Optimizing a prediction-based, mixed-asset portfolio including REITs

Fatim Z. Habbab, Michael Kampouridis
School of Computer Science and Electronic Engineering
University of Essex
Wivenhoe Park, United Kingdom
{fh20175, mkampo}@essex.ac.uk

***Abstract*—The real estate asset class has captured the attention of billions of global investors due to its ability to generate consistent returns and offer diversification benefits within a mixed-asset portfolio. Prior research has highlighted the advantages of including real estate in portfolio optimization. However, existing studies have primarily focused on historical data when addressing this optimization problem. This paper presents an analysis of the performance of a portfolio that incorporates real estate using price predictions derived from a Long Short-Term Memory (LSTM) model. To provide a comprehensive evaluation, we compare the performance of our portfolio against a benchmark portfolio consisting of stocks and bonds only. To this end, we run a genetic algorithm on the two portfolios. Our findings demonstrate a substantial improvement in the average risk-adjusted return of the portfolio that includes real estate with a magnitude of around 100%, highlighting the substantial value that real estate brings to a diversified portfolio. In this way, we propose a novel approach for showing the benefits of investing in real estate.**

***Index Terms*—mixed-asset portfolio, real estate, LSTM, risk-adjusted return**

I. Introduction

Real estate has long been recognized as a valuable component of a diversified investment portfolio [1]. Including real estate in a mixed-asset portfolio offers several potential benefits. First, real estate has historically exhibited low correlation with traditional asset classes such as stocks and bonds [2]. This low correlation can help improve portfolio diversification and reduce overall portfolio risk [3]. Second, real estate investments have the potential to provide consistent income streams through rental income, which can act as a hedge against inflation and provide stability during economic downturns [4]. Third, real estate has the potential for capital appreciation over the long term, as property values tend to increase over time [5].

In order to assess the added value of investing in real estate, it is useful to compare the performance of a portfolio that includes real estate to that of a portfolio that does not include it. Previous studies have examined such comparison and found evidence supporting the inclusion of real estate in mixed-asset portfolio. For instance, [6] found that portfolios that included real estate showed higher risk-adjusted returns compared to portfolios without real estate. In a more recent study, [7] conducted an analysis of U.S. investment portfolios and concluded that portfolios that incorporated real estate assets outperformed those that did not, both in terms of risk-adjusted returns and diversification benefits. Furthermore, [8] confirmed the positive impact of real estate inclusion on portfolio performance, emphasizing its ability to enhance risk-adjusted returns and improve diversification benefits.

However, those studies relied on historical data in calculating the optimal weights of a mixed-asset portfolio, and then applied those weights to an unseen test set [9]. A potential limitation of such approach is that prices in the test set might differ significantly compared to the prices in the training set [10]. As a result, weights computed using the training set might not fit the test set very well and thus lead to worse portfolio performance (i.e., increased risk and/or reduced return).

To alleviate the above issue, an alternative approach is to try and predict prices in the test set, and then perform the portfolio optimization task — i.e. calculating the optimal weights — directly in the test set [11]. The advantage of this approach is that we focus only on the data period we're interested in — i.e., the test set —; as a result, accurate predictions would closely reflect the prices in the test set, and thus lead to a more efficient portfolio selection. However, the quality of the results is very much dependent on the effectiveness of the price predictions.

Once we have obtained price predictions, we run a genetic algorithm (GA) to optimize a portfolio that includes real estate, stocks and bonds. In order to provide a comprehensive evaluation, we compare the performance of our portfolio to that of a portfolio consisting of stocks and bonds only. Our goal is to demonstrate that a portfolio including real estate outperforms a portfolio not including it. To this end, we evaluate financial metrics such as Sharpe ratio, returns, and risk and compare the results with the proposed benchmark — i.e., portfolio not including real estate.

For our experiments, we used daily prices downloaded from *Yahoo!Finance* for stocks and REITs, and *Investing.com* for bonds, referring to the period between January 2017 and January 2021, for financial instruments belonging to three asset classes — i.e., stocks, bonds, and real estate —, and to three countries — i.e., US, UK, and Australia. For each of the three markets, we used prices for five stocks, five bonds, and five REITs. Thus, we ran our experiments on a total of 90 datasets. All prices were expressed as USD, so as to account for currency risk.

The rest of this paper is organized as follows. Section II explains the methodology used in this study. The results of our experiments are presented in Section III, where we provide a detailed discussion of the results obtained by predicting asset prices using LSTM, and by running a GA to optimize our portfolios. Finally, Section IV summarizes the conclusions of the study.

