Comparative Analysis of Neural Networks for Asset Price Prediction

William Zhang
260865382
Department of Software Engineering
McGill University
Montreal, QC
william.zhang2@mail.mcgill.ca

November 3, 2019

Abstract

Financial asset pricing is one of the key driving forces in our economy, with various models such as Capital Asset Pricing Model (CAPM), attempting to understand, explain and predict future asset prices. However, due to its multi-factored nature, lack of linear representation and inherent noise in financial pricing, it is impossible to determine all underlying factors that influence an asset price. In order to solve these issues, machine learning techniques are used to establish relationships across multiple factors and determine their effect on asset prices. The current state-of-the art models for time series analysis are Multilayered Perceptrons (MLP), Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) . With the last two models able to map time dependencies in order to more accurately represent various economic movements. While all three models are able to accurately map non-linear relationships, they do so using differing techniques. The three techniques are compared on their accuracy, their loss using mean squared error, and their performance in average daily returns. Based on their results, it is found that RNNs are the overall best model for asset price prediction.

***K*eywords** Asset Price Prediction · Machine Learning · Neural Network · Long Short Term Memory · Recurrent Neural Network

1 Introduction

The state of the economy is constantly in movement. Predicting the current and future states of the economy can drastically reduce market inefficiencies. However, there are many views on how and why the market moves like it does. One common perspective on the economy is that it follows a cyclical nature, with periods of expansion and recession, but with an overall upwards trajectory. These cycles are of varying lengths, some being short term, lasting a couple of months, and others spanning multiple years. This perspective on the economy indicates that there are underlying trends which can be picked up and used to forecast the future state of the economy.

1.1 Traditional Models

Capital Asset Pricing Model (CAPM), first introduced by Sharpe in 1964[1], is widely regarded as the standard model used to allocate assets effectively. This technique uses variance, covariance, correlation and other factors between an asset and the financial market to asses and predict the optimal allocation based on its risk. The underlying assumption is that stocks that have big and frequent movements in their price carry a higher risk on investment and therefore should have a higher expected return.

Building on top of the work done by Sharpe, Fama and French developed a 5 factor model used to more accurately reflect stocks that were consistent outliers in their performance relative to the market[2]. Both these models showcase the issue in asset price prediction, as they fail to include crucial factors such as momentum. There are simply too many factors to include in a manually constructed model, therefore both the CAPM[1] and the Fama and French model[2] attempt to simplify the problem space, consequently only offering well constructed estimation model at best.

1.2 Related work

The literature on Machine Learning approaches in solving Financial Asset pricing is varied and expansive, touching multiple techniques and various markets. Although various deep learning techniques have been around for over half a century, it is only recently that computers have managed to perform calculations fast enough for machine learning to be viable.

However, one of the earliest neural networks applied in finance was done by Lawson[3]. Following suite, Welch and Goyal demonstrated in 2007 that many fundamental and technical factors could influence an asset price, such as dividend-price ratios[4]. Heaton et al. showed the effectiveness of machine learning models for other financial goals, such as portfolio optimization or risk management[5]. Although such objectives are not touched upon in this paper, the techniques he uses in his deep learning models are similar to those currently being used by financial institutions, including Long Short Term Memory models and auto-encoders. Ghysels et al. showcase how these machine learning models can not only effectively predict asset prices in the short term, they are able to detect moments of high volatility in the economy. These algorithms can therefore be used as a tool for investors to more effectively allocate their portfolio during moments of high volatility[6].

Finally, Ryll and Seidens aggregated and compared over 150 scholarly articles of machine learning techniques in financial market prediction, while also comparing with traditional stochastic techniques. Their findings demonstrate that machine learning models are indeed most effective, but the robustness of these models are yet to be evaluated over a longer period of time[7].

2 Background

2.1 Problem Statement

While various papers have explored the effectiveness of machine learning algorithms in asset price prediction, an in-depth comparison of the techniques themselves has rarely been conducted. Three popular deep learning models currently used to find non-linear relationships are the Multilayer Perceptron (MLP), the Recurrent Neural Network (RNN) and a Long Short Term Memory (LSTM) model.

These models map relationships between input and output parameters based on the training data. Herein lies the issue with comparing these models, where a models performance can vary drastically depending on the data used to train it. Therefore, despite the findings reported by Yu and Yan on the performance of their model[8], it is unreasonable to compare this to the performance that Kraus and Feuerriegel[9] found using their RNN model.

