks in Stock Price Prediction for Different Frequency Domains. Algorithms, 14(8):251, August 2021. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

[8] Piotr Fiszeder and Tin Fah Chung. What are the major limitations of Econometric Models like GARCH and TARCH?, 2020.

[9] Fatim Z. Habbab and Michael Kampouridis. Machine Learning for Real Estate Time Series Prediction. Sheffield, UK, July 2022.

[10] Fatim Z. Habbab and Michael Kampouridis. An in-depth investigation of five machine learning algorithms for optimizing mixed-asset portfolios including REITs. Expert Systems with Applications, 235:121102, 2024.

[11] Yuxuan Huang, Luiz Fernando Capretz, and Danny Ho. Machine Learning for Stock Prediction Based on Fundamental Analysis. In 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pages 01–10, Orlando, FL, USA, December 2021. IEEE

[12] Zura Kakushadze. 101 Formulaic Alphas, March 2016. arXiv:1601.00991 [q-fin].

[13] Achal Lama, Girish K. Jha, Ranjit K. Paul, and Bishal Gurung. Modelling and Forecasting of Price Volatility: An Application of GARCH and EGARCH Models. Agricultural Economics Research Review, 28(1):73, 2015.

[14] Nareit. Global Real Estate Investment, 2024.

[15] Thomas Oberlechner. Importance of technical and fundamental analysis in the European foreign exchange market. International Journal of Finance & Economics, 6(1):81–93, January 2001.

[16] Mehtabhorn Obthong, Nongnuch Tantisantiwong, Watthanasak Jeamwatthanachai, and Gary Wills. A survey on machine learning for stock price prediction: algorithms and techniques. pages 63–71, May 2020. Num Pages: 9.

[17] Stephen Ross. Fundamentals of Corporate Finance. March 2021.

[18] Michael Ruse. Charles Darwin’s Theory of Evolution: An Analysis. Journal of the History of Biology, 8(2):219–241, 1975. Publisher: Springer.

[19] Shruti Shakhla, Bhavya Shah, Niket Shah, Vyom Unadkat, and Pratik Kanani. Stock Price Trend Prediction Using Multiple Linear Regression. May 2020.

[20] Dhruhi Sheth and Manan Shah. Predicting stock market using machine learning: best and accurate way to know future stock prices. International Journal of System Assurance Engineering and Management, 14(1):1–18, February 2023.

[21] Paul Tostevin and Charlotte Rushton. Total Value of Global Real Estate: Property remains the world’s biggest store of wealth, September 2023. Section: Market trends.

[22] Igor Tulchinsky, editor. Finding Alphas: A Quantitative Approach to Building Trading Strategies. Wiley, 1 edition, September 2019.

[23] Saizhuo Wang, Hang Yuan, Leon Zhou, Lionel M. Ni, Heung-Yeung Shum, and Jian Guo. Alpha-GPT: Human-AI Interactive Alpha Mining for Quantitative Investment, July 2023. arXiv:2308.00016 [cs, q-fin].

[24] Ya-Ping Yuan, Jiong Sun, and Hong-Kun Zhang. Garch Models in Value-At-Risk Estimation for REIT. 2017.