II. Methodology

Our experiments aim to provide evidence that a mixed-asset portfolio including real estate can significantly outperform a mixed-asset portfolio not including real estate. This aim can be broken down into two subtasks: (i) use LSTM to predict the prices of REITs, bonds, and stocks, and (ii) use these predictions as an input to a genetic algorithm, which is going to optimize the weights of all assets in the portfolio.

Before applying the LSTM algorithm, we first needed to take several data pre-processing steps, which are presented in Section II-A. We then present the features that we included in the price prediction in Section II-B and the loss function, which is the same across all algorithms, in Section II-C. Afterwards, we briefly present the Python libraries we used to apply our machine learning algorithms in Section II-D. Lastly, we present the genetic algorithm setup in Section II-E.

A. Data preprocessing

Before being used for price prediction, each time series data is differenced and scaled. In this work, we adopt a first-order differencing process to transform our data. First-order differencing is a common technique used in time series analysis to remove the trend component from the data. It aims to transform a non-stationary time series into a stationary one. A stationary time series is one whose statistical properties, such as the mean and variance, remain constant over time. Stationarity is desirable because it simplifies the analysis and makes it easier to model the underlying patterns. By taking the difference between consecutive observations, $D_t = P_t - P_{t-1}$, we obtain a new time series that represents the changes between adjacent data points.

After obtaining the values of D_t, they are further transformed to fall within the range of 0 and 1 using the scaling equation outlined in Equations 1.

$$N_t = \frac{(D - D_{min})}{(D_{max} - D_{min})} \tag{1}$$

where N_t is the standardized value of each variable (in this case the differenced price D), and D_{min} and D_{max} are the minimum and maximum value for D respectively, over all data in each dataset.

B. Features

For our regression problem, we employ two types of features: past observations of a given time series, denoted as N_t, and technical analysis (TA) indicators. The past observations (N_{t-1}, N_{t-2}, N_{t-3}, ..., N_{t-T}) are incorporated as features, with the lag length determined based on the Akaike Information Criteria (AIC) optimization. AIC is a widely-used metric for model selection. Each dataset may have a different lag length, resulting in a varying number of features. Additionally, we incorporate five TA indicators: Simple Moving Average (SMA), Exponential Moving Average (EMA), Moving Average Convergence/Divergence (MACD), Bollinger Bands, and Momentum. These indicators help identify trends and assist in price prediction and have been shown to improve REITs price prediction accuracy [12]. The SMA represents the average of past prices, while the EMA assigns exponentially decaying weights to past observations. The MACD measures the difference between short-term and long-term EMAs. Bollinger Bands define an interval around the SMA, considering standard deviations from the mean. Momentum captures the difference between prices over a specific time period. These indicators provide valuable information for predicting future price movements.

C. Loss function

Our LSTM model evaluated by using out-of-sample predictions, rather than one-day-ahead predictions. The former is when today's N_t value ($t1$) is known and is used to forecast the value of tomorrow ($t2$). However, tomorrow's value is unknown and cannot be used to forecast the value two days ahead. Hence, this method uses the value forecast at time-step 1 to forecast the value at time-step 2, and so on. In the case of one-day-ahead forecasting, the price today (time-step 0) is known, and is used to forecast tomorrow's price (time-step 1). Then tomorrow's *real* price is used to forecast the price at time-step 2, and so on. This second method is expected to be more accurate, because we are using the actual values as features, instead of predictions. However, for portfolio optimization purposes using out-of-sample predictions would be more realistic as using one-day-ahead predictions would require rebalancing a portfolio on a daily basis for a time period of around 150 days which can lead to significant management costs.

For our problem, we use the *root mean square error* (RMSE) as the loss function, which is presented in Equation 2:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{|j|}(P_t - \hat{P}_t)^2}{|j|}}, \tag{2}$$

where P_t refers to the actual value of the price, $\hat{P}_t$ is its predicted value, and $|j|$ denotes the number of observations for each dataset j. Please note that as it was explained in Section II-A, the differenced and scaled values (i.e. D_t and N_t respectively) are reverted back to their original price values (i.e. P_t), so that the loss function can be calculated.

D. LSTM

To apply our LSTM algorithm we used the *keras*[1] library. In order to fit the algorithm to the training data we used the

[1] https://keras.io/getting_started/ Last access: January 2023

`keras.Sequential` method. The trainable hyperparameters were determined using a grid search method. Once the algorithms were fit to the training data, they were then applied to the test set by using the `predict` attribute of the relevant model.

E. Genetic algorithm

Evolutionary algorithms have been widely used for financial applications — e.g. [13, 14, 15] —, including portfolio optimization [16]. To tackle the portfolio optimization problem we consider in this paper, we use a particular type of evolutionary algorithm known as *genetic algorithm* (GA) [17]. Below we briefly discuss the GA we have used.