Not only is the data-set used to train the model very important, each model has a set of hyper-parameters that are incredibly important in their performance. Without these two factors being consistent with all three models, any findings from comparing them is simply irrelevant. Therefore, to compare these models, an identical dataset will be fed to all three, with any adjustments made to suit each algorithms dimension space, and all hyperparameters will remain constant. All three models will execute gradient descent on mean squared error (MSE), one of the three comparison criteria.

2.2 Comparison Criteria

Each model will be compared based on three metrics, their accuracy, their mean squared error and their performance. Each criteria will showcase not only the error margin of the model, but also the consistency of each prediction.

2.2.1 Accuracy

Accuracy is used as a means to demonstrate how often a model is able to predict if an asset is increasing or decreasing in price from the prediction interval. A reasonable baseline to assume for this metric would be 33%. This is because an asset price can stay stagnant and not move in short prediction intervals.

2.2.2 Mean Squared Error

While accuracy is to determine the direction of movement of a particular asset, the mean squared error is used to determine the average error in prediction relative to its market price. The MSE is calculated using the following formula, where n is all the number of assets being predicted, y is the actual price, and $\tilde{y}$ is the predicted price.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_i - \tilde{y_i})^2$$

2.2.3 Performance

The performance of each model can be categorized as the average return per trade. While metric is correlated with the accuracy and MSE of the model, it is able to demonstrate the performance of each model in financial markets. Important to note is that each trade will not take into consideration trading fees and is assumed to occur at closing time, once on each trading day. The performance of the model can be calculated as the daily return for each trade day.

3 Analysis

Each model will be trained on three stocks, AAPL, GOOGL and MSFT. The dataset contains daily prices since 2000 and contains the daily Open, Close, High, Low and Volume that each asset. The training data is roughly 80% of the data, containing 3968 data points, and the test set is 960 points.The training set is a time period of roughly 16 years, with the test set covering the remaining 4 years.

3.1 Multilayer Perceptron

The multilayer perceptron model has an advantage over the other models in terms of its simplicity. Because the model does not take into consideration temporal relationships its input dimension is far smaller which leads to a faster execution time. On average, the MLP model predicted assets with an accuracy of 37.50%, a MSE of 2.13×10^{-4} and an average daily return of 0.27%.

From these results, a MLP model would not be effective in predicting asset pricing. Not only is the accuracy barely above the baseline of 33%, the average daily return of 0.27% would be entirely ineffective when taking into account training costs. This can be confirmed when looking at the Figure 1 and 2, demonstrating how the model has a hard time converging to an optimal solution.

The test accuracy has such high variance that it fluctuates across the training accuracy. This demonstrates that the model is unable to converge successfully and that it is unable to detect the relationship between the current asset price and the future price. The MSE is steadily decreasing at the beginning, however quickly converges and stops decreasing. However, this local minimum is clearly ineffective in predicting asset prices as the accuracy is still incredibly volatile when the loss converges.

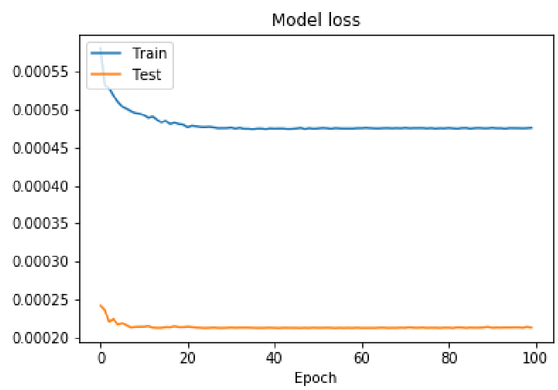

Figure 1: Validation error per epoch for MLP

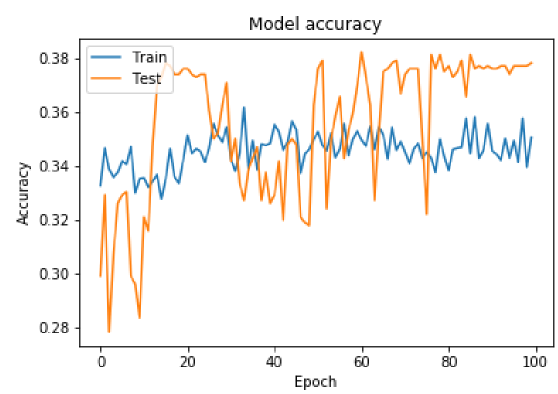

Figure 2: Accuracy per epoch for MLP

3.2 Recurrent Neural Network

The Recurrent Neural Network performs remarquably better than the MLP, with an accuracy of 81.25%, an MSE of 1.04×10^{-5} and a daily return of 2.88%. Not only are these results far better than the MLP, the returns generated from this model are very high. This could be explained with the temporal dependency that RNNs follow.