GA *chromosomes* (or, *individuals*) consist of N *genes* indicating the weights allocated to the N assets in the portfolio. The weight are real numbers in the interval $[0, 1]$, and their sum is equal to 1. For example, a GA individual that has the genotype [0.5 0.2 0.3] indicates that there are three assets, and the weight for those asset are 0.5, 0.2, and 0.3, respectively. Initially, all genes are assigned the same weight (in particular, $W_i = 1/N$ for each asset i), which are then evolved according to a set of operators.

We use *elitism*, *one-point crossover* and *one-point mutation*. After the application of crossover and mutation, we apply normalization to each GA individual, to ensure that the sum of weights remains equal to 1.

State-of-the-art methods for solving portfolio optimization problems have used many different metrics as fitness functions. In this paper, we use the *Sharpe ratio*, defined as the ratio of the difference between the average return and the risk-free rate, over the standard deviation of the returns, that is,

$$S = \frac{r - r_f}{\sigma_r}, \tag{3}$$

where r is the average return of the investment, r_f is the risk-free rate, and σ_r is the standard deviation of the returns.

III. Results

In this Section, we examine the experimental results in the form of RMSE distributional statistics (Section III-A), and summary statistics regarding the GA portfolio optimization results (Section III-B). It should be noted that all results are daily results. So when, for example, we present a seemingly "low" return of around 0.03%, its annual equivalent would be around 11.6%. [2]

A. RMSE

First, we compare the accuracy of predictions between two scenarios, one that includes REITs, and one that does not include REITs. Table I shows the summary statistics for two RMSE distributions, one for each of the two previously mentioned scenarios. For each of those distributions, we analyze the mean and standard deviation. As we can observe, the RMSE distribution in the first scenario shows lower RMSE average value compared to the second scenario, with a percentage difference of -46.43%. This indicates that including REITs in the analysis improves the accuracy of predictions. Furthermore, the RMSE distribution for the first scenario shows a noticeably lower standard deviation value compared to the second scenario, with a reduction of 50.79%. This suggests that incorporating REITs in the analysis leads to more accurate predictions with reduced variability.

In order to compare the RMSE distributions obtained, we performed a Kolmogorov-Smirnov (KS) test at the 5% significance level. The null hypothesis is that the compared RMSE distributions belong to the same continuous distribution. According to the test results, the adjusted p-value is equal to 1.94E-45, which indicates a statistically significant difference in the two distributions.

In summary, when analyzing the RMSE values, it becomes evident that incorporating REITs in the analysis improves the accuracy of predictions in terms of mean and standard deviation. The scenario of incorporating REITs consistently outperforms the scenario of not including REITs, suggesting that including REITs provides more precise predictions. From the KS test results, we observed that such difference is statistically significant.

B. GA portfolio optimization

After having analyzed the RMSE distributional statistics, we examine the expected portfolio performance for the above-mentioned scenarios. First, we examine the expected return distributions. From Table I, we can notice an increase in the expected return average of around 66.06%. We also notice a 54.11% reduction in the volatility of the expected return distribution, which indicates an increased concentration of values around the mean. This implies that including REITs in a mixed-asset portfolio might improve the overall portfolio return with a reduced volatility.

We also observe that the average expected risk tends to decrease when including REITs with a magnitude of around 33.21%. This implies that investing in REITs allows to reduce the overall portfolio risk. On the other hand, we notice that the standard deviation of the expected risk values tends to decrease with a magnitude of around 63.28%, which indicates an increased concentration of risk values around the mean.

Finally, we observe that the average Sharpe ratio increases when incorporating REITs, with a percentage difference of 103.71%. We also notice a slight increase in the volatility of 4.17%. This suggests that including REITs tends to have a marginal impact on the volatility the risk adjusted returns.

In order to compare the Sharpe ratio distributions obtained, we again performed a Kolmogorov-Smirnov (KS) test at the 5% significance level. Since we are making three comparisons, one for each metric (i.e., portfolio return, risk, and Sharpe ratio), we adjusted the p-values according to the Bonferroni's correction (e.g., $0.05/3 = 0.0167$). According to the test results, the adjusted p-value is equal to 1.55E-45 for all the considered metrics, which indicates a statistically significant difference in the compared distributions.

[2] $\text{AnnualizedReturn} = [(\text{DailyReturn} + 1)^{365} - 1] \times 100 = 11.6\%$.

TABLE I: RMSE and Sharpe ratio distributional statistics. Values in bold represent best results for each statistic.