The results of the RNN relative the MLP demonstrates the importance of temporal mapping and in the process, empirically confirms that the economy has underlying trends. This is to say that past performance is a good indicator for future performance, given a model that can accurately map their relationship.

Figures 3 and 4 demonstrate that the error and accuracy were both steadily decreasing and increasing respectively. Furthermore, while the loss of the model decreased in the first few epochs, each additional epoch was still very useful as demonstrated in the accuracy of the model consistently increasing even when the loss has converged.

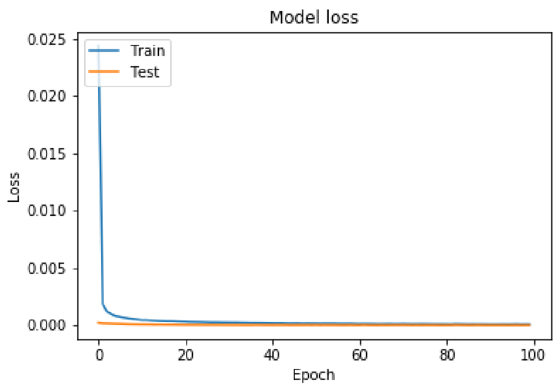

Figure 3: Validation error per epoch for RNN

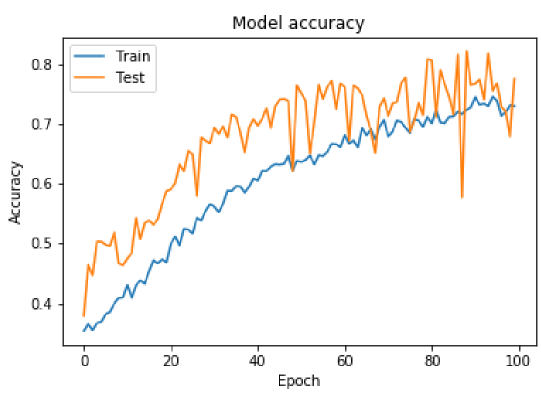

Figure 4: Accuracy per epoch for RNN

3.3 Long Short Term Memory

The Long Short Term Memory should in theory perform better than RNNs since their cells help solve the vanishing gradient issue that RNNs commonly experience. However, it is not the case with this dataset. With the LSTM having an accuracy of 36.88%, a MSE of 2.14×10^{-4}, and an average daily return of 0.13%.

These results demonstrate that either the model did not have time to converge adequately, the architecture of the model was insufficient in capturing the relations or the data set was too small for sufficient feature extraction.

However, based on the plots, the most likely cause is a model architecture that is unable to map good parameter relations. The loss was seen to converge, with the model making very small improvements per epoch. However, the sudden jump towards the end showcases that the model mapped its relations inaccurately.

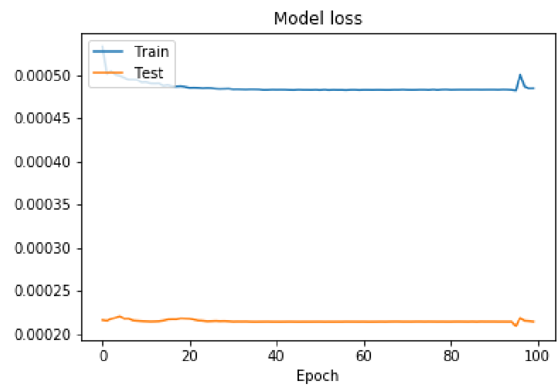

Figure 5: Validation error per epoch for LSTM

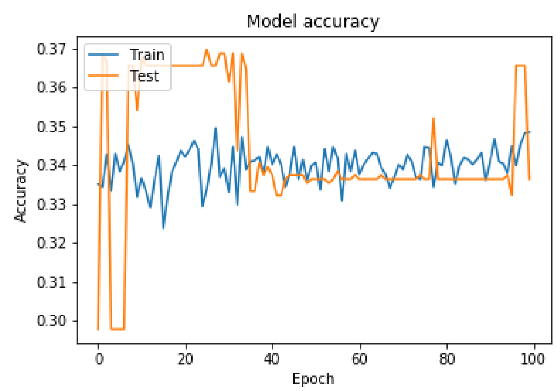

Figure 6: Accuracy per epoch for LSTM

3.4 Comparison

The following table showcases each models result for each criteria. Overall, it is clear that with the given dataset that an RNN is the best model. It is superior to the other three in all three criteria.

Model	Accuracy	MSE	Performance
MLP	37.50%	2.13×10^{-4}	0.27%
RNN	81.25%	1.04×10^{-5}	2.88%
LSTM	36.88%	2.14×10^{-4}	0.13%

Table 1: Comparison of model accuracy, MSE and performance

4 Conclusion

While various models exist in predicting asset prices, three of the most effective models in machine learning are the Multilayer Perceptron, Recurrent Neural Network and Long Short Term Memory. While all three models are effective in finding non-linear relationships between input and output variables, the unique nature of financial asset prices complicates such relationship. From its cyclical nature to its large problem space dimension, asset prices are incredibly difficult to predict accurately and precisely. However, from the three models described, the one most adept at learning and applying these relationships was clearly the RNN. With an accuracy of 81.25%, a MSE of 1.04×10^{-5} and an average daily return of 2.88%.

5 Recommendations

Given the results showcased by all three models, it is clear that the optimal solution would be a Recurrent Neural Net. However, it is important to note that these results are based on a strict set of hyperparameters and with a specific dataset.

5.1 Implementation

An initial implementation would be the 2-layer RNN used in the comparison. Each layer would have 256 units, and the look back window would be of 50 business days. The output of the model would also be a percentage difference between two business days. This is to apply a form of normalization to the data, preventing the exploding gradient problem. While this would be a good starting point, it is important to constantly test out different models as the machine learning landscape is constantly shifting.

To improve this base model, the first step would be to increase the number of inputs going into the model, as simply feeding the past price of the asset is missing a lot of key information. However, there are many possible ways to improve the model.

5.2 Future Research

While the RNN demonstrated very strong results in all three metrics, there are numerous areas which can be explored for improving these results such as hyperparameter tuning and improving the dataset.

Furthermore, there are many other criteria that can be used to compare models, which could help distinguish each model further. There is definitely far more research that can be conducted in finding other criteria to compare and other models to test.

5.2.1 Criteria

While the three criteria effectively capture most important aspects in asset price prediction, there are innumerable criteria that can be used for further analysis. Very important are recall and precision and recall, which are used to measure the rate of false positives and false negatives that the model produces. Another criterion that can be used is asset allocation, which is how the model allocates assets depending on their expected return. This criterion is more inline with financial analysis and would be very useful in allocating assets effectively.

5.2.2 Models

Despite the strong results showcased by the RNN, any shift in either the model or the data could greatly affect the results of each model. If the dataset comprised of more data sets, and a longer time window, it is most likely than an LSTM would then outperform the RNN. If the data had more factors rather than simply the asset price, then the MLP would have performed considerably better.

An important model that was left out are Convolutional Neural Networks (CNN), which utilizes kernels to break down two dimensional data. While traditionally used in image and video analysis, the space dependency used in CNNs could definitely be used in asset price prediction. Another model that has not been considered is the Q-Learning, which falls under reinforcement learning. In this case, the model would teach itself to trade such that it would generate the highest return.

Overall, there are many models that have not been analyzed nor included, as their metrics are too different. However, it is clear that machine learning will continue to have a strong impact in financial asset pricing.

References

[1] W. Sharpe, *Capital asset prices: a theory of market equilibrium under conditions of risk*. Journal of Finance, vol. 19, no. 3, pp. 425-442, 1964.

[2] E. Fama and K. French, *A Five-factor Asset Pricing Model*. Journal of Financial Economics, vol. 116, 10/29 2014.

[3] R. Lawrence, *Using neural networks to forecast stock market prices*. 01/12 1998.

[4] I. Welch and A. Goyal, *A Comprehensive Look at The Empirical Performance of Equity Premium Prediction*. The Review of Financial Studies, vol. 21, no. 4, pp. 1455-1508, 2007.

[5] J. Heaton, N. Polson, and J. H. Witte, *Deep learning for finance: deep portfolios*. Applied Stochastic Models in Business and Industry, vol. 33, no. 1, pp. 3-12, 2017.

[6] C. D'Hondt, R. De Winne, E. Ghysels, and S. Raymond, *Using neural networks to forecast stock market prices*. Available at SSRN 3415981, 2019.

[7] L. Ryll and S. Seidens, *Evaluating the Performance of Machine Learning Algorithms in Financial Market Forecasting: A Comprehensive Survey*. arXiv e-prints. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2019arXiv190607786R