	RMSE			Expected return		
Metric	**Without REITs**	**With REITs**	**% Difference**	**Without REITs**	**With REITs**	**% Difference**
Mean	36.29	**19.44**	-46.43%	5.41E-04	**8.99E-04**	66.06%
Std Dev	146.15	**71.93**	-50.79%	6.08E-05	**2.79E-05**	-54.11%
	Expected risk			**Sharpe ratio**		
Metric	**Without REITs**	**With REITs**	**% Difference**	**Without REITs**	**With REITs**	**% Difference**
Mean	5.54E-03	**3.70E-03**	-33.21%	7.26E-03	**1.48E-02**	103.71%
Std Dev	4.04E-04	**1.48E-04**	-63.28%	**5.33E-04**	5.55E-04	4.17%

In summary, when considering the portfolio return, risk, and Sharpe ratio distributions, we observe that including REITs in the analysis has a positive impact on the portfolio performance. It significantly improves the risk-adjusted distributions, as a result of an increased portfolio return and a reduced portfolio risk. The effect of REITs on risk-adjusted return distributions is significant, as shown by the KS test results.

IV. Conclusions

In our work, we evaluated the performance of a portfolio including REITs by comparing it against a portfolio that does not include REITs. From our experimental results, we noticed a significant improvement in the risk-adjusted performance of our portfolio which is highlighted by a greater average Sharpe ratio that doubles the average Sharpe ratio of a portfolio that does not include REITs. This can be related to a lower average RMSE that results from including REITs in the analysis. This suggests that including REITs in a portfolio including bonds and stocks can mitigate the greater portfolio risk caused by including stock investments.

While our results show that adding real estates to investment portfolios can have positive effect under the diversification perspective, further research can be done on different countries to further explore the opportunities of investing in real estate. Another opportunity for further research might be to extend the holding period for real estate portfolios.

References

[1] J. Y. Campbell and L. M. Viceira, “Strategic asset allocation: Portfolio choice for long-term investors,” *Oxford University Press*, 2006.

[2] D. M. Geltner, N. G. Miller, J. Clayton, and P. Eichholtz, *Commercial real estate analysis & investments*. Cengage Learning, 2017.

[3] J. Gordon and C. F. Sirmans, *Real estate finance and investments*. McGraw-Hill/Irwin, 2013.

[4] W. B. Brueggeman and J. D. Fisher, *Real estate finance and investments*. McGraw-Hill, 2011.

[5] A. Baum, *Commercial real estate analysis and investments*. Routledge, 2016.

[6] D. C. Ling and A. Naranjo, “Real estate and the capital asset pricing model: The integration of real estate into a mixed-asset portfolio,” *Real estate economics*, vol. 30, no. 3, pp. 361–382, 2002.

[7] Z. Chen and D. Ghosh, “Diversification benefits of including commodities, reits, and tips in mixed-asset portfolios: A simulation analysis,” *Journal of Real Estate Portfolio Management*, vol. 23, no. 2, pp. 191–201, 2017.

[8] F. Z. Habbab, M. Kampouridis, and A. A. Voudouris, “Optimizing mixed-asset portfolios involving REITs,” in *2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)*, pp. 1–8, IEEE, 2022.

[9] J. Lekander and M. Wilhelmsson, “Real estate and the financial crisis: How did the portfolio of swedish real estate firms perform during the crisis?,” in *ERES Conference*, 2015.

[10] F. Z. Habbab and M. Kampouridis, “Optimizing mixed-asset portfolios with real estate: why price predictions?,” in *2022 IEEE World Congress on Computational Intelligence (WCCI)*, pp. 1–8, IEEE, 2022.

[11] F. Z. Habbab and M. Kampouridis, “Machine Learning for Real Estate Time Series Prediction,” in *2022 UK Workshop on Computational Intelligence (UKCI)*, IEEE, 2022.

[12] F. Z. Habbab, M. Kampouridis, and T. Papastylianou, “Improving REITs time series prediction using ML and technical analysis indicators,” in *2023 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2023.

[13] S. Cramer, M. Kampouridis, and A. A. Freitas, “Decomposition genetic programming: An extensive evaluation on rainfall prediction in the context of weather derivatives,” *Applied Soft Computing*, vol. 70, pp. 208–224, 2018.

[14] M. Kampouridis and F. E. Otero, “Evolving trading strategies using directional changes,” *Expert Systems with Applications*, vol. 73, pp. 145–160, 2017.

[15] S. Cramer, M. Kampouridis, A. A. Freitas, and A. Alexandridis, “Stochastic model genetic programming: Deriving pricing equations for rainfall weather derivatives,” *Swarm and Evolutionary Computation*, vol. 46, pp. 184–200, 2019.

[16] M. Kampouridis, A. Brabazon, and M. O’Neill, “Applications of genetic programming to finance and economics: past, present, future,” *Genetic Programming and Evolvable Machines*, vol. 21 (1), pp. 33–53, 2020.

[17] D. E. Goldberg, “Genetic algorithms in search, optimization, and machine learning. addison,” *Reading*, 1989.